JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 8 (2020), Issue 4 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Guest Editorial

Reviews

Original Papers

Tutorial

Corrigenda and Addendas

Guest Editorial

COVID-19 Contact Tracing and Data Protection Can Go Together

Johannes Abeler[1*], PhD; Matthias Bäcker[2*], PhD, LLM; Ulf Buermeyer[3*], PhD, LLM; Hannah Zillessen[1*], MPhil

[1]Department of Economics, University of Oxford, Oxford, United Kingdom

[2]Rechts- und Wirtschaftswissenschaften, Johannes Gutenberg-Universität Mainz, Mainz, Germany

[3]Gesellschaft für Freiheitsrechte (German Society for Civil Rights), Berlin, Germany

[*]all authors contributed equally

Corresponding Author:
Johannes Abeler, PhD
Department of Economics
University of Oxford
Manor Road
Oxford
United Kingdom
Phone: 44 1865281440
Email: johannes.abeler@economics.ox.ac.uk

Related Article:

Comment on: https://mhealth.jmir.org/2020/4/e18936/

Abstract

We discuss the implementation of app-based contact tracing to control the coronavirus disease (COVID-19) pandemic and discuss its data protection and user acceptability aspects.

(JMIR Mhealth Uhealth 2020;8(4):e19359) doi:10.2196/19359

KEYWORDS

COVID-19; app; contact tracing; proximity tracing; privacy; data protection; Bluetooth

Why Is Contact Tracing Useful?

The coronavirus disease (COVID-19) pandemic is the greatest public health threat that the world has seen in the last 100 years. In response, countries have introduced various levels of "lockdown" to reduce the number of new infections. Lockdowns, however, come at a great cost to workers, firms, and families. Recent epidemiological models also predict that the epidemic will start anew, once the lockdown is lifted [1].

Scientists have thus discussed a second approach to keeping the epidemic in check: app-based contact tracing. Several apps are currently in development (eg, in the United Kingdom [2], by a pan-European initiative [3], and in a joint Google and Apple venture [4]), or have already been launched (eg, in Singapore [5]).

Why would such an app be useful at all? We still don't know many things about COVID-19. The data so far suggest, however, that about half of all infections occur before the dreaded symptoms of fever or a persistent cough appear. It is therefore not enough to quarantine people only after they show symptoms. To reduce infections, one would need to act quickly when a person is diagnosed with COVID-19 to find all people this person was in close proximity with. The risk of infection is highest if one has been within 1.5-2 m of an infected person for at least 10-15 minutes. If it could be determined who had been in such close proximity, then one could ask freshly infected, presymptomatic people to self-isolate and thus stop them from infecting more people. Mathematical models of the pandemic [6] show that fast contact tracing combined with a large-scale virus-testing program might be able to not just delay the epidemic but to stop it entirely. This would also mean that the lockdown measures currently in place around the world could be slowly loosened up again. However, such fast contact tracing is not possible manually. Only a digital, largely automatic solution would help.

Epidemiology Meets Data Protection

Some might argue that the demands of the COVID-19 crisis justify even extreme countermeasures. After all, this is about saving the lives and preserving the health of as many people as possible. Weakening data protection might be preferable to the far-reaching restrictions of personal freedom and to the

economic costs of the current lockdown. In keeping with this, many countries have started tracking their citizens' phones and using location data to monitor the spread of the virus as well as to enforce both lockdown and early isolation restrictions. The most prominent example of this is China [7], where entry to many public places is restricted to people who can show a green health code on their smartphones and thus demonstrate they have not been in contact with a confirmed case of COVID-19. However, countries like Israel [8] or South Korea [9] use location data as well—in the former case, to enforce quarantine rules and notify the contacts of an infected person, and in the latter to warn people before they enter "high risk" zones. However, even in the face of an existential threat, we should interfere with fundamental rights as little as possible. Among the effective approaches, we should choose the one that least compromises fundamental rights. In particular, we believe that swift and efficient contact tracing is possible without collecting extensive amounts of data in a central database.

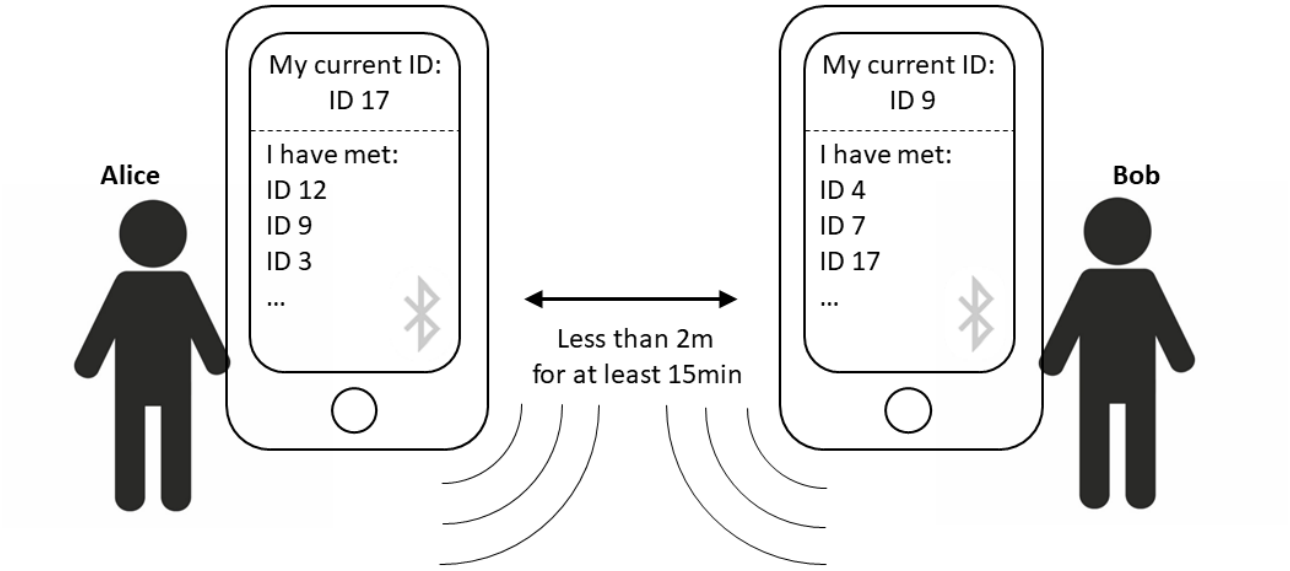
A contact tracing system can be set up in a way that would allow for most data processing to happen locally on users' mobile phones rather than on a central server. Only the notification of users who have been in contact with an infected person would need to be coordinated centrally. Even in this case, the necessary data could be processed in a way that would effectively preclude the central server from identifying users. The system would also not require collecting any location data.

The fundamental idea is simple: it does not matter *where* people get in contact with an infected person. Be it on the bus or at work—what matters is proximity to a contagious person. This means that particularly sensitive location data, such as GPS or radio cell data, is actually neither necessary nor useful. Instead, the only data that matters is whether two people have come into close enough contact to risk an infection. One example of this would be contact tracing based on "contact points" as suggested by Yasaka et al [10]. The smartphone app they propose would allow users to create "checkpoints" by generating a QR code that can be scanned by all other app users when joining their checkpoint. If checkpoints were created for any social interaction, be it among friends and family or in public spaces like a restaurant, then the app could use both this information and voluntary notifications from users should they be diagnosed with the virus to compute transmission graphs. These graphs in turn could let every user know if there were any possible transmission paths leading up to the checkpoints they visited and thus their risk of being infected. The app would not need any location data and in fact wouldn't even require users to register. It would, however, require the active participation of users who would need to either create or join a checkpoint whenever they get close to someone outside their household. Thus, this approach relies on high levels of vigilance and willingness to participate among at least a majority of the population—not only initially but also as the pandemic continues.

Another example for this kind of "privacy by design" COVID-19 tracing approach is the TraceTogether app [5] from the Singaporean government. Unlike the contact point system, it only requires users to enable Bluetooth on their phone. Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) by the European consortium [3], as well as Google and Apple's recently announced joint initiative [4], are following a very similar concept. We present a slightly modified version below.

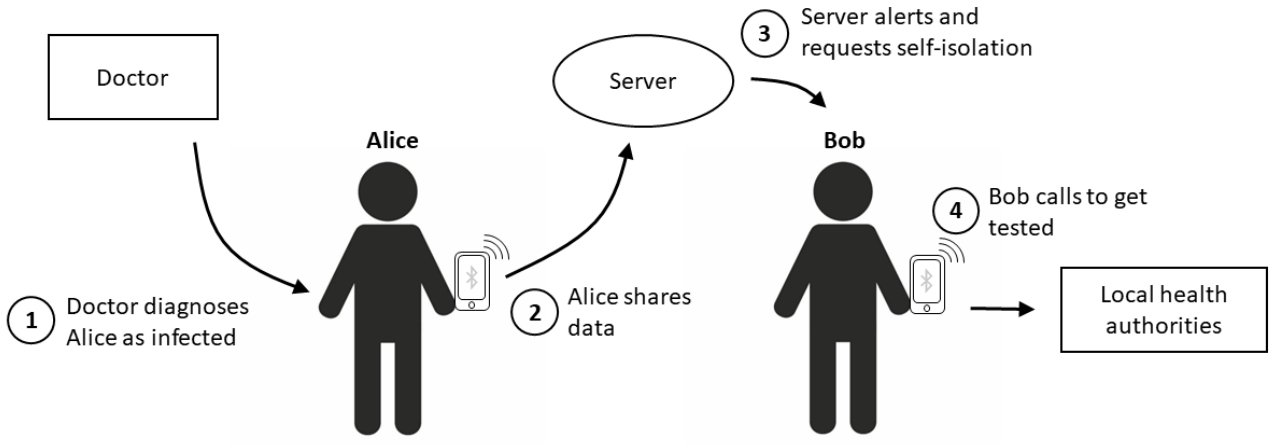
In order to detect whether two people have come into close enough physical proximity to risk an infection, one can use Bluetooth low energy technology. The general drawback of Bluetooth—that it can only reach across a few meters—becomes an advantage here. The tracking itself would work as follows: as many people as possible voluntarily install the app on their phone. The app cryptographically generates a new temporary ID every half hour. As soon as another phone with the same app is in close proximity, both phones receive the temporary ID of the respective other app and record it. This list of logged IDs is encrypted and stored locally on the users' phones (Figure 1).

Figure 1. A COVID-19 tracing approach via Bluetooth. Every mobile phone stores a list of mobile phones that were within 2 m for at least 15 minutes. IDs are temporary but can be decrypted by the server.



As soon as an app user is diagnosed with COVID-19, the doctor making the diagnosis asks the user to share their locally stored data with the central server (Figure 2). If the user complies, the central server receives information on all the temporary IDs the "infected" phone has been in contact with. The server is not able to decrypt this information in a way that allows for the identification of individuals. However, it is able to notify all affected phones. This is because the server does not need any personal data to send a message to someone's phone. The server only needs a so-called PushToken, a kind of digital address of an app installation on a particular phone. This PushToken is generated when the app is installed on the user's phone. At the same time, the app will send a copy of the PushToken, as well as the temporary IDs it sends out over time, to a central server. The server could be hosted, for example, by the Robert Koch Institute for Germany or by the National Health Service for the United Kingdom. This way, it would be possible to contact phones solely based on temporary IDs and PushTokens whilst completely preserving the privacy of the person using the phone.

Figure 2. A user can share their data with the server after receiving a COVID-19 diagnosis. The server then alerts all phones that have been in close proximity to the infected phone. The alerted people would still need to contact their local health authorities, as their identity is not linked to the app.



If a phone has been in close proximity to an "infected" phone, the user of that phone receives a notification together with the request to immediately go into quarantine at home. The user will then need to contact the local health authorities to get tested for the virus as soon as possible so that, depending on the outcome, the user is either able to stop quarantining or all their contacts can be informed (Figure 2).

During the entire process, no one learns the identity of the app user (eg, other users who got in close contact with them, the local health authorities, the central server) since the app is not linked to an identity. Location data is neither recorded nor stored at any point of the process.

As mentioned above, we did not come up with this concept. Singapore introduced a very similar app, and several European countries [3] are working on comparable apps as well. We do not agree with all aspects of the Singaporean app and their practice of contact tracing. For example, every app installation in Singapore is linked with the user's telephone number, making the user identifiable—something that is not strictly necessary and thus, for data protection reasons, should be rejected. Nevertheless, we like the general concept. The recently published PEPP-PT [3] looks promising and might prove to be a legitimate implementation of the privacy-friendly tracing approach outlined above.

Such an app could implement contact tracing much more effectively than a system that relies on radio cell or location data, since neither of these two data sources permit determining a person's position with the necessary precision of 2 m maximum. At the same time, such a concept would comply with existing data protection regulations. Finally, it would work even without users paying constant attention to potentially risky interactions as would be necessary in a contact-point system. Thus, this concept is potentially more robust to fatigue or inattentiveness.

Data Minimization Begets Acceptance

In the case of contact tracing, the approach that requires the least amount of data also seems to be the most effective epidemiologically. This is because an app like the one described above would be better suited to determine who actually was in close proximity than any of the other proposed solutions. Moreover, even digital contact-tracing systems need users to cooperate (by installing the app and carrying their phones with them) for any chance of success. Consequently, the effectiveness of any contact-tracing system depends on public support. There is reason to believe that the level of support can be increased by opting for a data-minimizing solution. A representative survey [11] across the United States, United Kingdom, Germany, Italy, and France shows that about 70% of respondents would install an app like the one described above on their phones (disclosure: co-author JA was also the lead author of the survey study). The reason most frequently brought up against an installation is the worry that the government could use the app as an excuse for greater surveillance after the end of the epidemic. If the government wants as many people as possible to install the app, it should take these concerns seriously and refrain from using location data. Contact tracing works without it.

Conclusion: Proportionality Instead of "Whatever It Takes"

In the current crisis, we will have to endure more and deeper encroachments on fundamental rights than we are used to. Still, there is no reason to tolerate such encroachments to a greater extent than strictly necessary. Even under the current time pressure, it is important to find solutions that minimize data processing as far as possible. We have shown above that this is possible for the case of contact tracing. As the pandemic progresses, many other challenges will emerge. For each of them, one will have to check which data processing is necessary to address them and which ones can be avoided.

Trying to find a data-minimizing solution does not just protect fundamental rights. Such solutions will often increase the effectiveness and efficiency of the respective data-processing system. Only if people trust a system—because it does not spy on them—will the system find broad support in the population.

Conflicts of Interest

None declared.

References

1. Ferguson N, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, Imperial College COVID-19 Response Team. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London 2020:1-20 [FREE Full text] [doi: 10.25561/77482]
2. Rapson J. HSJ. 2020 Mar 18. NHS developing coronavirus contact tracking app URL: https://www.hsj.co.uk/free-for-non-subscribers/nhs-developing-coronavirus-contact-tracking-app/7027163.article
3. PEPP-PT. Pan-European Privacy-Preserving Proximity Tracing. 2020. URL: https://www.pepp-pt.org/
4. Google Inc. Apple and Google partner on COVID-19 contact tracing technology. 2020 Apr 10. URL: https://www.blog.google/inside-google/company-announcements/apple-and-google-partner-covid-19-contact-tracing-technology/
5. Government of Singapore. TraceTogether. URL: https://www.tracetogether.gov.sg/
6. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 Mar 31 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
7. Associated Press. Time. 2020 Apr 02. Inside China's Smartphone 'Health Code' System Ruling Post-Coronavirus Life URL: https://time.com/5814724/china-health-code-smartphones-coronavirus/

8. Tidy J. BBC News. 2020 Mar 17. Coronavirus: Israel enables emergency spy powers URL: https://www.bbc.com/news/technology-51930681
9. Cuthbertson A. Independent. Coronavirus Apps Let People Avoid High-Risk Locations in South Korea URL: https://www.independent.co.uk/life-style/gadgets-and-tech/news/coronavirus-news-app-south-korea-latest-cases-deaths-location-a9371651.html
10. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
11. OSF. 2020 Mar 26. covid19_app_survey URL: https://osf.io/7vgq9/

Abbreviations

COVID-19: coronavirus disease
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing

Edited by G Eysenbach; submitted 14.04.20; this is a non–peer-reviewed article;accepted 15.04.20; published 20.04.20.

Please cite as:
Abeler J, Bäcker M, Buermeyer U, Zillessen H
COVID-19 Contact Tracing and Data Protection Can Go Together
JMIR Mhealth Uhealth 2020;8(4):e19359
URL: http://mhealth.jmir.org/2020/4/e19359/
doi:10.2196/19359
PMID:32294052

Review

Smartphone Apps for the Treatment of Mental Disorders: Systematic Review

Ignacio Miralles[1*], PhD; Carlos Granell[1*], PhD; Laura Díaz-Sanahuja[1*], PsyM; William Van Woensel[2*], PhD; Juana Bretón-López[1,3*], PhD; Adriana Mira[4*], PhD; Diana Castilla[3,4*], PhD; Sven Casteleyn[1*], PhD

[1]Universitat Jaume I, Castellón de la Plana, Spain

[2]Dalhousie University, Halifax, NS, Canada

[3]CIBER of Physiopathology of Obesity and Nutrition CIBERobn, Castellón, Spain

[4]Department of Personality, Evaluation and Psychological Treatment, University of Valencia, Valencia, Spain

[*]all authors contributed equally

Corresponding Author:
Ignacio Miralles, PhD
Universitat Jaume I
Av Vicent Sos Baynat, s/n
Castellón de la Plana, 12071
Spain
Phone: 34 964728000
Fax: 34 964729016
Email: mirallei@uji.es

Abstract

Background: Smartphone apps are an increasingly popular means for delivering psychological interventions to patients suffering from a mental disorder. In line with this popularity, there is a need to analyze and summarize the state of the art, both from a psychological and technical perspective.

Objective: This study aimed to systematically review the literature on the use of smartphones for psychological interventions. Our systematic review has the following objectives: (1) analyze the coverage of mental disorders in research articles per year; (2) study the types of assessment in research articles per mental disorder per year; (3) map the use of advanced technical features, such as sensors, and novel software features, such as personalization and social media, per mental disorder; (4) provide an overview of smartphone apps per mental disorder; and (5) provide an overview of the key characteristics of empirical assessments with rigorous designs (ie, randomized controlled trials [RCTs]).

Methods: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for systematic reviews were followed. We performed searches in Scopus, Web of Science, American Psychological Association PsycNET, and Medical Literature Analysis and Retrieval System Online, covering a period of 6 years (2013-2018). We included papers that described the use of smartphone apps to deliver psychological interventions for known mental disorders. We formed multidisciplinary teams, comprising experts in psychology and computer science, to select and classify articles based on psychological and technical features.

Results: We found 158 articles that met the inclusion criteria. We observed an increasing interest in smartphone-based interventions over time. Most research targeted disorders with high prevalence, that is, depressive (31/158,19.6%) and anxiety disorders (18/158, 11.4%). Of the total, 72.7% (115/158) of the papers focused on six mental disorders: depression, anxiety, trauma and stressor-related, substance-related and addiction, schizophrenia spectrum, and other psychotic disorders, or a combination of disorders. More than half of known mental disorders were not or very scarcely (<3%) represented. An increasing number of studies were dedicated to assessing clinical effects, but RCTs were still a minority (25/158, 15.8%). From a technical viewpoint, interventions were leveraging the improved modalities (screen and sound) and interactivity of smartphones but only sparingly leveraged their truly novel capabilities, such as sensors, alternative delivery paradigms, and analytical methods.

Conclusions: There is a need for designing interventions for the full breadth of mental disorders, rather than primarily focusing on most prevalent disorders. We further contend that an increasingly systematic focus, that is, involving RCTs, is needed to improve the robustness and trustworthiness of assessments. Regarding technical aspects, we argue that further exploration and

innovative use of the novel capabilities of smartphones are needed to fully realize their potential for the treatment of mental health disorders.

(JMIR Mhealth Uhealth 2020;8(4):e14897) doi:10.2196/14897

KEYWORDS

mental health; mental disorders; treatment; intervention; mHealth; smartphone; mobile phone; mobile apps; systematic review

Introduction

Background

The popularity of smartphones has skyrocketed over the last decade. Different sources estimate that around 35% of people worldwide own a smartphone; even much higher penetration rates are reported in developed countries (ranging from 76% in the United Kingdom to 95% in South Korea) [1,2]. Smartphones are fast becoming the most common mobile phone, even in emerging economies [2]. Smartphones distinguish themselves from other types of mobile phones in several ways: (1) improved modality (screen and sound), interaction, and computational resources, which support sophisticated software applications called (mobile) apps; (2) built-in mobile sensors, which allow apps to access various measurements, such as the user's current position, motion, ambient light, and sound; and (3) connectivity hardware (Wi-Fi and Bluetooth), which allows virtually ubiquitous internet connections, as well as connections to nearby wireless hardware (eg, headsets and physiological sensors). A variety of mobile apps have been developed, conveniently installable from so-called app stores, which address a wide range of personal, entertainment, and business needs. In 2017, 2.8 and 2.2 million apps were available from the Google Play and Apple App stores, respectively; collectively, these were downloaded a total of 178.1 billion times [3].

Researchers quickly realized the potential of mobile apps in health [4] and mental health [5], with systematic reviews on related research, that is, targeting mobile health (mHealth) apps, appearing as early as 2011 [5]. Whereas traditional telehealth [6] and cellphone-based [7] systems are limited to SMS, telephone, or video calls, smartphones present a more versatile, powerful, and personalized platform for a holistic set of care tasks, including patient screening, symptom and disorder assessment, psychoeducation, intervention delivery, progress monitoring, and relapse prevention [5]. By providing these health tasks via smartphone apps, albeit partially or combined with a therapist's intervention, a number of obstacles for mental health care are reduced such as therapist workload, lack of qualified personnel, geographic barriers, and attitudinal barriers to seek treatment. New opportunities arise as well, such as improving assessment by leveraging built-in smartphone sensors (eg, biofeedback and motion) and analyzing device usage, and providing ecological interventions directly to the patient when they are most needed, as determined by in situ assessments [8-14].

This is a timely evolution, as reported mental health problems are becoming increasingly prevalent. Trautmann et al [15] estimated that over 50% of the population of high- and middle-income countries suffers from at least one mental disorder in their life, with a significant impact on their quality of life and an overall annual economic cost of US $2.5 trillion (2010) and rising. According to the latest US annual survey [16], there is an estimated 12-month mental disorder prevalence of 18.3% among adults (4.2% for serious mental illness). Mobile mental health interventions have reported promising mental health outcomes [17-19], large acceptance rates by patients [20], and increased sustainability and preservation of treatment effects [21]. Hence, owing to their ability to reduce obstacles for mental health care, these interventions can be leveraged to meet present-day mental health challenges. Nevertheless, we find that the possibilities of current smartphone technology have only just been tapped, and further research is needed to explore them fully [22], as are studies to rigorously analyze the empirical effectiveness of these systems [22,23]. For driving and steering such future research, there is a continuous need to establish a state of the art, which comprehensively reviews current focal points on psychological (ie, type of disorder and evaluation) and technological factors (smartphone capabilities, technologies, and features used). Such a review should include both exploratory research, which investigates technological opportunities, and empirical research, which establishes robust empirical evidence for the efficacy of smartphone interventions. Previous mobile mental health reviews have become dated [24-26], while more recent studies only consider specific mental disorders; for example, cognitive impairment [27], alcohol and substance abuse [28], anxiety [17]; only consider technologies, for example, text messaging [29] and SMS messages [11]; or focus solely on efficacy, usability, and feasibility of interventions realized by mHealth (mobile health, referring to the use of mobile computing and communication technologies in health care [29]) interventions [11,17,27,30,31].

Objectives

We provide a systematic review that studies the recent (2013-2018) research on smartphone app-based interventions for mental disorders. Specifically, we aimed to analyze and summarize relevant research to (1) analyze the coverage of mental disorders in research articles per year; (2) study the types of assessment in research articles per mental disorder per year; (3) map the use of sensors, software features, and analytical capabilities of smartphones per mental disorder; (4) provide an overview of mobile smartphone apps per mental disorder; and (5) provide an overview of the key characteristics from empirical assessments with rigorous designs (ie, randomized controlled trials [RCTs]). As a counterbalance to our focus on smartphone interventions, we also briefly discuss potential risks such as lack of proven effectiveness, possibility for harm, and breach of privacy.

Methods

Search and Study Selection

This systematic review uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [32] as a guideline. We performed an extensive search of scientific databases, that is, Scopus, Web of Science (WoS), American Psychological Association (APA) PsycNET, and Medical Literature Analysis and Retrieval System Online (MEDLINE), using queries that combined search terms related to the psychological (eg, psychology, psychological, mental disorder and intervention) and the technological dimensions (eg, mobile device, smartphone and mHealth) using logical operators. All database-specific queries were semantically equivalent but formulated using the different syntaxes and technical support of the respective search engines. The queries were launched on March 9, 2018, covering results from 2013 until March 2018, and relaunched on July 13, 2019, to cover the full year of 2018. Keywords and queries can be found in Multimedia Appendices 1 and 2.

All resulting publications were downloaded, and duplicates were removed. All papers were equally divided among four multidisciplinary groups of two members, each comprising one computer scientist and one psychologist. Publications were initially screened based on the inclusion/exclusion criteria (IC/EC), using title, abstract, and keywords. Subsequently, papers that were still inconclusive, that is, after initially screening for their title, abstract, and keywords, were fully reviewed to check their eligibility using the IC/EC. Both during initial screening and full-text screening for eligibility, both team members processed the group's assigned papers independently and discussed their observations before making a final decision. In case of disagreement, a third reviewer was assigned, and a final decision was made collaboratively.

Inclusion Criteria

Articles fulfilling all the following IC were included in our systematic review: (*IC1*) Full research articles published in an international journal or conference proceedings between January 1, 2013, and December 31, 2018, written in English, and where a full text was available; (*IC2*) Primary research articles, that is, articles that produce first-hand contributions to the research field; (*IC3*) Articles explicitly describing the use of a smartphone app for the delivery of psychological intervention(s) for mental disorder(s), whereby (a) smartphones are used as delivery platform and at least one smartphone-specific feature is used, thus going beyond regular mobile phone features (eg, SMS messages and phone calls) and standard content delivery (eg, nonmobile and generic websites); (b) the targeted mental disorders are found in the Diagnostic and Statistical manual of Mental Disorders (DSM-5) [33]; and (*IC4*) Articles including either exploratory research (ie, investigating technological opportunities) or empirical research (ie, establishing robust empirical evidence). For *exploratory research*, an explicit description on the use of the smartphone app for a psychological intervention is required. For *empirical research*, there were no restrictions on study design. Study protocols were also included.

Exclusions Criteria

EC were all sources that do not comply with the IC: (*EC1*) All research articles published before 2013 or after 2018, not written in English, not published as a full paper in an international journal or conference. This excludes articles published in any other outlet, such as workshops, discussion forums, colloquia, patent descriptions, white papers, and other types of publications, for example, posters, demo papers, tutorial paper, editorials, or extended abstracts; (*EC2*) All secondary research articles, that is, articles that use primary research articles to derive results such as reviews, systematic maps, meta-analysis, synthesis, and comments; (*EC3*) Any article not explicitly describing the use of smartphones as the primary mode of delivering psychological interventions for mental disorders. This excludes articles addressing nonmental disorders (eg, cancer) or symptoms (eg, stress), as well as articles describing the use of other mobile devices (eg, wearables, smart watches, and tablets) or using smartphones only as a regular phone (eg, SMS messages and phone calls); and (*EC4*) Any article that only superficially describes the application of a smartphone app to a mental disorder—that is, without providing empirical evaluation data, or lacking a detailed description on the use of the smartphone app for delivering psychological interventions for mental disorders. This includes philosophical papers, vision papers, or papers solely focusing on a technical innovation without an accompanying mobile app and/or targeted mental disorder.

Classification of Studies

All included studies were classified according to technology- and psychology-related dimensions. Additionally, we recorded the name of the app as well.

The technology-related dimensions included the following: (1) built-in sensors: accelerometer, gyroscope, GPS, microphone, and camera; (2) software features: prompting (any kind of proactive prompting to the patient, for example, reminders, notifications, or motivational messages), health care provider communication (directly communicating with a health care provider through the mobile app), progress (allowing patients to monitor their progress throughout the intervention), assessment (capability to [psychologically] assess the patient, including self-assessment [eg, questionnaire] and automatic assessment [eg, based on smartphone usage patterns]), social (availability of social networking and peer communication, such as forums, chat, messaging, and sharing of experiences or information sources), personalization (ability to customize/personalize some aspects of the mobile app toward the patient), learning (any kind of learning material or support presented to the patient), in situ use (explicit support for using the mobile app in the patient's natural environment [ecological], that is, which allows real-time [momentary] interventions when they are most needed), gamification (use of game elements and principles), context awareness (capability of detecting the context/environment of the patient, for example, location, ambient sound, and text/call history), virtual reality (VR, use of virtual environments as delivery paradigm), and augmented reality (use of augmented environments as delivery paradigm); and (3) analytics: use of advanced software algorithms in the

mobile app or supporting infrastructure (ie, server side)—including machine learning, behavioral analysis, activity analysis, and spatial analysis.

The psychology-related dimensions included the following: (1) mental disorders: the considered mental disorders are based on DSM-5 [33]. In addition to the well-established diagnosis categories from DSM-5, we also considered a *suicidal behavior disorder/non suicidal self-injury* category, as this condition is very well represented in the literature and recognized as a condition for further study in DSM-5 (ie, likely to be included in future versions). In cases where the smartphone app focuses on multiple disorders, we distinguished between (a) comorbid disorders, that is, those specifically focusing on comorbidity, and (b) various disorders, that is, those delivering treatment(s) for different disorders (not co-occurring, that is, in different patients); and (2) approaches to psychotherapy: the different approaches to psychotherapy are based on the existing theories, which guide psychologists through the process of understanding patients and their mental disorders and developing solutions. Taking into account different treatment modalities and psychological frameworks, approaches to psychotherapy fall into eight broad categories: cognitive behavioral therapies, humanistic therapies, systemic therapies, psychoanalysis therapies, third wave therapies, transdiagnostic therapies, positive psychotherapy, and others.

Finally, the study-related dimensions included the assessment type: *Effect*, *Usability/user experience* (short: *Usability/UX*), *Effect and Usability/UX* and *No Assessment*. *Effect* indicates that the authors reported results about the smartphone app's effects on the participants' clinical symptomatology. *Usability and user experience*, as defined by ISO 9241-210:2010, that is, the International Standard on Ergonomics of human System Interaction [34], indicates that the authors assessed variables such as usability, user acceptance, opinion and satisfaction, feasibility, and intention to use. *Effect and Usability/UX* denotes that the authors assessed *Effect* as well as *Usability and UX*. Finally, *No Assessment* refers to those cases where no assessment was reported, for example, including study protocols or technical descriptions of the smartphone apps delivering psychological interventions for mental disorders.

Data and Software Availability

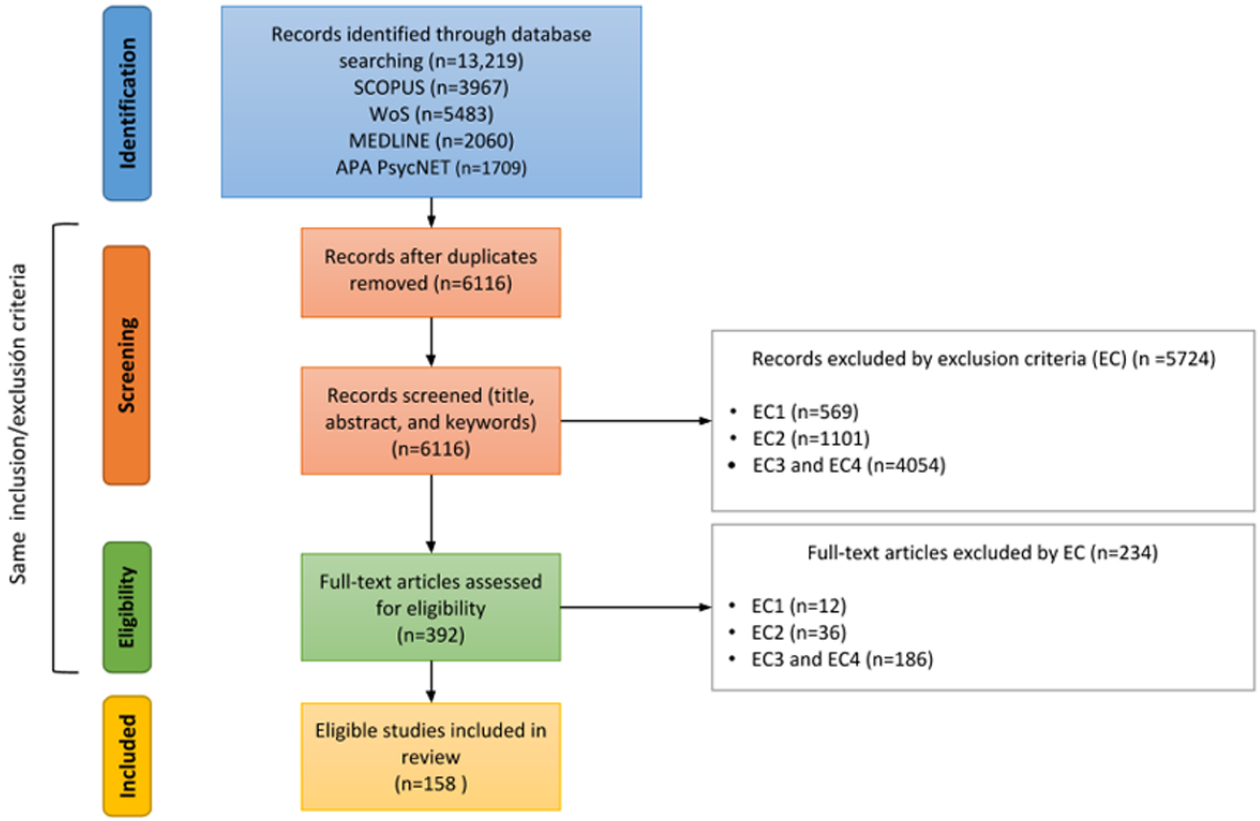
For transparency and reproducibility, we published the resulting data, code, and instructions on GitHub (San Francisco, California) and archived the work in Zenodo [35]. The GitHub repository includes a literate programming document that combines text, data preprocessing, analysis, and visualizations.

Results

Study Inclusion

Figure 1 shows the results of the systematic review processes according to the PRISMA data flow chart. During the *identification* phase, we identified 13,219 studies from the four different Web-based sources (Scopus, WoS, APA PsycNET, and MEDLINE), which we reduced to 6116 after removing duplicates. After the *screening* phase, that is, based on title, abstract, and keywords, we retained 392 articles. The *eligibility* assessment, that is, based on the full paper, led to a final set of 158 papers. More details can be found in Figure 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for the systematic review. WoS: Web of Science.



Evolution of Research and Types of Assessment

Figure 2 shows the temporal distribution of research over the study period 2013-2018, along with their reported assessment type. Overall, we observed a positive evolution of the amount of research over time, steadily increasing from only few (7) articles in 2013 to a much larger amount (60) in 2018.

In Table 1, we show the distribution of assessment types for the reviewed studies. The majority of articles (113/158, 71.5%) reported some kind of assessment. Looking at the distribution of assessment types over time (Figure 2; percentages), we observe an overall slow proportional increase of studies with an assessment (2015 appears to have been an outlier). Regarding the type of assessment, we observe that only a fifth of the articles with assessment (22/113, 19.5%) focus specifically on the effect of intervention on clinical symptomatology (15/113, 13.9% of all studies). Although we see an absolute increase over the last 2 years (in line with the overall increase of studies in general), the sharp increase in 2017 could not be confirmed in 2018 in proportional terms. The proportional amount of usability/UX assessments steadily rose over the years, with an outlier in 2016, where it counteracted a sharp drop in mixed assessments. Caution should be taken with interpreting and generalizing these results; additional data over a larger timeframe are needed.

Orthogonal to the general type of assessment, we also considered other characteristics of the assessment—that is, whether it features an RCT over a long timeframe, or a pilot RCT; or supplies less empirically rigorous results, such as qualitative studies, feasibility studies, case studies (eg, n of one clinical trials) or usability studies. From Table 1, we observe that only a small minority of studies performed an RCT assessment (25/113, 22.1% of all studies with assessment; 25/158, 15.8% overall) and only a handful of papers (7/113, 6.2%; 7/158, 4.4% overall) performed a pilot RCT. Moreover, only a minority of all RCTs (9/25, 36%; 9/158, 5.7% overall) and pilot RCTs (2/7, 29%; 2/158, 1.3% overall) were focused specifically on effect assessments.

Figure 2. Temporal trend and number of articles published per assessment type.

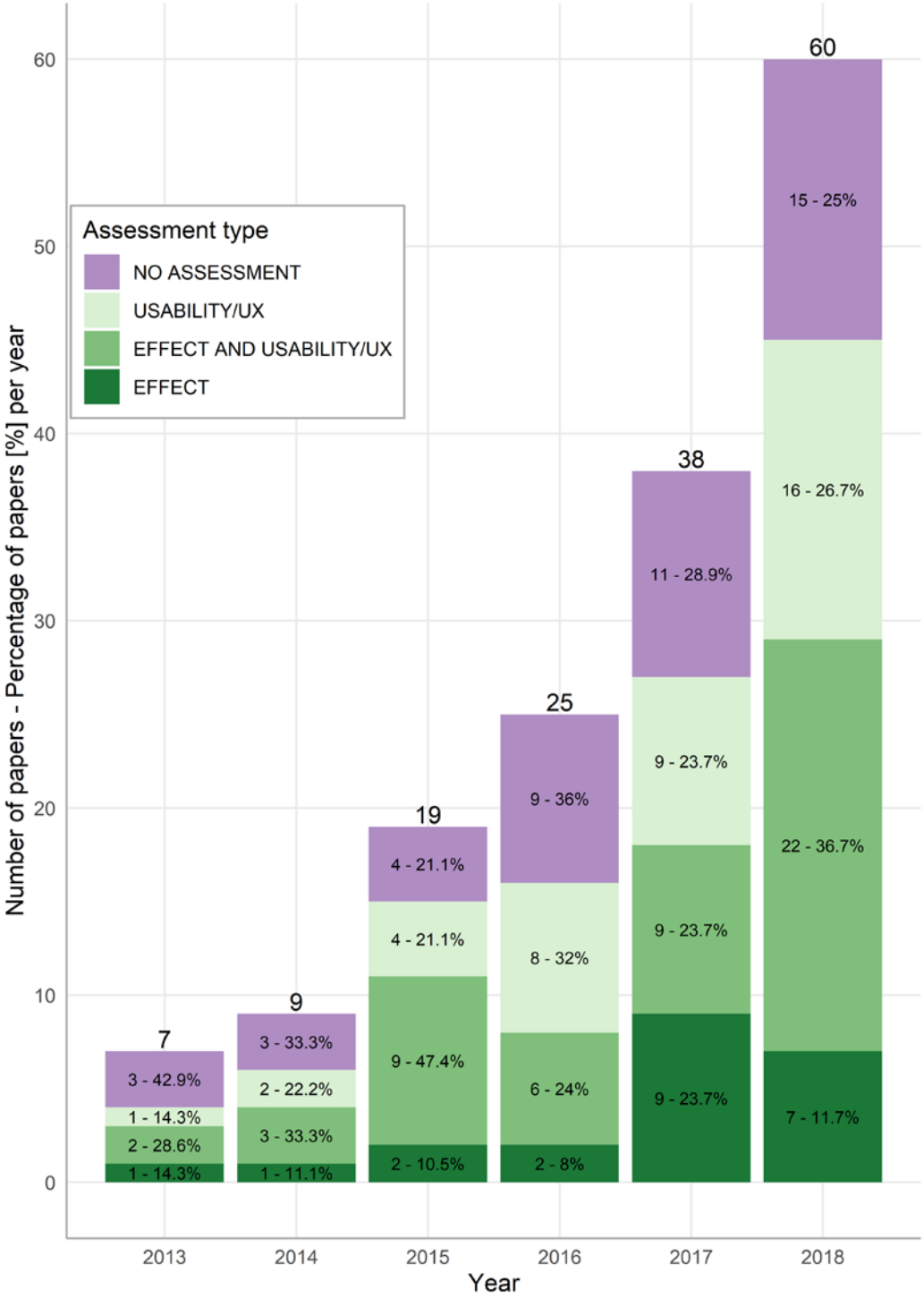


Table 1. Distribution of assessment type.

Assessment types	All, n (%)	RCT[a], n (%)	Pilot RCT, n (%)
No assessment (total)	45 (28.5)[b]	N/A[c]	N/A
Assessment (total)	113 (71.5)[b]	25 (22.1)[d]	7 (6.2)[d]
Usability/UX[e]	40 (35.4)[d]	2 (8)[f]	0 (0)[f]
Effect + usability/UX	51 (45.1)[d]	14 (56)[f]	5 (71)[f]
Effect	22 (19.5)[d]	9 (36)[f]	2 (29)[f]

[a]RCT: randomized controlled trial.

[b]Percentage based on the total number of studies (N=158).

[c]N/A: not applicable.

[d]Percentage based on the number of studies with an assessment (N=113).

[e]UX: user experience.

[f]Percentage based on the number of RCT studies (N=25) and Pilot RCT studies (N=7), respectively.

Covered Mental Disorders

Figure 3 shows the number of studies per mental disorder, ranked in ascending order and subcategorized according to the type of assessment. *Depressive disorders* (31/158, 19.6%) is the most commonly addressed mental disorder. Note that the category of *various disorders* includes apps addressing multiples disorders, where serious mental illness, *depressive and anxiety disorders* are most represented. Collectively, the top six mental disorders account for 73.4% (116/158) of all studies included in the search. Regarding *comorbid disorders*, we point out that the majority of papers were related to a specific dual pathology, that is, where a psychological disorder coexisted with the abuse of substances. One case in this category was focused on *neurodevelopmental disorders and elimination disorders*. For all remaining mental disorders from DSM-5 [33] (not shown in Figure 3), we did not find studies that met our IC.

Figure 3. Distribution of articles per mental disorder, categorized according to assessment type. Aggregated results of assessment types: No assessment 45/158 (28.5%), Usability/UX 40/158 (25.3%), Effect + usability/UX 51/158 (32.3%), Effect 22/158 (13.9%).

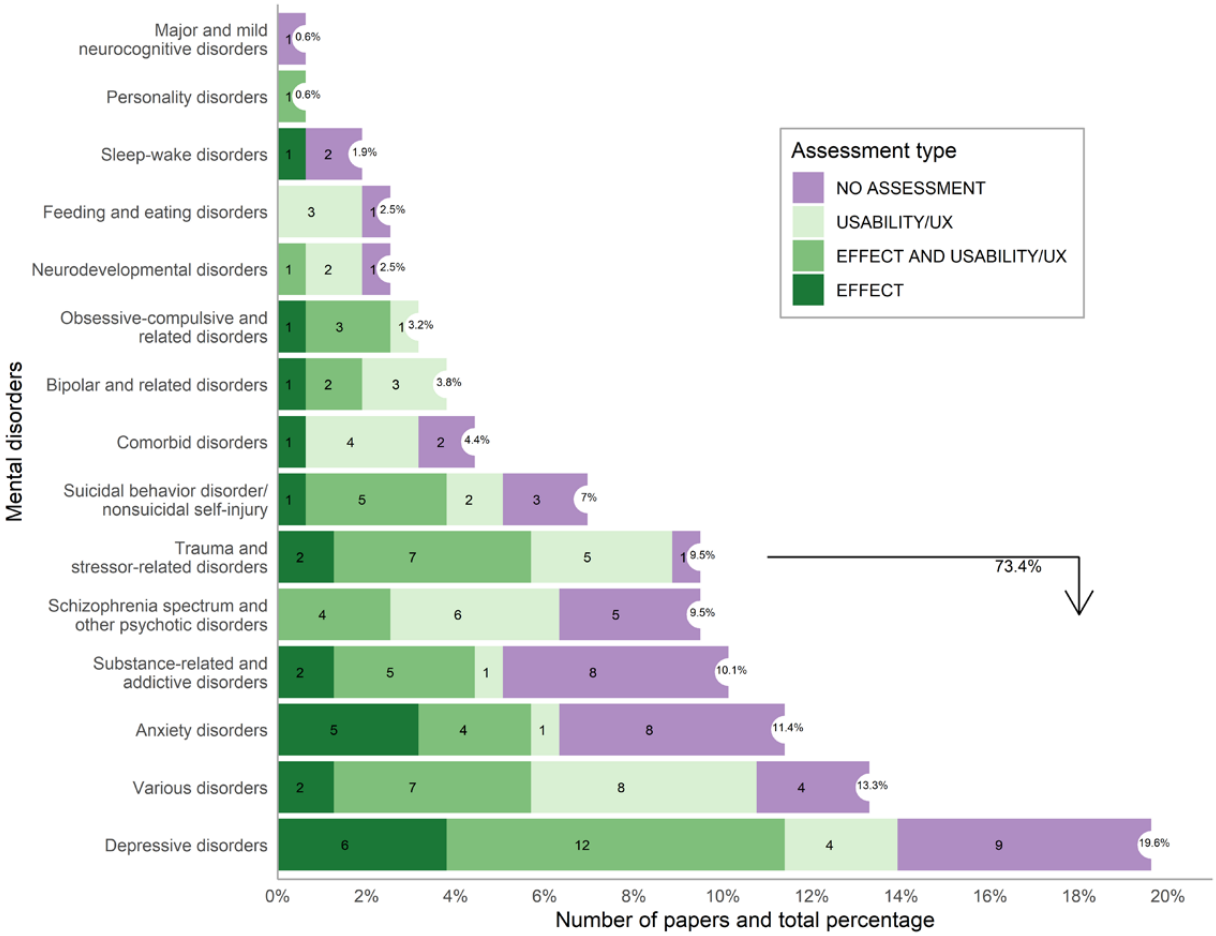
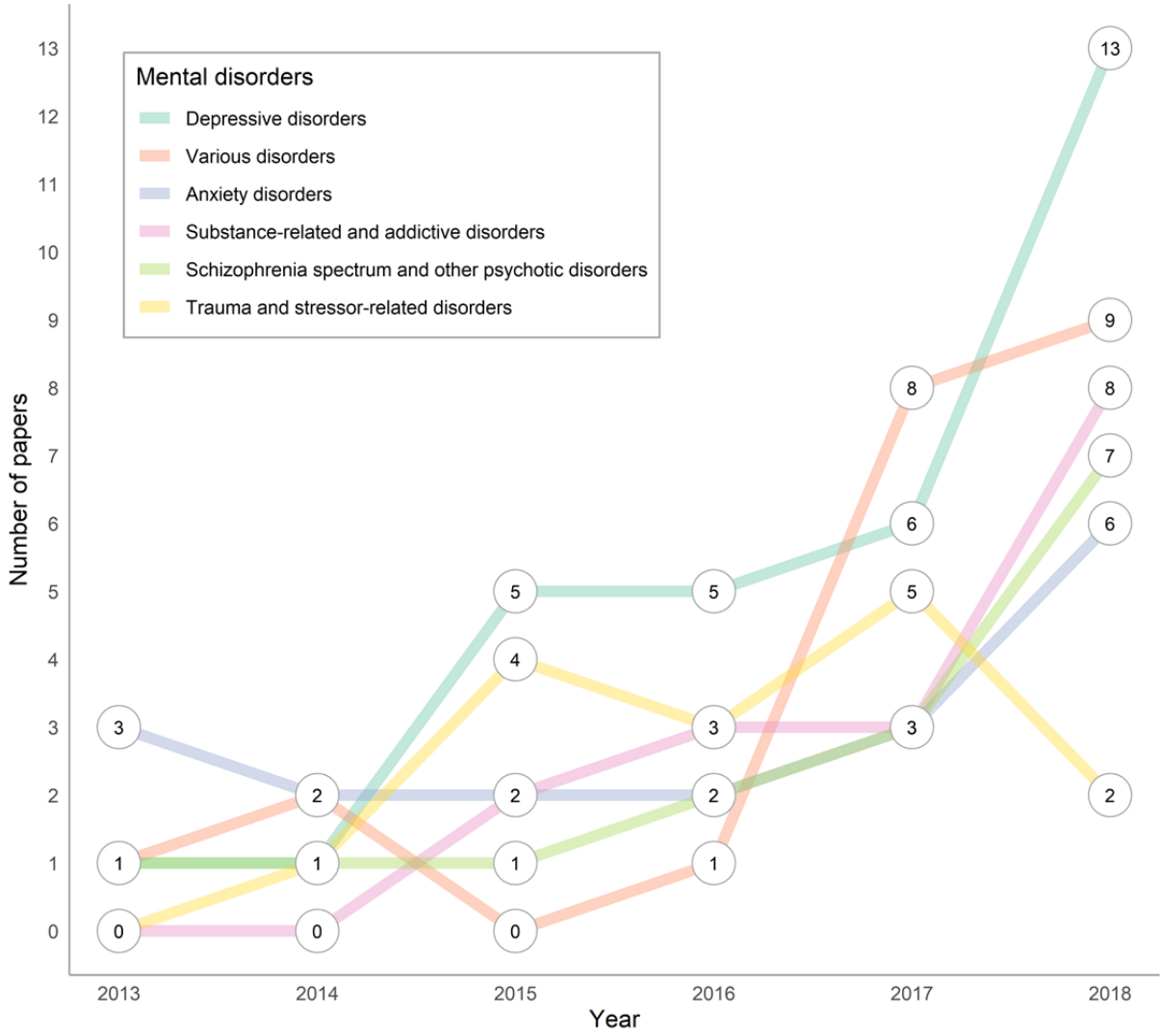


Figure 4 shows the temporal trend of the top six mental disorders targeted by studies over the period 2013-2018. Overall, we observed an increasing number of published articles related to the top six mental disorders over time, with a significant increase since 2015. We also noted that the relative ranking of the top six mental disorders is largely maintained since 2015, with two notable exceptions: *trauma and stressor-related disorders* sharply decreased in 2018, and *various disorders* (ie, the app can be utilized to target multiple independent [noncomorbid] disorders) significantly increased in the last 2 years, reaching the first and second positions, respectively. Finally, we point out the doubling of research on *depressive disorders*—which was already well researched previously—in 2018, and the fact that research on *substance-related and addictive disorders* only started in 2015, yet it has been steadily growing since to reach the third position in 2018.

Figure 4. Temporal distribution of articles published for the top 6 mental disorders.



Relation Between Assessment Type and Mental Disorder

In Figure 3, we observe multiple effect assessments for all top six disorders, except for *schizophrenia spectrum and other psychotic disorders* (0/15, 0%). The *depressive* and *anxiety disorders* are relatively well studied for effect assessment; 19% (6/31) and 28% (5/18) of assessments pertain to effect, respectively. On the other hand, effect is assessed only marginally for *trauma and stressor-related disorder* (13.3%), *substance-related and addictive disorder* (12%, 2/16), and *various disorders* (9%, 2/21). For less addressed disorders (ie, not in the top six), we only see one or no effect assessment.

Regarding other types of assessment, no clear patterns can be observed, and we fall back to individual observations. Remarkable are the high number of *mixed* assessments for *trauma and stressor-related disorders* (47%, 7/15) and to a lesser extent *depressive disorders* (39%, 12/31); the low number of pure usability/UX assessments for *anxiety disorders* (5%, 1/18), *substance-related and addictive disorders* (6%, 1/16), and *depressive disorders* (13%, 4/31); and the large number of articles without any assessment for *anxiety disorders* (44%, 8/18), which contrasts the high number of effect assessments.

Coverage of Technical Features Per Mental Disorder

Figure 5 plots the technology-related dimensions, namely, software features implemented by the studied apps (in orange), the utilized built-in sensors (in green), and analytics (in blue), vs the type of mental disorders. In doing so, the figure shows to which extent, and for which disorder(s), the state of the art is leveraging hardware- and/or software-related smartphone capabilities. Within each technology-related dimension (X axis), features are ranked by their decreasing popularity over all mental disorders (left-right; occurrence count is shown at the top of each column); mental disorders (Y axis) are similarly ordered by decreasing popularity in literature (bottom-up).

Figure 5. Bubble plot representing technology-related dimensions (software features—orange; built-in sensors—green; analytics—blue) vs mental disorders. Bubble size corresponds with the number of articles.

We note that larger bubbles tend to be concentrated at the bottom of the graph, as the most popular mental disorders have a higher number of articles, which also tend to cover more varied technical dimensions. The 7×12 vertical rectangle illustrates the top seven software features that are being leveraged for the majority of mental disorders (ie, 12 out of 15). These features are mostly related to intervention-specific features, such as learning and in situ use, and communication features such as prompting. The 9×6 rectangle shows nine software features that have full coverage for the top six mental disorders; it also includes social and gamification features. Regarding the delivery paradigm, virtual and augmented realities are each leveraged for only one mental disorder. Finally, regarding built-in sensors and analytics, we observe a much lower and dispersed coverage without clear patterns (especially for analytics). GPS stands out, with a relatively wide spread over mental disorders (ie, 9 out of 15).

Concrete Studies Per Mental Disorder and App Name

For the benefit of the research community, Table 2 lists the concrete studies (by app name, when available; if the app name was not found, we put N/A) per mental disorder. Apps that are the subject of multiple studies are shown in *italics*. On the one hand, it can be observed that, independent of the mental disorder, most studies utilized a custom-made app, which was not being assessed in other studies. On the other hand, this implies that there exists a wide variety of apps, even for the same mental disorder. Remarkably, for *trauma and stressor-related disorders*, most apps were assessed in multiple studies. Highlighting some interesting cases, Koroko-App (*depressive disorders*) is an app that was rigorously tested both for effect and usability/UX; a study protocol was published, followed by two RCTs assessing effect and usability/UX issues. Other apps, such as post-traumatic stress disorder coach and Simply Yoga (*trauma and stressor-related disorders*), combine RCTs with other type(s) of assessments for effect and usability/UX. Some apps such as the Blue Ice app (*suicidal behavior disorders/nonsuicidal self-injury*) published assessments using non-RCT designs.

Table 2. Apps and studies grouped by mental disorder (apps covered in multiple publications are in italics).

Mental disorder	References by app
Major and mild neurocognitive disorders	Rico [36]
Personality disorders	EMOTEO [37]
Sleep-wake disorders	Sleepcare [38], SleepIO [39,40]
Feeding and eating disorders	Jorvie [41], Student Bodies–Eating Disorders [42], *Recovery Record* [43,44]
Neurodevelopmental disorders	iCanLearn [45], LifePal [46], My MFG [47], TimeOut [48]
Obsessive-compulsive and related disorders	Geo-Feedback App [49], GGOC [50], Live OCD Free [51], Mayo Clinic Anxiety Coach [52], RAW HAND [53]
Bipolar and related disorders	MyT [54], PRISM [55], *SIMPLe* [56-59]
Comorbid disorders	CASA-CHESS [60], Enuresis Trainer [61], Learn To Quit [62], SMI-CM [63], Stay Quit Coach [64], Stop-Cannabis [65], N/A[a] [66]
Suicidal behavior disorder/nonsuicidal self-injury	BackUp; mEMA [67], BeyondNow [68], BRITE [69], iBobbly [70], *Virtual Hope Box (VHB)* [71,72], N/A [73,74], *Blue Ice* [75-77]
Trauma and stressor-related disorders	RELAX [78], N/A [79], *Life Armor; PE Coach; Positive Activity Jackpot; Eventful; Tactical Breather; VHB; Daily Yoga; Simply Yoga* [79,80],*PE Coach* [81-84], *PTSD Coach* [85-91]
Schizophrenia spectrum and other psychotic disorders	iCOPE [92], MindFrame [93], movisenseXS [94], RealLife Exp [95], SlowMo [96], TechCare [97], Temstem [98], *Actissist* [99,100], *FOCUS* [101,102], *Heal Your Mind* [103], *PRIME* [104,105]
Substance-related and addictive disorders	Drink Less [106], Fit&Sobber [107], Mind the Moment [108], S-Health [109], SEVA [110], SmartQuit [111], Smoke Mind [112], Social-Local-Mobile [113], *A-CHESS* [114,115], *CET App* [116,117], *Kick.it* [118,119], *Smart-T* [120,121]
Anxiety disorders	Agoraphobia Free; Stress Free [122], Ångesthjälpen [123], AnxietyCoach [124], CBT Assistant [125], Challenger [126], Lantern [127], PsychAssist [128], Public Speech Trainer (PST) [129], SmartCAT [130], *GET.ON PAPP* [131,132], N/A [133-139]
Various disorders	ACT Daily APP [140], FOCUS [141], Headspace [142], iBobbly [143], iCare-stress [144], IntelliCare Suite [145], MoodMission [146], MoodTrainer [147], myCompass [148], PeerTECH App [149], Pocket Skills [150], Sinasprite [151], SmartCAT [152], SPIRIT [153], The Moment [154], TODAY! [155], Wellframe [156], WellWave [157], N/A [149], *WorkingWell* [158,159]
Depressive disorders	7Cups [160], Be Good to Yourself [18], BlueWatch [161], Dcombat [162], Get Happy Program [163], HeadGear [164], iCare Prevent [165], MedLink [166], Mobile Sensing and Support [167], MoodHacker [168], Moodivate [169], MyGamePlan [170], PRIME-D [171], Push-D [172], SocioEmpathy [173], SPSRS [174], SuperBetter [175], The Sound Advice [176], Thought Challenger [177], TODAC [178], *Kokoro-App* [179-181], N/A [19,80,182-187]

[a]N/A: app name not available/not mentioned.

Overview of Randomized Controlled Trial Assessments

In Multimedia Appendix 3, we provide a comprehensive table with key variables of (pilot) RCT studies, including their main characteristics and results. We found a total of 32 rigorous assessment designs, with 7 pilot RCTs (22%) and 25 RCTs (78%). Below, we discuss the psychotherapy approaches (ie, treatment modalities) found in these assessments, temporal evolution of assessment types in RCTs, and notable observations on covered mental disorders.

Regarding approaches to psychotherapy, we found that 80.4% (127/158) of all included studies follow a cognitive behavioral therapy (CBT) psychological framework; 21.3% (27/127) of these additionally include third wave therapy techniques such as mindfulness or acceptance, or commitment therapy techniques. There are 13 apps fully based on a third wave therapy. Two apps are based on behavioral activation and physical activity. For the remaining mobile apps, no information regarding their psychological framework was reported in the article. When specifically considering the pilot RCT and the RCT studies, we observed the same trend: the majority is based on the CBT psychological framework, followed by studies which combine CBT with third wave therapy techniques. When looking at the overall temporal evolution of RCT studies, we noticed an even spread of combined effect and usability/UX evaluations over time. For effect studies, however, we observed that the majority took place in the last 2 years.

The most commonly addressed disorders by RCT studies were *depressive disorders* (9/25, 36% of the RCT studies). The majority of these studies showed that participants who received intervention apps significantly improved their symptoms (depression, anxiety, etc) compared with the waiting list [18,175], alternative care [168], or control conditions [178,180]. Furthermore, studies with follow-ups showed that the treatment benefits were maintained [168,178]. One of the RCTs compared an intervention app with treatment as usual (ie, therapist); the results showed that, at posttreatment, the clinical variables did not differ between groups [186]. Another study compared two

different apps, each featuring a different approach to psychotherapy (behavioral activation and mindfulness); the results showed that both apps were useful and did not differ significantly from one another [157]. The second most addressed disorder involved *anxiety disorders* (4/25, 16% of the RCT studies). Here, the results showed the same trend as for depressive disorders; participants who received intervention apps improved their symptoms significantly compared with the waiting list [123,126,139]. Furthermore, in studies with follow-ups, treatment benefits were maintained [126,139]. One of the studies compared two different intervention apps with different intervention targets, that is, agoraphobia vs general anxiety symptoms, for managing agoraphobia [122]. The results showed reductions in symptom severity over time that was statistically significant without differences between both apps.

For *schizophrenia spectrum and other psychotic disorders* (3/25, 12% of the RCT studies), we found that 2 RCTs used the same app [104,105]. Results showed significant improvements in clinical symptoms posttrial compared with the waiting list, and also good acceptability [104,105]. A third app also showed positive results at posttreatment [100]. For *substance-related and addictive disorders*, we found one RCT; the study found improvements in alcohol outcomes for the participants [106]. Several other disorders were also only covered by one RCT. In case of *sleep-wake disorders*, the intervention app produced significant improvements in insomnia severity and sleep efficiency compared with the waiting list [38]. For *suicidal behavior disorder/nonsuicidal self-injury disorders*, the intervention app reported a significantly improved ability to cope with unpleasant emotions and thoughts compared with the control group [71]. In case of *bipolar and related disorders*, participants in the intervention app group showed significantly greater reductions in depressive symptoms [55]. For *trauma and stressor-related disorders*, RCT studies showed slightly less promising results. Usage of an intervention app did not result in significantly better outcomes compared with other active control conditions [79]; still, outcomes were better when compared with the waiting list condition [85].

Discussion

Principal Findings

In general, we infer a growing interest in utilizing smartphone apps for delivering psychological treatments, with research increasing from only a few (7) articles in 2013 to an order of magnitude more (60) in 2018. This is a promising trend, as these apps can complement therapist-led psychological treatments and, hence, increase their efficacy and availability. When delegating (part of) psychological treatment to smartphone-based interventions, the need for face-to-face sessions and manual follow-up is decreased, which, in turn, lowers costs and reduces waiting lists in the public health system. According to the mental health workforce breakdown (by the World Health Organization region), there are only 4.6 psychologists per 100,000 inhabitants in Europe [188]. For Spain, studies have reported wait times of more than 45 days before the first psychological assistance by a clinical psychologist or psychiatrist [188], and a frequency of face-to-face sessions of around once a month [189]. Moreover, leveraging smartphones' capabilities enables ecological momentary interventions (EMI), whereby patients are able to access psychological care when and where they need it most, in their natural environment and daily routines [190-192]. Below, we discuss our concrete observations on the assessment types of included studies, coverage of mental disorders, and technical features.

Evolution of Research and Types of Assessment

It is a promising sign that overall, the number of articles with some sort of assessment is slowly increasing. Furthermore, we observe that proportionally, there is a much higher number of studies with an evaluation of only usability/UX, compared with only effect. Usability factors have been widely recognized as key factors to enhance the acceptance of information and communication technologies (ICT) tools; on the basis of the technology acceptance model, authors have suggested that the intention to use a product in the future is strongly correlated with its ease of use [193,194]. Hence, initial efforts to research and ensure the usability of new ICT tools are essential. At the same time, we observe an overall much lower number (less than half compared with usability/UX) of studies that explicitly assesses the effect of smartphone interventions on clinical symptomatology (despite a peak in 2017). Yet, it is specifically this type of studies, focusing on the (long-term) clinical effects of the intervention, that are needed to demonstrate efficacy, and increase therapists' and patients' trust in smartphone-based interventions.

Moreover, RCTs, which are considered the gold standard of experiment design in mental health (and medicine in general), are only minimally represented in the literature (22/158, 15.8% of articles overall). Among them, we see an even spread of combined effect and usability/UX evaluations over time; for studies specifically focusing on effect, however, the majority took place in the last 2 years (with a peak in 2017). This is a promising sign, although there are still relatively few effect studies (see Multimedia Appendix 3). Furthermore, the most commonly addressed disorders using the RCT methodology are depressive disorders, followed by anxiety disorders. It is, thus, important to carry out more RCTs to prove mental health apps' efficacy in treating other mental disorders, and to study the satisfaction and experience of the patients using these apps. Moreover, to draw rigorous and trustworthy conclusions on the clinical efficacy of smartphone apps, more long-term RCT studies will be needed (eg, to better measure the effects of attrition). Similarly, we observe a distinct lack of cross-validation studies, with only a few apps having been studied in multiple articles (19/138, 13.8%). One could note that this phenomenon is correlated with the lack of rigorous long-term studies on smartphone interventions—a single, multiyear study would warrant multiple articles for a single smartphone intervention on study protocol, usability evaluation, and effect studies at multiple intervals. Particularly when utilizing novel technological features, rigorous assessment studies are needed to validate their potential for psychological interventions and encourage further research in the field. A stronger cooperation between research groups could increase the resources needed for such long-term psychological intervention studies.

Covered Mental Disorders

To an extent, the coverage of mental disorders in the relevant literature seems to be in line with their real-world prevalence. This holds, in particular, for depressive and anxiety disorders, commonly called emotional disorders [195]; they (1) represent the first and third most covered disorders in the literature (we point out that *various disorders* include apps addressing multiples disorders, where depressive and anxiety disorders, in addition to serious mental illness, are most represented), with the research on *depressive disorders* being doubled in 2018; and (2) they are known to affect the most people worldwide. For mental disorders with highest prevalence among people [196,197], lifetime prevalence has been estimated at 28.8% for anxiety disorders, 20.8% for mood disorders (including 16.6% for depressive disorders, which are a mood disorder), 24.8% for impulse-control disorders, and 14.6% for substance use disorders. Estimated 12-month prevalence follows a similar trend: anxiety disorders are the most prevalent class with 18.1%, followed by mood disorders with 9.5% (including 6.7% for depressive disorders), impulse-control disorders (8.9%), and substance disorders (3.8%) [196,197]. Hence, according to the psychological literature [198,199], the three most prevalent mental disorders include anxiety, mood (including depressive disorders), and substance disorders. Indeed, these similarly make up our top four of most covered mental disorders in smartphone intervention studies. Depressive and anxiety disorders reduce a patient's psychosocial functioning and quality of life [198,200], and are associated with important personal, social, and economic repercussions [199,201]. Other ICT technologies for delivering psychological treatments, such as internet and Web-based programs, are also mostly focused on depressive and anxiety disorders [202,203]; this might also have had an influence on the proliferation of smartphone-based interventions. Although determining the underlying factor(s) behind the distribution of addressed mental disorders in the literature is certainly an interesting exercise, we consider this beyond the scope of this paper.

Beyond depressive and anxiety disorders, the literature is heavily focused on only a small number of disorders; six mental disorders account for approximately 73.4% (116/158) of research. On the other hand, more than half of the categories of mental disorders listed in DSM-5 (15) are fully excluded or very scarcely represented (<3%). Clearly, there is an opportunity, as well as an acute need, to pay more attention to the whole breadth of mental disorders—that is, including those that are less prevalent—to help as many people as possible. Some of these less prevalent disorders, such as personality disorders, often have a higher severity that may lead to extreme consequences. For instance, borderline personality disorder affects only 2% to 6% of the population [204,205], but its mortality rate by suicide is one of the highest in the world among people with psychiatric disorders [206].

Coverage of Technical Dimensions

When looking at technical dimensions, more traditional software features (see 7×12 vertical rectangle in Figure 5) are much more utilized than the novel sensing or analytical capabilities of smartphones. One may argue that these top seven features, which involve intervention-specific features (eg, learning) and communication features (eg, prompting), do not offer a significant advancement over the prior state of the art. Indeed, many previous studies that leveraged (nonsmartphone) mobile phones supported learning by displaying psychoeducational content [56,58,140], receiving tips/reminders via SMS [47,149], using (bidirectional) SMS communication to perform (in situ) assessment [149], or telephone calls to health care providers [75,81,145]. Notwithstanding these observations, even this rather conservative transition to smartphones has enabled interventions that are out of reach for classic mobile phones. Research leveraging smartphones have exploited larger screen resolutions and multimedia capabilities to provide multimodal learning materials, using audio and video guides [122,184], pictures [71,75-77,87,154], audio [71,76,87], music [75-77], and video [71,77,174]. Some authors have leveraged the improved connectivity and ubiquity of smartphones to offer access to entire Web-based libraries of learning materials [152,184]; others utilize in-app prompting as intervention techniques, for example, sending reminders to use the app [76,85,116,178], motivational messages [47], or messages from the therapist [80,103]. We found studies that exploit the improved interactivity of smartphones to provide interactive quizzes for training skills and improved learning [62,152], assessments for panic attacks [133], suicidal intentions [143], symptoms of various disorders [54,120,121], and communication with therapists [37,58,92,133] or other users [79,104] through message/chat. Furthermore, aside from being better supported by smartphone capabilities, many of these psychological smartphone interventions are available at the touch of a button, instead of relying on receipt of SMS or phone calls.

That said, most studies still only scratch the surface of advanced smartphone capabilities. This is particularly apparent in the relatively low coverage of context awareness, that is, leveraging sensors to detect and react upon the current state of patients and their environments. We argue that such context awareness is a key ingredient of true EMI. Indeed, although EMI are meant to proactively issue suitable therapeutic interventions at the right time and place, most EMI studies consider smartphones merely as a tool for manually accessing interventions, or receiving predefined interventions at set time intervals, at any moment and place. We found a very limited number of smartphone-based studies leveraging external sensors for recording physiological parameters: measuring heart rate for detecting physiological arousal in the context of anger management [78], for instance. Similarly, we found very low coverage of analytics-based studies that could support advanced context awareness, for example, learning and assessing mental states based on physiological, environment, activity, and/or behavioral contexts. In our opinion, studies that progressively use internal and/or external smartphone sensors, possibly combined with advanced analytics, are a useful step toward realizing the full potential of EMI, where relevant events are detected through analysis of sensor readings, and acted upon by triggering suitable, personalized interventions when they are needed. The general hesitance toward context-aware EMI could be explained by the lack of validated computerized psychological models for assessing mental states based on patient context (including physiological, environment, and activity factors), as well as the need to

combine technologically advanced solutions (ie, use of sensors, context awareness, and analytics).

Through their improved modalities (screen and sound), interactivity, and computational resources, smartphones enable novel intervention delivery paradigms, including virtual and augmented realities. However, these have found very limited coverage in the literature. We found a few individual studies, for example, utilizing a mobile VR system to help patients coping with agoraphobia by guiding an avatar through real-life simulations in a game-based setting [122]. Beyond exposure-type therapies, Repetto et al [137] utilized VR techniques to cope with generalized anxiety disorders, leveraging biofeedback to regulate features of the virtual world (eg, current heart rate). As mentioned before, the seeming lack of nontraditional intervention methods may be because of the lack of validated psychological models for supplying evidence-based VR or augmented reality, and/or the technical difficulty of novel delivery paradigms.

Given the lack of studies on these topics, we believe that there lies a huge potential for future research in utilizing technologically advanced solutions (ie, sensors, context awareness, and alternative delivery paradigms) to deliver smartphone-based psychological interventions tailored to the patient's current health context.

Barriers to Implementation and Patient Risks

Despite our advocation for further research in the utilization of smartphone features to advance treatments for mental disorders, we note that technological innovation should not constitute a goal in itself. It must provide a distinct advantage toward patient care, such as improved mental health care access, that is, a broader reach and lower barriers; increased assessment frequency and accuracy; lower cost; improved efficacy; and immediate access to care, when and where the patient needs it most. These intended benefits must be balanced with possible adverse effects, and the use of treatment modalities, including advanced technical features, needs to be carefully contemplated.

Here, we point out possible risks and barriers to implementation of smartphone-based interventions. The lack of research evidence on the *effectiveness of mobile health apps* is likely the most important issue to consider [207,208]. As shown in Table 1, only a limited number of smartphone apps are validated using an RCT, and we observed a distinct lack of cross-validation studies, with only a few apps having been assessed in multiple studies. We further point out that this analysis only covers apps presented in the literature and not the many thousands of other, nonvalidated apps in popular app stores. There is a clear risk in using nonvalidated apps, as there is no evidence that they have any therapeutic effect, and they may even worsen the patient's condition. For example, Baron et al [209] noted that low-accuracy sleep trackers to self-diagnose sleep disturbances caused patients to be overly concerned on getting the perfect sleep (*orthosomnia*), which may have exacerbated their insomnia.

The ability to protect the *privacy and confidentiality* of patient information is pivotal as well [207,210-212] and may form a barrier to adoption. Many health-related apps collect a large amount of demographic, medical, and lifestyle information, as well as data on users' daily routines and practices [213]. Some apps have even been found to collect user data unrelated to the app's purpose [211]. Despite the importance of privacy with respect to (mental) health data, several researchers have found that a significant portion (31%-49%) of studied mental health apps does not include a privacy policy [210,214]. Collected information may be distributed to (third-party) services for storage and analysis. Moreover, unregulated apps may even pass on this information to unidentified parties, for example, for advertising purposes [210,211]. In a study of depression and smoking cessation mobile apps, it was observed that the majority (81%) of the studied apps transmitted data for advertising and marketing purposes, but only 59% of the apps disclosed this in their privacy policy [210]. In all the aforementioned cases, there is a risk that personal data are being transmitted over insecure network connections [215,216]. Furthermore, there is the possibility of loss, theft, or malfunction of the mobile device [5], or cybercriminals specifically targeting health information [207], all of which could result in the loss of sensitive data.

Toward addressing these two issues, some authors suggest that mental health professionals should screen the apps they recommend for privacy issues [217]; however, this does not seem like a feasible or robust solution. Olff [218] proposed a disclaimer for apps that have not been validated. Taking it one step further, certification processes for mobile (mental) health apps have recently been instigated in the European Union (EU) and the United States, even though they do not apply to all apps. Under the 2017 EU Medical Device Regulation [219], mobile apps with a medical intended purpose require a CE marking. Devices (apps) may be differently classified depending on the purpose and risk they pose, yielding more or less stringent regulations regarding quality, control, or development process. The EU General Data Protection Regulation [220] has furthermore motivated a number of app developers to improve transparency on privacy policy, although confusion remains on its applicability outside the EU [211]. Similarly, in the United States, the Food and Drug Administration policy for mobile medical apps [221] stipulates that any software that is utilized for the diagnosis of disease or other conditions, or the cure, mitigation, treatment, or prevention of disease, constitutes a medical device, regardless of the platform on which it is run (eg, desktop or mobile). The applicable regulations depend on the particular functionality. Given the recency of these (updated) regulations, their rate of adoption and effects on efficacy, safety, privacy, and usability of mHealth apps remain to be seen, and we expect the regulations to be further clarified, refined, and extended in scope in the future.

Further barriers to implementation include a lack of sufficient general, digital, and health literacy levels [207] and a digital divide [222], the view and attitude of practitioners and patients toward the use of mobile mental health interventions [223,224], availability and awareness of evidence-based apps [225], economic and other associated costs [226], and user acceptance and usability [226]. For further elaboration on barriers and facilitators to implementation of mental health care apps, we refer the readers to Lipschitz et al [225] and Simblett et al [226].

Recommendations for the Research Community

In line with our findings, we propose four recommendations for the research community to further develop and advance the field of smartphone-based psychological interventions:

- Attention for less covered disorders: The majority of research (approximately 73%) pertains to the top six covered mental disorders, four of which coincide with highly prevalent mental disorders. On the other hand, more than half of the DSM-5 recognized mental disorders are not or very scarcely covered. We call upon the research community to invest into covering the full breadth of mental disorders.
- Attention for advanced technical and software-based solutions: Many smartphone-based psychological interventions merely translate traditional and electronic health (eHealth) solutions to smartphones; that is, these interventions do not fully exploit their capabilities. Concretely, the use of sensors and corresponding context awareness, particularly to promote EMI, the exploration of alternative delivery paradigms such as virtual or augmented reality, and more advanced analytical methods are scarcely investigated. We call upon the research community to explore beyond traditional strategies, toward leveraging advanced technological features to improve mHealth interventions.
- Multidisciplinary approaches: To fully exploit the smartphone's capability as a pervasive, ubiquitously connected, sensor-packed computing platform to deliver innovative, real-time, and in situ psychological interventions, both the domain knowledge of psychologists and the technical expertise of computer scientists are needed. Hence, we call for multidisciplinary collaborations as to not let technical difficulties, or lack of psychological knowledge on mental disorders, hinder advances and novelties in the field.
- Validation toward effect: Although we uncovered, at least in absolute numbers, a slight increase in effect validations and RCT-based effect assessments during the last 2 years combined, they are still underrepresented (particularly RCTs). Hence, there exists a need to rigorously validate smartphone-based psychological treatments for effect. Especially when utilizing advanced technical features (eg, context awareness, analytics, and alternative delivery paradigms), effect validation may increase trust and spark further research in such novel types of interventions. We call upon the research community to augment efforts in rigorous effect assessment, to allow transfer of research into practice.

Despite our call for research in technical innovation and its broader applicability and validation, we note that the eventual use of such advanced technical features—or any technological aid in psychological interventions for that matter—needs to be carefully balanced with the characteristics and needs of the individual patient.

Strengths and Limitations

The main strength and novelty of this study is that it explored and summarized, considering a wide range of technical characteristics, the current state of the art in smartphone-based interventions for mental disorders. We hereby provide a broad overview of the field (1) covering the full spectrum of mental disorders as classified in the latest version of DSM, rather than focusing on a specific mental illness as done in previous studies; and (2) exposing technical features used to realize smartphone-based treatments. Consequently, this contribution is highly innovative as a synergetic study targeting mental health research and recent developments in mobile sensing and computing. Further strengths of this study include the use of four different bibliographic sources for a comprehensive coverage of the research and literature, and the methodological process based on pairs of multidisciplinary researchers for the selection, validation, and classification of the literature.

As any systematic study, search term specification may lack other relevant terms not considered by the authors, and searches only covered the literature published in English. Therefore, there always exists a risk to not fully identify all relevant studies. Classification of studies may also be prone to error. To reduce this risk, we used pairs of researchers from different disciplines with a requirement of interrater agreement.

Conclusions

We presented a comprehensive systematic review of the state of the art in smartphone-based psychological interventions, with a synergetic focus on psychology-related issues, such as mental disorders and type of assessment, as well as technological features, such as software features and device sensors. Our results show a rapid increase over recent years in the number of psychological interventions for various mental disorders using smartphone-based apps. It captures how depressive and anxiety disorders are primarily covered, in line with their real-world prevalence. The top six of mental disorders together account for approximately three-quarters of coverage in the literature, while over half are not or very scarcely covered. This implies the need for further research on smartphone interventions for the full breadth of mental disorders to help as many affected people as possible. On the technical side, the review highlights a group of software features related to intervention (eg, learning and in situ use) and communication (eg, prompting) deployed in smartphone interventions that mostly mimic more traditional mobile phone and eHealth solutions. More innovative use of smartphones' capabilities, such as sensing, alternative delivery paradigms, and advanced analytics, are only scarcely present in the literature, despite their potential for advancing solutions such as EMI. With regard to studies including an assessment, we found that there is an overall slow proportional increase, with significantly more usability/UX compared with effect studies. RCT studies are still a small minority. They mostly deal with depressive and anxiety disorders. Over the last 2 years, there are promising yet inconclusive signs of more effect studies.

Acknowledgments

IM was funded by the Spanish Ministry of Education, Culture and Sports (grant reference FPU14/00020). CG and SC were funded by the Ramon y Cajal Programme of the Spanish government (grant references RYC-2014-16913 and RYC-2014-16606 respectively). LD was funded by the Conselleria de Educación, Investigación, Cultura y Deporte ORDRE 86/2016, de 21 de desembre [2016/10396] (grant reference ACIF/2017/181). This study was supported by the *SYMPTOMS* project of the *Programa Estatal de I+D+i Orientada a los Retos de la Sociedad* of the Spanish government (grant number RTI2018-099939-B-I00) and by the UJI-A2016-14 Program, Project 16I336.01/1 (Universitat Jaume I).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords.
[DOCX File , 14 KB - mhealth_v8i4e14897_app1.docx]

Multimedia Appendix 2
Search queries.
[DOCX File , 15 KB - mhealth_v8i4e14897_app2.docx]

Multimedia Appendix 3
Key variables for randomized controlled trials or pilot randomized controlled trial studies.
[DOCX File , 41 KB - mhealth_v8i4e14897_app3.docx]

References

1. Statista. 2019. Number of Smartphone Users Worldwide From 2016 to 2021 (in billions) URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2019-09-01]
2. Taylor K, Silver L. Pew Research Center. 2019. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewglobal.org/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf [accessed 2019-09-01]
3. Clement J. Statista. Mobile App Usage - Statistics & Facts URL: https://www.statista.com/topics/1002/mobile-app-usage/ [accessed 2018-03-09]
4. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
5. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol 2011;42(6):505-512. [doi: 10.1037/a0024485]
6. Koch S. Home telehealth--current state and future trends. Int J Med Inform 2006 Aug;75(8):565-576. [doi: 10.1016/j.ijmedinf.2005.09.002] [Medline: 16298545]
7. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
8. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
9. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
10. Huxley CJ, Atherton H, Watkins JA, Griffiths F. Digital communication between clinician and patient and the impact on marginalised groups: a realist review in general practice. Br J Gen Pract 2015 Dec;65(641):e813-e821 [FREE Full text] [doi: 10.3399/bjgp15X687853] [Medline: 26622034]
11. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
12. Pepin R, Segal DL, Coolidge FL. Intrinsic and extrinsic barriers to mental health care among community-dwelling younger and older adults. Aging Ment Health 2009 Sep;13(5):769-777. [doi: 10.1080/13607860902918231] [Medline: 19882416]
13. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999-2010. Health Aff (Millwood) 2013 Oct;32(10):1723-1730 [FREE Full text] [doi: 10.1377/hlthaff.2013.0133] [Medline: 24101061]
14. Twomey C, O'Reilly G, Meyer B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: A meta-analysis. Psychiatry Res 2017 Oct;256:371-377. [doi: 10.1016/j.psychres.2017.06.081] [Medline: 28686935]

15. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016 Sep;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
16. Bagalman E, Cornell AS. Federation Of American Scientists. 2018 Jan 19. Prevalence of Mental Illness in the United States: Data Sources and Estimates URL: https://fas.org/sgp/crs/misc/R43047.pdf [accessed 2020-02-05]
17. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
18. Lüdtke T, Pult L, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
19. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
20. Proudfoot J, Parker G, Pavlovic DH, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
21. Andersson E, Steneby S, Karlsson K, Ljótsson B, Hedman E, Enander J, et al. Long-term efficacy of internet-based cognitive behavior therapy for obsessive-compulsive disorder with or without booster: a randomized controlled trial. Psychol Med 2014 Oct;44(13):2877-2887. [doi: 10.1017/S0033291714000543] [Medline: 25066102]
22. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
23. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
24. Ehrenreich B, Righter B, Rocke DA, Dixon L, Himelhoch S. Are mobile phones and handheld computers being used to enhance delivery of psychiatric treatment? A systematic review. J Nerv Ment Dis 2011 Nov;199(11):886-891. [doi: 10.1097/NMD.0b013e3182349e90] [Medline: 22048142]
25. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, 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]
26. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
27. Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, et al. Categorizing health outcomes and efficacy of mHealth apps for persons with cognitive impairment: a systematic review. J Med Internet Res 2017 Aug 30;19(8):e301 [FREE Full text] [doi: 10.2196/jmir.7814] [Medline: 28855146]
28. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
29. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
30. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 6;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
31. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
33. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Washington, DC: American Psychiatric Publishing; 2013.
34. International Organization for Standardization. 2010. ISO 9241-210:2010 Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems URL: https://www.iso.org/standard/52075.html [accessed 2020-02-05]
35. Granell C, Bretón-López J, Casteleyn S, Castilla D, Díaz L, Mira A, et al. Zenodo. 2020 Jan 15. Reproducibility Package for 'Smartphone apps for the treatment of mental disorders: a systematic review' URL: https://doi.org/10.5281/zenodo.3609267 [accessed 2020-02-05]

36. Al-Asad RY, Al-Zuhair AJ, Esmacili SE. Smart Phone Based Facial and Text Recognition Application (RICO). In: Proceedings of the Iranian Conference on Electrical Engineering. 2018 Presented at: ICEE'18; May 8-10, 2018; Mashhad, Iran p. 125-129. [doi: 10.1109/ICEE.2018.8472658]
37. Prada P, Zamberg I, Bouillault G, Jimenez N, Zimmermann J, Hasler R, et al. EMOTEO: a smartphone application for monitoring and reducing aversive tension in borderline personality disorder patients, a pilot study. Perspect Psychiatr Care 2017 Oct;53(4):289-298. [doi: 10.1111/ppc.12178] [Medline: 27439663]
38. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
39. Espie CA, Luik AI, Cape J, Drake CL, Siriwardena AN, Ong JC, et al. Digital Cognitive Behavioural Therapy for Insomnia versus sleep hygiene education: the impact of improved sleep on functional health, quality of life and psychological well-being. Study protocol for a randomised controlled trial. Trials 2016 May 23;17(1):257 [FREE Full text] [doi: 10.1186/s13063-016-1364-7] [Medline: 27216112]
40. Kyle S, Hurry ME, Emsley R, Luik AI, Omlin X, Spiegelhalder K, et al. Effects of digital Cognitive Behavioural Therapy for Insomnia on cognitive function: study protocol for a randomised controlled trial. Trials 2017 Jun 17;18(1):281 [FREE Full text] [doi: 10.1186/s13063-017-2012-6] [Medline: 28623947]
41. Kolar D, Hammerle F, Jenetzky E, Huss M. Smartphone-Enhanced Low-Threshold Intervention for adolescents with Anorexia Nervosa (SELTIAN) waiting for outpatient psychotherapy: study protocol of a randomised controlled trial. BMJ Open 2017 Oct 22;7(10):e018049 [FREE Full text] [doi: 10.1136/bmjopen-2017-018049] [Medline: 29061627]
42. Nitsch M, Dimopoulos CN, Flaschberger E, Saffran K, Kruger JF, Garlock L, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972] [Medline: 26753539]
43. Tregarthen JP, Lock J, Darcy AM. Development of a smartphone application for eating disorder self-monitoring. Int J Eat Disord 2015 Nov;48(7):972-982. [doi: 10.1002/eat.22386] [Medline: 26213130]
44. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR Mhealth Uhealth 2018 Jun 22;6(6):e10253 [FREE Full text] [doi: 10.2196/10253] [Medline: 29934285]
45. Zaffke A, Jain N, Johnson N, Ul Alam MA, Magiera M, Ahamed SI. iCanLearn: A Mobile Application for Creating Flashcards and Social Stories for Children with Autism. In: Proceedings of the International Conference on Smart Homes and Health Telematics. 2014 Presented at: ICOST'14; June 25–27, 2014; Denver, CO, USA p. 225-230 URL: https://doi.org/10.1007/978-3-319-14424-5_25 [doi: 10.1007/978-3-319-14424-5_25]
46. Skillen KL, Donnelly MP, Nugent CD, Booth N. LifePal: A Mobile Self-management Tool for Supporting Young People with Autism. In: Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing. 2016 Presented at: MEDICON'16; March 31-April 2, 2016; Paphos, Cyprus p. 1174-1179 URL: https://doi.org/10.1007/978-3-319-32703-7_227 [doi: 10.1007/978-3-319-32703-7_228]
47. Chacko A, Isham A, Cleek AF, McKay MM. Using mobile health technology to improve behavioral skill implementation through homework in evidence-based parenting intervention for disruptive behavior disorders in youth: study protocol for intervention development and evaluation. Pilot Feasibility Stud 2016;2:57 [FREE Full text] [doi: 10.1186/s40814-016-0097-4] [Medline: 27965873]
48. Flobak E, Jensen DA, Lundervold AJ, Nordgreen T, Chen LH, Guribye F. Towards Technology-Based Interventions for Improving Emotional and Cognitive Control. In: Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA'18; April 21–26, 2018; Montréal, QC, Canada. [doi: 10.1145/3170427.3188603]
49. Olbrich H, Stengler K, Olbrich S. Smartphone based Geo-Feedback in obsessive compulsive disorder as facilitatory intervention: a case report. J Obsessive Compuls Relat Disord 2016;8:75-78 [FREE Full text] [doi: 10.1016/j.jocrd.2016.01.001]
50. Pascual-Vera B, Roncero M, Doron G, Belloch A. Assisting relapse prevention in OCD using a novel mobile app-based intervention: a case report. Bull Menninger Clin 2018;82(4):390-406. [doi: 10.1521/bumc.2018.82.4.390] [Medline: 30589573]
51. Boisseau CL, Schwartzman CM, Lawton J, Mancebo MC. App-guided exposure and response prevention for obsessive compulsive disorder: an open pilot trial. Cogn Behav Ther 2017 Nov;46(6):447-458. [doi: 10.1080/16506073.2017.1321683] [Medline: 28565937]
52. Whiteside SP, Ale CM, Douglas KV, Tiede MS, Dammann JE. Case examples of enhancing pediatric OCD treatment with a smartphone application. Clin Case Stud 2014;13(1):80-94 [FREE Full text] [doi: 10.1177/1534650113504822]
53. Hong JS, Kim SM, Aboujaoude E, Han DH. Investigation of a mobile 'Serious Game' in the treatment of obsessive-compulsive disorder: a pilot study. Games Health J 2018 Oct;7(5):317-326. [doi: 10.1089/g4h.2017.0158] [Medline: 30129775]
54. Wenze SJ, Armey MF, Weinstock LM, Gaudiano BA, Miller IW. An open trial of a smartphone-assisted, adjunctive intervention to improve treatment adherence in bipolar disorder. J Psychiatr Pract 2016 Nov;22(6):492-504 [FREE Full text] [doi: 10.1097/PRA.0000000000000196] [Medline: 27824786]

55. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord 2015 Mar 15;174:23-30 [FREE Full text] [doi: 10.1016/j.jad.2014.10.053] [Medline: 25479050]
56. Hidalgo-Mazzei D, Reinares M, Mateu A, Juruena MF, Young AH, Pérez-Sola V, et al. Is a SIMPLe smartphone application capable of improving biological rhythms in bipolar disorder? J Affect Disord 2017 Dec 1;223:10-16. [doi: 10.1016/j.jad.2017.07.028] [Medline: 28711743]
57. Hidalgo-Mazzei D, Vieta E, Colom F. Self-Monitoring and Psychoeducation in Bipolar Patients with a Smart-phone application (SIMPLe) project: Preliminary results from a feasibility study. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. 2015 Presented at: MOBIHEALTH'15; October 14–16, 2015; London p. 227-230 URL: https://doi.org/10.4108/eai.14-10-2015.2261770 [doi: 10.4108/eai.14-10-2015.2261770]
58. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]
59. Hidalgo-Mazzei D, Reinares M, Mateu A, Nikolova VL, Del Mar Bonnín C, Samalin L, et al. OpenSIMPLe: a real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. J Affect Disord 2018 Dec 1;241:436-445. [doi: 10.1016/j.jad.2018.08.048] [Medline: 30145515]
60. Muroff J, Robinson W, Chassler D, López LM, Gaitan E, Lundgren L, et al. Use of a smartphone recovery tool for Latinos with co-occurring alcohol and other drug disorders and mental disorders. J Dual Diagn 2017;13(4):280-290. [doi: 10.1080/15504263.2017.1348649] [Medline: 28692420]
61. Tajima-Pozo K, Ruiz-Manrique G, Montañes F. Treating enuresis in a patient with ADHD: application of a novel behavioural modification therapy. BMJ Case Rep 2014 Jun 10;2014:pii: bcr2014203912 [FREE Full text] [doi: 10.1136/bcr-2014-203912] [Medline: 24916977]
62. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
63. Minami H, Brinkman HR, Nahvi S, Arnsten JH, Rivera-Mindt M, Wetter DW, et al. Rationale, design and pilot feasibility results of a smartphone-assisted, mindfulness-based intervention for smokers with mood disorders: Project mSMART MIND. Contemp Clin Trials 2018 Mar;66:36-44 [FREE Full text] [doi: 10.1016/j.cct.2017.12.014] [Medline: 29288740]
64. Herbst E, Pennington D, Kuhn E, McCaslin SE, Delucchi K, Batki SL, et al. Mobile technology for treatment augmentation in veteran smokers with posttraumatic stress disorder. Am J Prev Med 2018 Jan;54(1):124-128. [doi: 10.1016/j.amepre.2017.08.016] [Medline: 29074319]
65. Benarous X, Edel Y, Consoli A, Brunelle J, Etter J, Cohen D, et al. Ecological momentary assessment and smartphone application intervention in adolescents with substance use and comorbid severe psychiatric disorders: study protocol. Front Psychiatry 2016;7:157 [FREE Full text] [doi: 10.3389/fpsyt.2016.00157] [Medline: 27703435]
66. Matalenas LA, McLaughlin AC, Chen Y, Daughters SB. Developing a smartphone application for the life enhancement treatment for substance use (lets act): designing for motivation and feedback. Proc Hum Factors Ergon Soc Annu Meet 2015;59(1):1100-1104 [FREE Full text] [doi: 10.1177/1541931215591158]
67. Nuij C, van Ballegooijen W, Ruwaard J, de Beurs D, Mokkenstorm J, van Duijn E, et al. Smartphone-based safety planning and self-monitoring for suicidal patients: rationale and study protocol of the CASPAR (Continuous Assessment for Suicide Prevention And Research) study. Internet Interv 2018 Sep;13:16-23 [FREE Full text] [doi: 10.1016/j.invent.2018.04.005] [Medline: 30206514]
68. Melvin GA, Gresham D, Beaton S, Coles J, Tonge BJ, Gordon MS, et al. Evaluating the feasibility and effectiveness of an Australian safety planning smartphone application: a pilot study within a tertiary mental health service. Suicide Life Threat Behav 2019 Jun;49(3):846-858 [FREE Full text] [doi: 10.1111/sltb.12490] [Medline: 29999193]
69. Kennard BD, Goldstein T, Foxwell AA, McMakin DL, Wolfe K, Biernesser C, et al. As Safe as Possible (ASAP): a brief app-supported inpatient intervention to prevent postdischarge suicidal behavior in hospitalized, suicidal adolescents. Am J Psychiatry 2018 Sep 1;175(9):864-872 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17101151] [Medline: 30021457]
70. Tighe J, Shand F, Ridani R, Mackinnon A, de La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
71. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A Virtual Hope Box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 1;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
72. Chen JI, Smolenski DJ, Dobscha SK, Bush NE, Denneson LM. Correlates of mental health smartphone application use among patients with suicidal ideation. J Technol Hum Serv 2018;36(4):191-207. [doi: 10.1080/15228835.2018.1502119]
73. 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 Nov 20;14:396 [FREE Full text] [doi: 10.1186/1745-6215-14-396] [Medline: 24257410]

74. O'Brien KH, 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]
75. Stallard P, Porter J, Grist R. Safety, acceptability, and use of a smartphone app, BlueIce, for young people who self-harm: protocol for an open Phase I trial. JMIR Res Protoc 2016 Nov 16;5(4):e217 [FREE Full text] [doi: 10.2196/resprot.6525] [Medline: 27852560]
76. Stallard P, Porter J, Grist R. A Smartphone App (BlueIce) for Young People Who Self-Harm: Open Phase 1 Pre-Post Trial. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e32 [FREE Full text] [doi: 10.2196/mhealth.8917] [Medline: 29382632]
77. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018 Feb 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
78. Morland LA, Niehaus J, Taft C, Marx BP, Menez U, Mackintosh MA. Using a mobile application in the management of anger problems among veterans: a pilot study. Mil Med 2016 Sep;181(9):990-995. [doi: 10.7205/MILMED-D-15-00293] [Medline: 27612342]
79. Mackintosh MA, Niehaus J, Taft CT, Marx BP, Grubbs K, Morland LA. Using a mobile application in the treatment of dysregulated anger among veterans. Mil Med 2017 Nov;182(11):e1941-e1949. [doi: 10.7205/MILMED-D-17-00063] [Medline: 29087863]
80. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
81. 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]
82. Kuhn E, Crowley JJ, Hoffman JE, 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;46(6):437-443 [FREE Full text] [doi: 10.1037/pro0000051]
83. 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 [FREE Full text] [doi: 10.1037/mil0000083]
84. Reger G, Browne KC, Campellone TR, Simons C, Kuhn E, Fortney JC, et al. Barriers and facilitators to mobile application use during PTSD treatment: Clinician adoption of PE coach. Prof Psychol Res Pr 2017;48(6):510-517 [FREE Full text] [doi: 10.1037/pro0000153]
85. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
86. Possemato K, Kuhn E, Johnson EM, Hoffman JE, Brooks E. Development and refinement of a clinician intervention to facilitate primary care patient use of the PTSD Coach app. Transl Behav Med 2017 Mar;7(1):116-126 [FREE Full text] [doi: 10.1007/s13142-016-0393-9] [Medline: 27234150]
87. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
88. 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;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
89. 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]
90. Kuhn E, van der Meer C, Owen JE, Hoffman JE, Cash R, Carrese P, et al. PTSD Coach around the world. Mhealth 2018;4:15 [FREE Full text] [doi: 10.21037/mhealth.2018.05.01] [Medline: 29963560]
91. Cernvall M, Sveen J, Johannesson KB, Arnberg F. A pilot study of user satisfaction and perceived helpfulness of the Swedish version of the mobile app PTSD Coach. Eur J Psychotraumatol 2018;9(Suppl 1):1472990 [FREE Full text] [doi: 10.1080/20008198.2018.1472990] [Medline: 29805783]
92. de Almeida RS, Martins C, Marques A, Benevides D, Costa A, Queirós C, et al. Development of a hybrid application for psychotic disorders self-management. In: Advances in Intelligent Systems and Computing. Volume 615. Cham: Springer; 2017:229-237.
93. Terp M, Jørgensen R, Laursen BS, Mainz J, Bjørnes CD. A smartphone app to foster power in the everyday management of living with schizophrenia: qualitative analysis of young adults' perspectives. JMIR Ment Health 2018 Oct 1;5(4):e10157 [FREE Full text] [doi: 10.2196/10157] [Medline: 30274966]

94. Bell IH, Fielding-Smith SF, Hayward M, Rossell SL, Lim MH, Farhall J, et al. Smartphone-based ecological momentary assessment and intervention in a coping-focused intervention for hearing voices (SAVVy): study protocol for a pilot randomised controlled trial. Trials 2018 May 2;19(1):262 [FREE Full text] [doi: 10.1186/s13063-018-2607-6] [Medline: 29720208]
95. Bell IH, Fielding-Smith SF, Hayward M, Rossell SL, Lim MH, Farhall J, et al. Smartphone-based ecological momentary assessment and intervention in a blended coping-focused therapy for distressing voices: Development and case illustration. Internet Interv 2018 Dec;14:18-25 [FREE Full text] [doi: 10.1016/j.invent.2018.11.001] [Medline: 30510910]
96. Garety PA, Ward T, Freeman D, Fowler D, Emsley R, Dunn G, et al. SlowMo, a digital therapy targeting reasoning in paranoia, versus treatment as usual in the treatment of people who fear harm from others: study protocol for a randomised controlled trial. Trials 2017 Nov 2;18(1):510 [FREE Full text] [doi: 10.1186/s13063-017-2242-7] [Medline: 29096681]
97. Husain N, Gire N, Kelly J, Duxbury J, McKeown M, Riley M, et al. TechCare: mobile assessment and therapy for psychosis - an intervention for clients in the Early Intervention Service: a feasibility study protocol. SAGE Open Med 2016;4:2050312116669613 [FREE Full text] [doi: 10.1177/2050312116669613] [Medline: 27790373]
98. Jongeneel A, Scheffers D, Tromp N, Nuij C, Delespaul P, Riper H, et al. Reducing distress and improving social functioning in daily life in people with auditory verbal hallucinations: study protocol for the 'Temstem' randomised controlled trial. BMJ Open 2018 Mar 6;8(3):e020537 [FREE Full text] [doi: 10.1136/bmjopen-2017-020537] [Medline: 29511020]
99. Bucci S, Barrowclough C, Ainsworth J, Morris R, Berry K, Machin M, et al. Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. Trials 2015 Sep 10;16:404 [FREE Full text] [doi: 10.1186/s13063-015-0943-3] [Medline: 26357943]
100. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
101. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
102. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
103. Kim S, Lee GY, Yu HY, Jung EI, Lee JY, Kim SY, et al. Development and feasibility of smartphone application for cognitive-behavioural case management of individuals with early psychosis. Early Interv Psychiatry 2018 Dec;12(6):1087-1093. [doi: 10.1111/eip.12418] [Medline: 28516480]
104. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: a cognitive neuroscience-informed mobile app intervention to enhance motivated behavior and improve quality of life in recent onset schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
105. Schlosser DA, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]
106. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
107. Abrantes AM, Blevins C, Lindsay C, Battle CL, Buman MP, Agu E, et al. Formative work in the development of a physical activity smartphone app targeted for patients with alcohol use disorders. Psychol Sport Exerc 2019;41:162-171. [doi: 10.1016/j.psychsport.2018.02.007]
108. Leonard NR, Silverman M, Sherpa DP, Naegle MA, Kim H, Coffman DL, et al. Mobile health technology using a wearable sensorband for female college students with problem drinking: an acceptability and feasibility study. JMIR Mhealth Uhealth 2017 Jul 7;5(7):e90 [FREE Full text] [doi: 10.2196/mhealth.7399] [Medline: 28687533]
109. Liang D, Han H, Du J, Zhao M, Hser Y. A pilot study of a smartphone application supporting recovery from drug addiction. J Subst Abuse Treat 2018 May;88:51-58 [FREE Full text] [doi: 10.1016/j.jsat.2018.02.006] [Medline: 29606226]
110. Quanbeck A, Gustafson DH, Marsch LA, Chih MY, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
111. 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 Dec;21(12):998-1004 [FREE Full text] [doi: 10.1089/tmj.2014.0232] [Medline: 26171733]
112. Alsharif AH, Philip N. Cognitive Behavioural Therapy Embedding Smoking Cessation Program Using Smart Phone Technologies. In: Proceedings of the 2015 5th World Congress on Information and Communication Technologies. 2015 Presented at: WICT'15; December 14-16, 2015; Marrakesh, Morocco p. 134-139. [doi: 10.1109/wict.2015.7489660]
113. Jódar-Sánchez F, Hernández LC, Núñez-Benjumea FJ, González MA, Conde JM, Calderón CL, et al. Using the social-local-mobile app for smoking cessation in the SmokeFreeBrain project: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 6;7(12):e12464 [FREE Full text] [doi: 10.2196/12464] [Medline: 30522992]

114. Dennis ML, Scott CK, Funk RR, Nicholson L. A pilot study to examine the feasibility and potential effectiveness of using smartphones to provide recovery support for adolescents. Subst Abus 2015;36(4):486-492 [FREE Full text] [doi: 10.1080/08897077.2014.970323] [Medline: 25310057]
115. Scott CK, Dennis ML, Gustafson DH. Using smartphones to decrease substance use via self-monitoring and recovery support: study protocol for a randomized control trial. Trials 2017 Aug 10;18(1):374 [FREE Full text] [doi: 10.1186/s13063-017-2096-z] [Medline: 28797307]
116. Mellentin AI, Nielsen B, Nielsen AS, Yu F, Stenager E. A randomized controlled study of exposure therapy as aftercare for alcohol use disorder: study protocol. BMC Psychiatry 2016 Apr 21;16:112 [FREE Full text] [doi: 10.1186/s12888-016-0795-8] [Medline: 27098817]
117. Mellentin AI, Stenager E, Nielsen B, Nielsen AS, Yu F. A smarter pathway for delivering cue exposure therapy? The design and development of a smartphone app targeting alcohol use disorder. JMIR Mhealth Uhealth 2017 Jan 30;5(1):e5 [FREE Full text] [doi: 10.2196/mhealth.6500] [Medline: 28137701]
118. Lawn S, van Agteren J, Zabeen S, Bertossa S, Barton C, Stewart J. Adapting, pilot testing and evaluating the Kick.it app to support smoking cessation for smokers with severe mental illness: a study protocol. Int J Environ Res Public Health 2018 Feb 3;15(2):pii: E254 [FREE Full text] [doi: 10.3390/ijerph15020254] [Medline: 30720772]
119. van Agteren JE, Lawn S, Bonevski B, Smith BJ. Kick.it: the development of an evidence-based smoking cessation smartphone app. Transl Behav Med 2018 Mar 1;8(2):243-267. [doi: 10.1093/tbm/ibx031] [Medline: 29447386]
120. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: Evaluation of feasibility and effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
121. Hébert ET, Stevens EM, Frank SG, Kendzor DE, Wetter DW, Zvolensky MJ, et al. An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors. Addict Behav 2018 Mar;78:30-35 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.026] [Medline: 29121530]
122. Christoforou M, Fonseca JA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. J Med Internet Res 2017 Nov 24;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
123. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the Internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
124. Carper MM. Multimedia field test thinking about exposures? There's an app for that!. Cogn Behav Pract 2017 Feb;24(1):121-127 [FREE Full text] [doi: 10.1016/j.cbpra.2016.11.001] [Medline: 28757791]
125. Michelle TQ, Jarzabek S, Wadhwa B. CBT Assistant: mHealth App for Psychotherapy. In: Proceedings of the 2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite. 2014 Presented at: GHTC-SAS'14; September 26-27, 2014; Trivandrum, India p. 135-140 URL: https://doi.org/10.1109/GHTC-SAS.2014.6967572 [doi: 10.1109/ghtc-sas.2014.6967572]
126. Boettcher J, Magnusson K, Marklund A, Berglund E, Blomdahl R, Braun U, et al. Adding a smartphone app to internet-based self-help for social anxiety: a randomized controlled trial. Comput Hum Behav 2018;87:98-108. [doi: 10.1016/j.chb.2018.04.052]
127. Szigethy E, Solano F, Wallace M, Perry DL, Morrell L, Scott K, et al. A study protocol for a non-randomised comparison trial evaluating the feasibility and effectiveness of a mobile cognitive-behavioural programme with integrated coaching for anxious adults in primary care. BMJ Open 2018 Jan 13;8(1):e019108 [FREE Full text] [doi: 10.1136/bmjopen-2017-019108] [Medline: 29331971]
128. Clough BA, Casey LM. Therapy on the move: the development of a therapeutic smartphone application. Int J Cyber Behav Psychol Learn 2015;5(1):33-41 [FREE Full text] [doi: 10.4018/ijcbpl.2015010103]
129. Stupar-Rutenfrans S, Ketelaars LE, van Gisbergen MS. Beat the fear of public speaking: mobile 360° video virtual reality exposure training in home environment reduces public speaking anxiety. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):624-633. [doi: 10.1089/cyber.2017.0174] [Medline: 29039704]
130. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate Cognitive Behavioral Therapy skills practice in child anxiety treatment: Open clinical trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
131. Stegemann SK, Ebenfeld L, Lehr D, Berking M, Funk B. Development of a Mobile Application for People With Panic Disorder as Augmentation for an Internet-Based Intervention. In: Proceedings of the 2013 Federated Conference on Computer Science and Information Systems. 2013 Presented at: FedCSIS'13; September 8-11, 2013; Kraków, Poland p. 1331-1337 URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84892495861&partnerID=40&md5=538ed7c5ceab852771deb97f8eb43ce0
132. Ebenfeld L, Stegemann SK, Lehr D, Ebert DD, Jazaieri H, van Ballegooijen W, et al. Efficacy of a hybrid online training for panic symptoms and agoraphobia: study protocol for a randomized controlled trial. Trials 2014 Nov 4;15:427 [FREE Full text] [doi: 10.1186/1745-6215-15-427] [Medline: 25370504]

133. Cruz L, Rubin J, Abreu R, Ahern S, Eldardiry H, Bobrow DG. A Wearable and Mobile Intervention Delivery System for Individuals With Panic Disorder. In: Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia. 2015 Presented at: MUM'15; November 30 - December 2, 2015; Linz, Austria p. 175-182 URL: https://doi.org/10.1145/2836041.2836058 [doi: 10.1145/2836041.2836058]
134. Lindner P, Ivanova E, Ly KH, Andersson G, Carlbring P. Guided and unguided CBT for social anxiety disorder and/or panic disorder via the Internet and a smartphone application: study protocol for a randomised controlled trial. Trials 2013 Dec 18;14:437 [FREE Full text] [doi: 10.1186/1745-6215-14-437] [Medline: 24351088]
135. Kumar S, Bell MJ, Juusola JL. Mobile and traditional cognitive behavioral therapy programs for generalized anxiety disorder: a cost-effectiveness analysis. PLoS One 2018;13(1):e0190554 [FREE Full text] [doi: 10.1371/journal.pone.0190554] [Medline: 29300754]
136. Benton SA, Heesacker M, Snowden SJ, Lee G. Therapist-assisted, online (TAO) intervention for anxiety in college students: TAO outperformed treatment as usual. Prof Psychol Res Pr 2016;47(5):363-371 [FREE Full text] [doi: 10.1037/pro0000097]
137. Repetto C, Gaggioli A, Pallavicini F, Cipresso P, Raspelli S, Riva G. Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. Pers Ubiquit Comput 2013;17(2):253-260 [FREE Full text] [doi: 10.1007/s00779-011-0467-0]
138. Brown RK, Petty S, O'Malley S, Stojanovska J, Davenport MS, Kazerooni EA, et al. Virtual reality tool simulates MRI experience. Tomography 2018 Sep;4(3):95-98 [FREE Full text] [doi: 10.18383/j.tom.2018.00023] [Medline: 30320208]
139. Stolz T, Schulz A, Krieger T, Vincent A, Urech A, Moser C, et al. A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. J Consult Clin Psychol 2018 Jun;86(6):493-504. [doi: 10.1037/ccp0000301] [Medline: 29781648]
140. Levin ME, Haeger J, Pierce B, Cruz RA. Evaluating an adjunctive mobile app to enhance psychological flexibility in acceptance and commitment therapy. Behav Modif 2017 Nov;41(6):846-867. [doi: 10.1177/0145445517719661] [Medline: 28689449]
141. Ben-Zeev D, Brian RM, Jonathan G, Razzano L, Pashka N, Carpenter-Song E, et al. Mobile health (mHealth) versus clinic-based group intervention for people with serious mental illness: a randomized controlled trial. Psychiatr Serv 2018 Sep 1;69(9):978-985. [doi: 10.1176/appi.ps.201800063] [Medline: 29793397]
142. Mistler LA, Ben-Zeev D, Carpenter-Song E, Brunette MF, Friedman MJ. Mobile mindfulness intervention on an acute psychiatric unit: feasibility and acceptability study. JMIR Ment Health 2017 Aug 21;4(3):e34 [FREE Full text] [doi: 10.2196/mental.7717] [Medline: 28827214]
143. Povey J, Mills PP, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
144. Mitrpanont J, Phandhu-Fung J, Klubdee N, Ratanalaor S, Mitrpanont T. iCare-Stress: An Integrated Mental Health Software. In: Proceedings of the 2017 2nd International Conference on Information Technology. 2017 Presented at: INCIT'17; November 2-3, 2017; Nakhonpathom, Thailand p. 1-6. [doi: 10.1109/incit.2017.8257889]
145. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017 Jan 5;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
146. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: MoodMission. Cogn Behav Pract 2018 Nov;25(4):496-514. [doi: 10.1016/j.cbpra.2018.07.002]
147. Addepally SA, Purkayastha S. Mobile-Application Based Cognitive Behavior Therapy (CBT) for Identifying and Managing Depression and Anxiety. In: Proceedings of the International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. 2017 Presented at: DHM'17; July 9–14, 2017; Vancouver, Canada p. 3-12 URL: https://doi.org/10.1007/978-3-319-58466-9_1 [doi: 10.1007/978-3-319-58466-9_1]
148. Proudfoot J, Clarke J, Birch MR, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
149. Fortuna KL, DiMilia PR, Lohman MC, Bruce ML, Zubritsky CD, Halaby MR, et al. Feasibility, acceptability, and preliminary effectiveness of a peer-delivered and technology supported self-management intervention for older adults with serious mental illness. Psychiatr Q 2018 Jun;89(2):293-305 [FREE Full text] [doi: 10.1007/s11126-017-9534-7] [Medline: 28948424]
150. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21 - 26, 2018; Montreal, Canada p. 1-15. [doi: 10.1145/3173574.3173972]
151. Almodovar AS, Surve S, Axon DR, Cooper D, Nahata MC. Self-directed engagement with a mobile app (Sinasprite) and its effects on confidence in coping skills, depression, and anxiety: retrospective longitudinal study. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e64 [FREE Full text] [doi: 10.2196/mhealth.9612] [Medline: 29549066]

152. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
153. Bauer AM, Hodsdon S, Hunter S, Choi Y, Bechtel J, Fortney JC. Lessons from the deployment of the SPIRIT app to support Collaborative care for rural patients with complex psychiatric conditions. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:772-780 [FREE Full text] [doi: 10.1145/3123024.3125610] [Medline: 29075683]
154. Huang ST, Sano A, Kwan CM. The Moment: A Mobile Tool for People With Depression or Bipolar Disorder. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. 2014 Presented at: UbiComp'14 Adjunct; September 13-17, 2014; Seattle, USA p. 235-238 URL: https://doi.org/10.1145/2638728.2638784 [doi: 10.1145/2638728.2638784]
155. Fleming JB, Hill YN, Burns MN. Usability of a culturally informed mHealth intervention for symptoms of anxiety and depression: feedback from young sexual minority men. JMIR Hum Factors 2017 Aug 25;4(3):e22 [FREE Full text] [doi: 10.2196/humanfactors.7392] [Medline: 28842389]
156. Fortuna KL, Lohman MC, Gill LE, Bruce ML, Bartels SJ. Adapting a psychosocial intervention for smartphone delivery to middle-aged and older adults with serious mental illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828 [FREE Full text] [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]
157. Mueller NE, Panch T, Macias C, Cohen BM, Ongur D, Baker JT. Using smartphone apps to promote psychiatric rehabilitation in a peer-led community support program: pilot study. JMIR Ment Health 2018 Aug 15;5(3):e10092 [FREE Full text] [doi: 10.2196/10092] [Medline: 30111526]
158. Nicholson J, Wright SM, Carlisle AM. Pre-post, mixed-methods feasibility study of the WorkingWell mobile support tool for individuals with serious mental illness in the USA: a pilot study protocol. BMJ Open 2018 Feb 6;8(2):e019936 [FREE Full text] [doi: 10.1136/bmjopen-2017-019936] [Medline: 29437756]
159. Nicholson J, Wright SM, Carlisle AM, Sweeney MA, McHugo GJ. The WorkingWell mobile phone app for individuals with serious mental illnesses: proof-of-concept, mixed-methods feasibility study. JMIR Ment Health 2018 Oct 25;5(4):e11383 [FREE Full text] [doi: 10.2196/11383] [Medline: 30361199]
160. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: Feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
161. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
162. Giosan C, Cobeanu O, Mogoaşe C, Szentagotai A, Mureşan V, Boian R. Reducing depressive symptomatology with a smartphone app: study protocol for a randomized, placebo-controlled trial. Trials 2017 May 12;18(1):215 [FREE Full text] [doi: 10.1186/s13063-017-1960-1] [Medline: 28494802]
163. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs computer. BMC Psychiatry 2013;13(1):49 [FREE Full text] [doi: 10.1186/1471-244x-13-49]
164. Deady M, Johnston DA, Glozier N, Milne D, Choi I, Mackinnon A, et al. A smartphone application for treating depressive symptoms: study protocol for a randomised controlled trial. BMC Psychiatry 2018 Jun 1;18(1):166 [FREE Full text] [doi: 10.1186/s12888-018-1752-5] [Medline: 29859060]
165. Weisel KK, Zarski A, Berger T, Schaub MP, Krieger T, Moser CT, et al. Transdiagnostic tailored internet- and mobile-based guided treatment for major depressive disorder and comorbid anxiety: study protocol of a randomized controlled trial. Front Psychiatry 2018;9:274 [FREE Full text] [doi: 10.3389/fpsyt.2018.00274] [Medline: 30022954]
166. Mohr DC, Stiles-Shields C, Brenner C, Palac H, Montague E, Kaiser SM, et al. MedLink: a mobile intervention to address failure points in the treatment of depression in general medicine. Int Conf Pervasive Comput Technol Healthc 2015 May;2015:100-107 [FREE Full text] [doi: 10.4108/icst.pervasivehealth.2015.259042] [Medline: 26640740]
167. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
168. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
169. Dahne J, Lejuez CW, Kustanowitz J, Felton JW, Diaz VA, Player MS, et al. Moodivate: a self-help behavioral activation mobile app for utilization in primary care-Development and clinical considerations. Int J Psychiatry Med 2017 Mar;52(2):160-175 [FREE Full text] [doi: 10.1177/0091217417720899] [Medline: 28792292]
170. Gordon M, Henderson R, Holmes JH, Wolters MK, Bennett IM, SPIRIT (Stress in Pregnancy: Improving Results with Interactive Technology) Group. Participatory design of ehealth solutions for women from vulnerable populations with perinatal depression. J Am Med Inform Assoc 2016 Jan;23(1):105-109. [doi: 10.1093/jamia/ocv109] [Medline: 26342219]

171. Schlosser DA, Campellone TR, Truong B, Anguera JA, Vergani S, Vinogradov S, et al. The feasibility, acceptability, and outcomes of PRIME-D: A novel mobile intervention treatment for depression. Depress Anxiety 2017 Jun;34(6):546-554 [FREE Full text] [doi: 10.1002/da.22624] [Medline: 28419621]
172. Mehrotra S, Sudhir P, Rao G, Thirthalli J, Srikanth T. Development and pilot testing of an internet-based self-help intervention for depression for Indian users. Behav Sci (Basel) 2018 Mar 22;8(4):pii: E36 [FREE Full text] [doi: 10.3390/bs8040036] [Medline: 29565278]
173. Panatagama AP, Pratama GP, Wibawa DY. SocioEmpathy: A Social-Sensitivity Application to Reduce Stress and Depression of Divorce or Domestic Violence Victims. In: Proceedings of the 2018 6th International Conference on Information and Communication Technology. 2018 Presented at: ICoICT'18; May 3-5, 2018; Bandung, Indonesia p. 92-97. [doi: 10.1109/icoict.2018.8528755]
174. Takahashi K, Takada K, Hirao K. Feasibility and preliminary efficacy of a smartphone application intervention for subthreshold depression. Early Interv Psychiatry 2019 Feb;13(1):133-136. [doi: 10.1111/eip.12540] [Medline: 29356332]
175. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
176. Caplan S, Lovera AS, Liberato PR. A feasibility study of a mental health mobile app in the Dominican Republic: the untold story. Int J Ment Health 2018;47(4):311-345. [doi: 10.1080/00207411.2018.1553486]
177. Stiles-Shields C, Montague E, Lattie EG, Schueller SM, Kwasny MJ, Mohr DC. Exploring user learnability and learning performance in an app for depression: usability study. JMIR Hum Factors 2017 Aug 11;4(3):e18 [FREE Full text] [doi: 10.2196/humanfactors.7951] [Medline: 28801301]
178. Hur JW, Kim B, Park D, Choi SW. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemed J E Health 2018 Sep;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
179. Watanabe N, Horikoshi M, Yamada M, Shimodera S, Akechi T, Miki K, Steering Committee of the Fun to Learn to Act and Think through Technology Project. Adding smartphone-based cognitive-behavior therapy to pharmacotherapy for major depression (FLATT project): study protocol for a randomized controlled trial. Trials 2015 Jul 7;16:293 [FREE Full text] [doi: 10.1186/s13063-015-0805-z] [Medline: 26149441]
180. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: Randomized controlled trial. J Med Internet Res 2017 Nov 3;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
181. Furukawa TA, Horikoshi M, Fujita H, Tsujino N, Jinnin R, Kako Y, et al. Cognitive and behavioral skills exercises completed by patients with major depression during smartphone cognitive behavioral therapy: Secondary analysis of a randomized controlled trial. JMIR Ment Health 2018 Jan 11;5(1):e4 [FREE Full text] [doi: 10.2196/mental.9092] [Medline: 29326098]
182. Hung GL, Yang PC, Wang CY, Chiang JH. A Smartphone-Based Personalized Activity Recommender System for Patients with Depression. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. 2015 Presented at: MOBIHEALTH'15; October 14-16, 2015; London p. 253-257 URL: https://doi.org/10.4108/eai.14-10-2015.2261655 [doi: 10.4108/eai.14-10-2015.2261655]
183. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
184. Giosan C, Mogoaşe C, Cobeanu O, Tătar AS, Mureşan V, Boian R. Using a smartphone app to reduce cognitive vulnerability and mild depressive symptoms: Study protocol of an exploratory randomized controlled trial. Trials 2016 Dec 28;17(1):609 [FREE Full text] [doi: 10.1186/s13063-016-1740-3] [Medline: 28031038]
185. Yoo JS, Kim IS, Gwon JA. Design and implementation of smartphone application for measurement and management of depressive emotions in adolescents using cognitive behavioral therapy. In: HCI International 2017 – Posters' Extended Abstracts. Cham: Springer; 2017:317-323.
186. Lüdtke T, Westermann S, Pult LK, Schneider BC, Pfuhl G, Moritz S. Evaluation of a brief unguided psychological online intervention for depression: a controlled trial including exploratory moderator analyses. Internet Interv 2018 Sep;13:73-81 [FREE Full text] [doi: 10.1016/j.invent.2018.06.004] [Medline: 30206522]
187. Vázquez FL, Torres A, Díaz O, Páramo M, Otero P, Blanco V, et al. Cognitive behavioral intervention via a smartphone app for non-professional caregivers with depressive symptoms: study protocol for a randomized controlled trial. Trials 2018 Jul 31;19(1):414 [FREE Full text] [doi: 10.1186/s13063-018-2793-2] [Medline: 30064466]
188. World Health Organization. Mental Health Atlas 2017. Geneva: World Health Organization; 2018.
189. González-González S, González M, García MC, Ruiz M, Rojo MV, Sandoya M, et al. Psychological treatments features and outcomes in Spanish public mental health centres. Int J Psychol Psychol Ther 2014;14(1):17-32 [FREE Full text]
190. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
191. Kazdin AE. Technology-based interventions and reducing the burdens of mental illness: perspectives and comments on the special series. Cogn Behav Pract 2015;22(3):359-366 [FREE Full text] [doi: 10.1016/j.cbpra.2015.04.004]

192. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
193. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
194. Huang TL, Liao S. A model of acceptance of augmented-reality interactive technology: the moderating role of cognitive innovativeness. Electron Commer Res 2015;15(2):269-295 [FREE Full text] [doi: 10.1007/s10660-014-9163-2]
195. Bullis JR, Boettcher H, Sauer‐Zavala S, Farchione TJ, Barlow DH. What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention. Clin Psychol Sci Pract 2019 Mar 14;26(2):e12278. [doi: 10.1111/cpsp.12278]
196. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
197. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
198. Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. Am J Psychiatry 2005 Jun;162(6):1171-1178. [doi: 10.1176/appi.ajp.162.6.1171] [Medline: 15930066]
199. Andlin-Sobocki P, Wittchen HU. Cost of anxiety disorders in Europe. Eur J Neurol 2005 Jun;12(Suppl 1):39-44. [doi: 10.1111/j.1468-1331.2005.01196.x] [Medline: 15877777]
200. Mendlowicz MV, Stein MB. Quality of life in individuals with anxiety disorders. Am J Psychiatry 2000 May;157(5):669-682. [doi: 10.1176/appi.ajp.157.5.669] [Medline: 10784456]
201. Cuijpers P, Beekman AT, Reynolds CF. Preventing depression: a global priority. J Am Med Assoc 2012 Mar 14;307(10):1033-1034 [FREE Full text] [doi: 10.1001/jama.2012.271] [Medline: 22416097]
202. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
203. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
204. Gunderson JG. Borderline personality disorder: ontogeny of a diagnosis. Am J Psychiatry 2009 May;166(5):530-539 [FREE Full text] [doi: 10.1176/appi.ajp.2009.08121825] [Medline: 19411380]
205. Lenzenweger MF, Lane MC, Loranger AW, Kessler RC. DSM-IV personality disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007 Sep 15;62(6):553-564 [FREE Full text] [doi: 10.1016/j.biopsych.2006.09.019] [Medline: 17217923]
206. Lieb K, Zanarini MC, Schmahl C, Linehan MM, Bohus M. Borderline personality disorder. Lancet 2004;364(9432):453-461. [doi: 10.1016/S0140-6736(04)16770-6] [Medline: 15288745]
207. Giota KG, Kleftaras G. Mental health apps: innovations, risks and ethical considerations. Ehealth Telecommun Syst Netw 2014;3(3):19-23. [doi: 10.4236/etsn.2014.33003]
208. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
209. Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: are some patients taking the quantified self too far? J Clin Sleep Med 2017 Feb 15;13(2):351-354 [FREE Full text] [doi: 10.5664/jcsm.6472] [Medline: 27855740]
210. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 5;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
211. Parker L, Halter V, Karliychuk T, Grundy Q. How private is your mental health app data? An empirical study of mental health app privacy policies and practices. Int J Law Psychiatry 2019;64:198-204. [doi: 10.1016/j.ijlp.2019.04.002] [Medline: 31122630]
212. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
213. Ackerman L. Privacy Rights Clearinghouse. 2013. Mobile Health and Fitness Applications and Information Privacy: Report to California Consumer URL: https://privacyrights.org/sites/default/files/pdfs/mobile-medical-apps-privacy-consumer-report.pdf [accessed 2020-02-07]
214. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
215. Harris KD. State of California - Office of the Attorney General. 2013 Jan. Privacy on the Go: Recommendations for the Mobile Ecosystem URL: http://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/privacy_on_the_go.pdf [accessed 2018-03-09]

216. Mense A, Steger S, Sulek M, Jukic-Sunaric D, Mészáros A. Analyzing privacy risks of mHealth applications. Stud Health Technol Inform 2016;221:41-45 [FREE Full text] [Medline: 27071873]
217. Kazdin AE, Rabbitt SM. Novel models for delivering mental health services and reducing the burdens of mental illness. Clin Psychol Sci 2013;1(2):170-191. [doi: 10.1177/2167702612463566]
218. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
219. EUR-Lex. Regulation (EU). (2017). 2017/745 of the European parliament and of the council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 [accessed 2018-03-09]
220. EUR-Lex. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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 URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2020-02-05]
221. US Food and Drug Administration, US Department of Health and Human Services, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Food and Drug Administration. Policy for Device Software Functions and Mobile Medical Applications. Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/media/80958/download [accessed 2019-10-01]
222. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Info Syst J 2018;28(6):995-1019 [FREE Full text] [doi: 10.1111/isj.12179]
223. Chiauzzi E, Newell A. Mental health apps in psychiatric treatment: a patient perspective on real world technology usage. JMIR Ment Health 2019 Apr 22;6(4):e12292 [FREE Full text] [doi: 10.2196/12292] [Medline: 31008711]
224. Osma J, Sprenger M, Mettler T. Introduction of e-mental health in national health systems – a health professionals' perspective. Health Policy Technol 2017;6(4):436-445. [doi: 10.1016/j.hlpt.2017.07.001]
225. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
226. Simblett S, Matcham F, Siddi S, Bulgari V, di San Pietro CB, López JH, RADAR-CNS Consortium. Barriers to and facilitators of engagement With mHealth technology for remote measurement and management of depression: Qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]

Abbreviations

APA: American Psychological Association
CBT: cognitive behavioral therapy
DSM: Diagnostic and Statistical Manual of Mental Disorders
EC: exclusion criteria
eHealth: electronic health
EMI: ecological momentary intervention
EU: European Union
IC: inclusion criteria
ICT: information and communication technologies
MEDLINE: Medical Literature Analysis and Retrieval System Online
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
UX: user experience
VR: virtual reality
WoS: Web of Science

Edited by J Car; submitted 03.06.19; peer-reviewed by J Osma López, B Barry, L Martinengo, A Perez-Navarro; comments to author 11.07.19; revised version received 05.12.19; accepted 20.01.20; published 02.04.20.

Please cite as:
Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, Castilla D, Casteleyn S
Smartphone Apps for the Treatment of Mental Disorders: Systematic Review
JMIR Mhealth Uhealth 2020;8(4):e14897
URL: https://mhealth.jmir.org/2020/4/e14897
doi:10.2196/14897
PMID:32238332

Review

Intervention and Evaluation of Mobile Health Technologies in Management of Patients Undergoing Chronic Dialysis: Scoping Review

Yang Yang[1], MSc, MBET; Helen Chen[1], PhD; Hammad Qazi[1], PhD; Plinio P Morita[1], PhD

School of Public Health and Health Systems, Faculty of Applied Health Sciences, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Yang Yang, MSc, MBET
School of Public Health and Health Systems
Faculty of Applied Health Sciences
University of Waterloo
Toby Jenkins Applied Health Research Building
Waterloo, ON, N2L 3G1
Canada
Phone: 1 (226) 2200127
Email: y24yang@uwaterloo.ca

Abstract

Background: Studies have shown the effectiveness and user acceptance of mobile health (mHealth) technologies in managing patients with chronic kidney disease (CKD). However, incorporating mHealth technology into the standard care of patients with CKD still faces many challenges. To our knowledge, there are no reviews on mHealth interventions and their assessments concerning the management of patients undergoing dialysis.

Objective: This study provided a scoping review on existing apps and interventions of mHealth technologies in adult patients undergoing chronic dialysis and identified the gaps in patient outcome assessment of mHealth technologies in the literature.

Methods: We systematically searched PubMed (MEDLINE), Scopus, and the Cumulative Index to Nursing and Allied Health Literature databases, as well as gray literature sources. Two keywords, “mHealth” and “dialysis,” were combined to address the main concepts of the objectives. Inclusion criteria were as follows: (1) mHealth interventions, which are on a smartphone, tablet, or web-based portals that are accessible through mobile devices; and (2) adult patients (age ≥18 years) on chronic dialysis. Only English papers published from January 2008 to October 2018 were included. Studies with mHealth apps for other chronic conditions, based on e-consultation or videoconferencing, non-English publications, and review papers were excluded.

Results: Of the 1054 papers identified, 22 met the inclusion and exclusion criteria. Most studies (n=20) were randomized controlled trials and cohort studies. These studies were carried out in 7 countries. The main purposes of these mHealth interventions were as follows: nutrition or dietary self-monitoring (n=7), remote biometric monitoring (n=7), web-based portal (n=4), self-monitoring of in-session dialysis-specific information (n=3), and self-monitoring of lifestyle or behavioral change (n=1). The outcomes of the 22 included studies were organized into five categories: (1) patient satisfaction and acceptance, (2) clinical effectiveness, (3) economic assessment, (4) health-related quality of life, and (5) impact on lifestyle or behavioral change. The mHealth interventions showed neutral to positive results in chronic dialysis patient management, reporting no to significant improvement of dialysis-specific measurements and some components of the overall quality of life assessment. Evaluation of these mHealth interventions consistently demonstrated evidence in patients’ satisfaction, high level of user acceptance, and reduced use of health resources and cost savings to health care services. However, there is a lack of studies evaluating safety, organizational, sociocultural, ethical, and legal aspects of mHealth technologies. Furthermore, a comprehensive cost-effectiveness and cost-benefit analysis of adopting mHealth technologies was not found in the literature.

Conclusions: The gaps identified in this study will inform the creation of health policies and organizational support for mHealth implementation in patients undergoing dialysis. The findings of this review will inform the development of a comprehensive service model that utilizes mHealth technologies for home monitoring and self-management of patients undergoing chronic dialysis.

(JMIR Mhealth Uhealth 2020;8(4):e15549) doi:10.2196/15549

KEYWORDS

mobile health; renal dialysis; health technology assessment; patient outcome assessment

Introduction

Background

Mobile technologies have changed the way individuals communicate and are also transforming the health ecosystem by providing patients and health care providers a wide range of supportive tools for monitoring and managing health information, thereby facilitating better delivery of health care services [1]. As a subdomain in the digital health field, mobile health (mHealth) is defined as "medical and public health practice through the use of mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs) and other wireless devices" [2]. More specifically, the mobile technology refers to the "wireless devices and sensors (including mobile phones) that are intended to be worn, carried, or accessed by the person during normal daily activities" [3]. This definition was adopted in this study.

According to the World Health Organization, mHealth technologies have four major application scenarios, including chronic disease monitoring, health management, web-based diagnosis and treatment, and medical appointment scheduling [2]. mHealth technologies have been used to monitor and manage chronic conditions, such as in heart disease, diabetes, hypertension, stroke, asthma, dementia, chronic pain, chronic pulmonary obstructive disease, and chronic kidney disease [4-7]. Dialysis is a medical intervention for patients with end-stage chronic kidney disease. These patients have already suffered a loss of independence by spending a significant amount of time on dialysis treatment in-center or at home [8,9]. No other chronic health condition has such enormous physical and cognitive effects on patients' daily lives [9]. Solutions aiming at reducing time on visiting dialysis centers are especially beneficial to this population, which reflects patient-centered care principle as well as improves patients' quality of life [10]. mHealth technologies can expedite patient communication and facilitate home monitoring and self-management, with the potential of improving patient's overall well-being. However, there is no systematic review on existing literature on mHealth solutions and their effectiveness and benefits in adult chronic dialysis patient management.

The Health Technology Assessment Core Model (HTA-CM), a general framework that facilitates international collaboration of HTA, divides the assessment outcomes into 9 domains, including "health problem and current use of the technology; description and technical characteristics of technology; safety; clinical effectiveness; costs, economic evaluation; ethical analysis; organizational aspects; social aspects; and legal aspects" [11]. The 9-domain HTA-CM has been used in this study, with adult patients undergoing chronic dialysis as a use case for mHealth technologies. Siddique et al [12] evaluated apps of managing CKD using the Mobile App Rating Scale. However, the evaluation of these apps was done by a team of reviewers, mainly on the construction of apps instead of the efficacy and benefits of using apps in patient care and clinical settings.

Objectives

This study aimed to provide a comprehensive review of existing mHealth interventions for adult patients undergoing chronic dialysis and to identify the gaps in the outcome evaluation of mHealth technologies in the literature. The objectives of this study are to (1) summarize the categories of interventions and their main functions of mHealth technologies in the management of adult patients undergoing chronic dialysis through reviewing the existing literature in a systematic approach, (2) examine how these mHealth interventions are evaluated, and (3) identify gaps in the assessment of mHealth technologies in this population.

Methods

Study Design

This study used the five-stage methodological framework for scoping reviews as outlined by Arksey and O'Malley [13], involving (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) collating, summarizing, and reporting the results. The 27-item Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) extension for scoping reviews was employed as the protocol for this study [14].

Identifying the Research Questions

The study population was adult patients undergoing chronic dialysis, and intervention types were services utilizing mHealth technology. Research questions were developed based on an initial literature search and were refined iteratively during the discussions among the research team. The research questions were as follows: (1) what mHealth apps exist to support adult chronic dialysis patients, and (2) how are these mHealth solutions evaluated?

Identifying Relevant Studies

A systematic search strategy was employed to identify relevant literature to the research questions. Two keywords, "mHealth" and "dialysis," were combined to address the two main concepts of the research questions (see Multimedia Appendix 1). The search was performed on PubMed (MEDLINE), Scopus, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases. Furthermore, gray literature was explored from the Canadian Agency for Drugs and Technologies in Health (CADTH)'s "Grey Matters," Health Quality Ontario, Food and Drug Administration, Ottawa Hospital Research Institute, Pan Canada HTA Collaborative, International Information Network on New and Emerging Health Technologies, and Google Scholar and Google search (see Multimedia Appendix 2). In the CADTH's "Grey Matters," the HTA section was used for all countries to find relevant papers. The first 40 records of Google Scholar and Google search results, as well as those from searching the FDA database, were

screened for relevant papers. Two independent researchers (YY and MZ) searched databases and gray literature for references of identified papers published from January 2008 up to October 15, 2018.

Selecting Relevant Papers for the Review

The inclusion and exclusion criteria for this study are listed in Textbox 1. Papers retrieved from each database were imported into the RefWorks software, a web-based reference management software produced by Ex Libris (Jerusalem, Israel), a ProQuest company. Duplicates were removed, and the title and abstract of each paper were screened by YY and JS. They assessed each paper as *included*, *excluded*, or *unsure*. Where there is uncertainty in achieving an agreement (ie, those marked as "unsure" or classified into different categories by 2 reviewers), a full-text review was conducted to determine whether they should be included. The individual screening results were compared, and discrepancies were resolved by consensus through discussion among the research team members. The full-text review was conducted on papers that met the inclusion criteria from the title and abstract review.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria were as follows:

- Adult patients both males and females (age ≥ 18 years)
- Receiving chronic dialysis (in-center or at home)
- Using mobile health (mHealth) interventions
- Papers found using search strategy started from January 2008 to October 2018 appearing at this time frame
- English papers only

Exclusion criteria were as follows:

- Participants aged <18 years (pediatric, adolescent)
- mHealth interventions based on e-consultation or videoconferencing
- Study protocols have no preliminary results
- Non-English publications
- Review papers

Charting the Data

The research team collaboratively developed the data charting form and determined the variables for extraction. The descriptive charting information includes (1) paper general description: first author and year, study design, study location, and patient population; and (2) intervention-specific information: intervention and mHealth app purpose, main functions, delivery method, duration and follow-up period, data collected, outcomes measured, and findings. Data extracted during full-text review of all selected papers are summarized in Multimedia Appendix 3.

Collating, Summarizing, and Reporting the Results

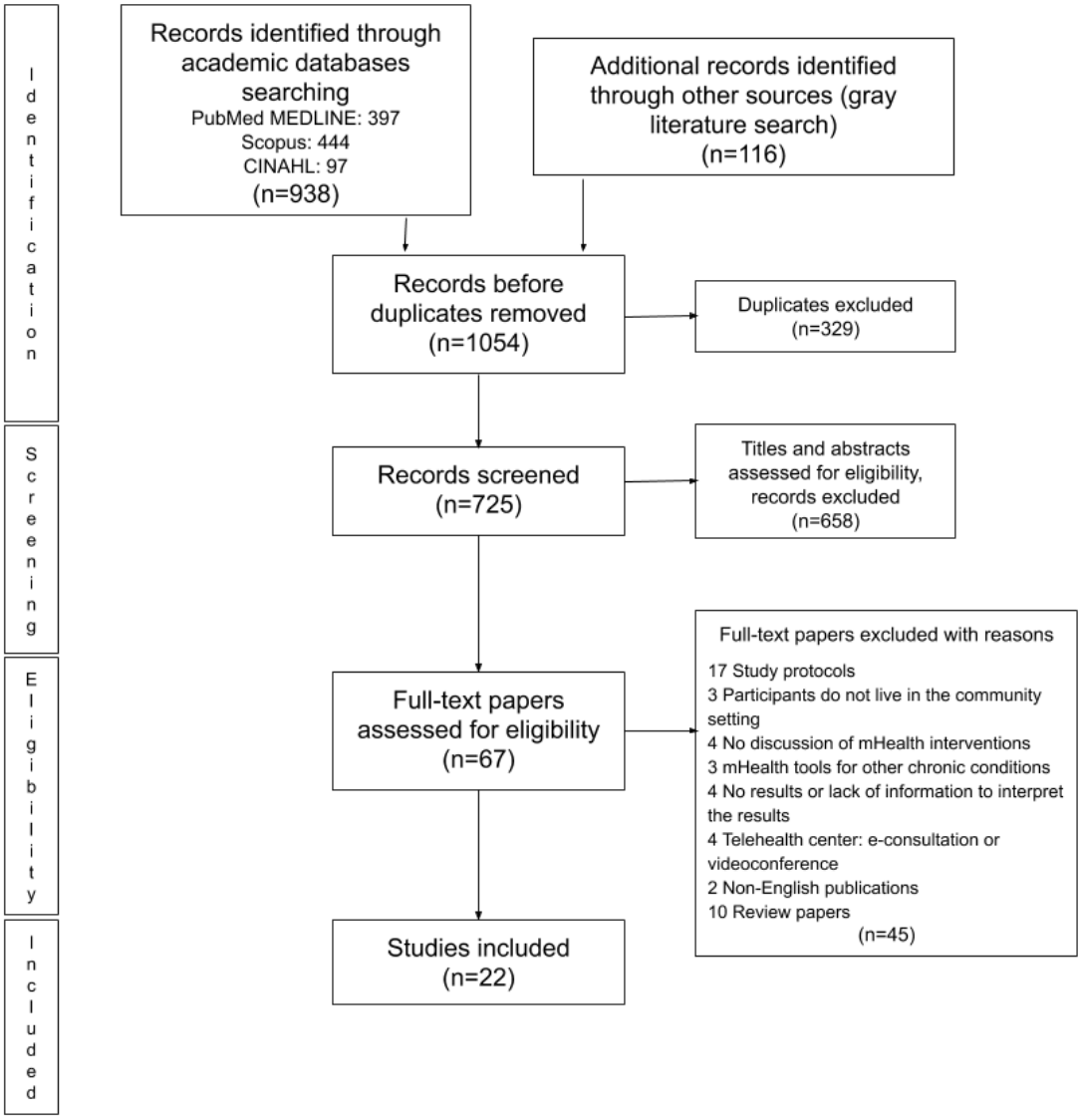
The general description of the reviewed papers was collated according to the descriptive characteristics of studies. The thematic content analysis of interventions and associated outcomes for each study was performed by the research team after a concurrent review of the charted data. First, codes were developed and applied to analyze the data. The coded segments of all the charted data were then created with color-coded quotations. Furthermore, the code summary was organized into an Excel table for thematic content analysis. This table was sorted by codes and density, looking for recurrent patterns that were addressed by the included papers, both across the whole dataset to compare the included studies and within each study until the key themes were identified. The research findings were summarized with the outlined objectives of this study.

Results

Selection of Included Papers

A total of 938 published papers were retrieved from the PubMed (MEDLINE), Scopus, and CINAHL databases, including one paper that was identified through a reference review of identified papers. The gray literature search resulted in 116 additional papers. Of the 1054 identified papers, 329 duplicates were excluded. Of the remaining 725 papers, 67 were selected for a full-text review based on the screening results of titles and abstracts. After the full-text review, 22 papers were included in this scoping review. Figure 1 shows the PRISMA flow diagram.

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow diagram. mHealth: mobile health.



General Characteristics of the Included Studies

Among 22 reported studies, cohort study was the most common study type (12 prospective and 4 retrospective studies), followed by randomized controlled trials (RCTs, 4 studies), 1 mixed method study, and 1 case study. The general characteristics of these studies are presented in Table 1.

Table 1. General characteristics of the included studies.

Characteristics	Studies (n=22)
Study design	
Randomized controlled trial	4
Cohort (prospective)	12
Cohort (retrospective)	4
Others	2
Origin of study	
United States	13
Canada	2
Australia	2
India	2
Others	3
Year of publication	
2014-2018	13
2008-2013	9

Types of Current Mobile Health Interventions

Dialysis patient management services enabled by mHealth technology fall into five main categories: (1) nutrition or dietary self-monitoring (n=7; [15-21]), (2) remote biometric monitoring (RBM) (n=7; [22-28]), (3) web-based portal (n=4; [29-32]), (4) self-monitoring of in-session dialysis-specific information (n=3; [33-35]), and (5) self-monitoring of lifestyle or behavioral change (n=1; [36]). These mHealth interventions were delivered using various technology platforms, including smartphone (7 studies), home-monitoring or telemonitoring unit or telemetry system (6 studies), tablet (5 studies), PDA (4 studies), and fitness tracker (1 study). The mHealth interventions were applied to various dialysis modalities, including hemodialysis (HD, 10 studies), peritoneal dialysis (9 studies) and home hemodialysis (HHD, 4 studies). The study duration ranged from 2 weeks to 42 months.

Outcome Evaluation of Current Mobile Health Interventions

The data collected by these studies included patient characteristics, dialysis in-session indicators, physiological measurements, quality-of-life indicators, health care resource utilization, change of lifestyle or behavior, and usage and user perceptions on mHealth interventions. The thematic content analysis identified five main themes in the evaluation of mHealth interventions: (1) patient satisfaction and acceptance, (2) clinical effectiveness, (3) economic assessment, (4) health related quality of life (HRQoL), and (5) impact on lifestyle or behavioral change. Table 2 lists outcome measures of current mHealth interventions.

Table 2. Outcome evaluation of current mobile health interventions.

Outcome measures (themes)	Studies (n=22)	Main functions (type of use)	Sample sizes (range)	Study duration (range)
Patient satisfaction and acceptance	14	Nutrition or dietary self-monitoring (6); self-monitoring of in-session dialysis-specific information (3); web-based portal (2); RBM[a] (2); self-monitoring of lifestyle or behavioral change (1)	(9, 241)	2 weeks to >15 months
Clinical effectiveness	9	Nutrition or dietary self-monitoring (4); RBM (3); self-monitoring of in-session dialysis-specific information (1); web-based portal (1)	(1, 2424)	30 days to 2 years
Economic assessment	8	RBM (4); web-based portal (3); self-monitoring of in-session dialysis-specific information (1)	(22, 269)	6 to 42 months
HRQoL[b]	5	RBM (2); web-based portal (2); self-monitoring of in-session dialysis-specific information (1)	(20, 60)	2 weeks to 17 months
Impact on lifestyle or behavioral change	5	Nutrition or dietary self-monitoring (4); self-monitoring of lifestyle or behavioral change (1)	(1, 72)	30 days to 6 months

[a]RBM: remote biometric monitoring.

[b]HRQoL: heart-related quality of life.

The studies revealed that mHealth technology has a beneficial effect on dialysis-specific measures. mHealth solutions helped to improve behaviors and reduce the utilization of health resources. Patients in these studies expressed satisfaction and acceptance toward mHealth technologies in dialysis care.

Patient Satisfaction and Acceptance

Patients' attitude toward mHealth interventions was assessed by Likert evaluation questionnaires at the end of the study period. Questions on the questionnaires include ease of use, reliability and performance of the app, perceived usefulness, and overall impression. The evaluation demonstrated affirmative results, reporting positive responses concerning satisfaction and acceptability (8 studies), the ability to understand and use the app (5 studies), confidence in the treatment (1 study), as well as access and accessibility (1 study).

A consistent rating of satisfaction and high level of acceptability was found in studies for self-monitoring of nutrition or dietary, in-session dialysis-specific information, RBM, and online portals, with median satisfaction scores rated over 4 of 5 and similar acceptability scores (approaching 4 of 5) [16,17,22,25,31,33,34]. In addition to high adherence throughout the follow-up reported by Stark et al [19], some studies also revealed that many participants would continue using mHealth interventions beyond the study periods [21,22,35,36]. Patients' willingness to continue using mHealth apps beyond the study periods indicated a high level of acceptability to the mHealth intervention in this cohort. One study reported a positive feedback from the nurses about satisfaction with the remote monitoring system for patients undergoing HHD, referring it a "time-saving tool" [34], and another study demonstrated improved interactions with patients [18].

Of the 5 studies that reported participants' ability to use the app, all the study results demonstrated that the apps were easy to use. This includes interfaces and scanner input mechanism [15], helpful feedback and easy-to-understand instructions [16,17], easy-to-follow interface demonstration while browsing the data [33], and app and web portals that are easy to navigate [31,32].

The use of mHealth technologies, especially facilitating RBM, was found associated with increased confidence in health care activities and reduced negative attitudes toward dialysis care [25]. One study reported patients' confidence in the treatment, with increases in confidence in the treatment and decreased perception of "being a burden to the family" or "having kidney disease takes too much time from the patient's life" [25]. Furthermore, one study evaluated access and accessibility, and most of the participants felt that they had a positive impression on getting access to a renal specialist [31].

Clinical Effectiveness

Nine of the included studies measured dialysis-specific information, including interdialytic weight gain (IWG), blood pressure (BP) pre- and postdialysis, ultrafiltration rate, laboratory tests (hemoglobin, serum potassium, serum phosphorus, serum parathyroid hormone [PTH], serum albumin level), and technique failure rates at baseline and follow-up. The results were identified from these studies, demonstrating neutral to positive results in patients using mHealth interventions, compared with their peers receiving standard care. Most of the results (6/9) showed nonsignificant differences between the baseline and follow-up in the intervention groups, whereas a few studies (3/9) demonstrated significant improvement of dialysis-specific measurement results throughout the intervention periods.

Four studies that examined IWG had inconclusive results, showing nonsignificant differences between the baseline and follow-up period [33], a strong trend for relatively lower IWG (P=.06) in the intervention groups during follow-up [16], a significant mean reduction in weight gain between the groups at the end of the study [26], and a superimposed linear regression of the reduction of average daily IWG for 1 participant during the follow-up in a case study [20]. Except for the case study, the differences in their sample sizes (n=20, n=44, and n=120, respectively) and respective study locations in three different regions (Japan, United States, and Germany, respectively) may have resulted in these inconclusive findings.

Hand et al [18] reported a significant change in high serum PTH between baseline and follow-up, its relatively longer follow-up of 6 months as opposed to 3 months and 6 weeks in other studies, as well as its 3-arm trial designed to separate the effects of algorithm from those of additional care time may account for its statistical significance. Neumann et al [36] showed significant differences on unfiltered ultrafiltration rate, duration on dialysis, and BP between groups at the end of the study. The relatively larger sample size (n=120) as compared with other studies with less than 50 participants may contribute to its significant results. No other studies with laboratory tests identified significant differences in the proportions of patients in the target ranges for hemoglobin, serum potassium, serum phosphorus, or serum albumin level between the start and end of the studies.

In addition to patients' physiological measurements, 2 studies reported significant results in adjusted hazard ratios of all-cause attrition and 5-year survival comparing intervention groups vs matched controls [23,30].

Economic Assessment

Eight of the included studies measured financial aspects, primarily from changes in use of hospital or health care services, including the number of hospitalizations or emergency room (ER) or clinic visits (4 studies), the number of days in the hospital and associated costs (4 studies), as well as nursing time and travel time or distance (3 studies). These studies consistently demonstrated the economic benefit of mHealth technology in cost and resource reduction to health care service providers.

Four studies that evaluated the impact on the utilization of hospital or health care resources reported significant cost savings per study day by reduced number of hospital visits or ER visits and fewer hospital days [27,28], reduced frequency of nurse-initiated telephone contacts required for medical alerts [27], reduced outpatient visit claim payment amounts [24], as well as lower medication cost [29]. The other 4 studies assessed the change in nursing time and telephone usage, as well as clinical interventions from generated alerts using mHealth apps. The results showed that remote reviews contributed to reduced

telephone use [31] and savings of nursing and patient travel times and travel distance from the avoidance of conducting home and (or) unit visits [32,34]. These remote reviews also resulted in changes to dialysis prescriptions, mainly related to adjustments in patients' dry weight [32,34], avoidance of admissions, medication changes, referrals or advice from a dietician or pharmacy, and self-management at home [22].

Health-Related Quality of Life

Health-related quality of life (HRQoL) was measured using the Kidney Disease Quality of Life-36 (KDQOL-36; 2 studies), the 36-Item Short Form Health Survey (SF-36; 2 studies), and the EuroQol Five Dimensions Questionnaire (1 study), with inconclusive study results. Most of the studies (4/5) showed no significant improvement in quality of life (QoL) throughout the study periods. One study demonstrated a considerable improvement on the Social Functioning Score, but all the other KDQOL scores had no significant difference in the intervention group [33]. Another study revealed mixed results with improved SF-12 Physical Health Composite scores, but slightly decreased scores on symptom or problem list, effects of kidney disease, burden of kidney disease, and SF-12 Mental Health Composite at exit [22].

The mixed results on the effects of HRQoL may partly be attributed to the use of varying assessment instruments in different studies; for example, the 2 cohort studies reported QoL utilizing SF-36 and KDQOL-36 separately. SF-36 is recognized as a generic assessment tool that measures QoL for any type of disease, whereas KDQOL-36 is a disease-specific instrument assessing patients with kidney disease. The first 12 items of KDQOL-36 consist of core items that are equivalent to SF-12, and the remaining items assess the burden symptoms and effects of kidney disease [22].

Impact on Lifestyle or Behavioral Change

Five of the included studies reported patients' lifestyle or behavioral change by measuring sodium intake (2 studies), calcium intake (2 studies), self-efficacy on diet proportion estimation (2 studies), or physical activity level and sleep (1 study). Of the 5 studies, 4 studies reported a decreasing trend of calcium and sodium intake and an increasing dietary self-efficacy. One study showed no significant difference in physical activity levels for patients undergoing HD using an off-the-shelf fitness tracker in the intervention group who received the feedback, compared with the control group with no feedback provided [36].

Welch et al [16] showed a significant decrease in sodium intake and calories, a marginal decrease in calories (P=.09) across all patients in the intervention group, and a marginal decrease in protein (P=.08) among active users who used mHealth app more than half of the time during the study.

Two studies that reported self-efficacy had slightly different results. One reported that most of the participants had an improvement in their self-efficacy in pre- and poststudy assessments [17], whereas another one showed no significant improvement between groups or over time in participants' self-efficacy [16]. It revealed that improved behaviors in daily life might not be supported by self-efficacy. While examining perceived control, Welch et al [16] demonstrated a significant group difference with time, where a higher perceived control was reported in the intervention group than the control group at the end of self-monitoring period.

Discussion

Principal Findings

This scoping review summarized the literature on current mHealth technologies in chronic dialysis patient management, providing the types of intervention, outcome evaluation, and the gaps in outcome evaluation of mHealth technology that can be addressed in future research. The high satisfaction ratings are in line with other studies evaluating adherence of mHealth tools in renal transplant recipients [37,38], in elderly, low-income, and vulnerable patient population [8].

Advantages of Mobile Health Technologies for Chronic Dialysis

From the patients' perspectives, mHealth interventions demonstrated a great potential to facilitate the monitoring of symptoms, improve self-management–related physical or psychosocial consequences and associated comorbidities, maintain compliance with dialysis prescription, and improve lifestyle while receiving chronic dialysis [39]. From the health care providers' perspective, rapid advancement in mobile technology enables real-time data capture and exchange between patient self-monitoring devices and a remote monitoring system. This creates opportunities to analyze the data and provide prompt feedback to patient-generated alerts [40].

Nayak et al [40] summarized the existing remote monitoring platforms used to support home dialysis modalities, identifying the system capacity being "two-way, rapid, real-time communication to help troubleshoot problems." Similarly, Jeddi et al [41] demonstrated that the features and functionalities of computerized systems on self-management outcomes of patients with CKD include "inform, record, display, communicate, remind or alert and guide." Digital IT solutions, particularly, mHealth interventions, have a unique advantage to meet this need.

The 22 included studies were multifaceted with respect to general characteristics and intervention-specific mHealth technologies. In a review paper, Havas et al [37] identified 10 aspects of self-management interventions from a patient's perspective, suggesting a complex and multifactorial framework of self-management in patients with CKD. Our findings are in line with some domains, such as getting into routines and using reminder systems, monitoring weight, tracking fluid and food, and modifying lifestyle (ie, self-monitoring of physical activity and sleep).

In this review, 10 of the included studies utilized mHealth interventions for RBM and self-monitoring of in-session dialysis, which is an essential and integral part of self-management in patients undergoing dialysis. This provides a good indication of good interest and motivation of using mHealth technologies facilitating home monitoring and self-care. The web-based portals that have been identified from this review

provide good evidence as mHealth interventions being utilized by both the patients undergoing dialysis as well as their clinical coordinators. Similarly, using commercially available fitness trackers can quantify physical activity levels and sleep data in patients undergoing dialysis, providing the opportunity of promoting a healthier lifestyle in this population.

Patients undergoing dialysis are often required to make significant changes in their dietary intake. A review paper reveals that mHealth nutrition apps usually lead to better adoption of self-monitoring and changes in dietary intake compared with conventional techniques [42]. Due to the complication of dietary adjustments, the restriction on fluid and dietary intake is a major stressor for this cohort [43]. The mHealth interventions for self-management of dietary intake have become increasingly available, assisting in recording food and fluid intake for monitoring or assessing nutrition [43].

Gaps Identified in Outcome Evaluation of Mobile Health Technologies

As part of the multidisciplinary assessment, the HTA-CM introduces 9 domains for assessing the outcomes of health technologies, including "health problem and current use of the technology; description and technical characteristics of technology; safety; clinical effectiveness; costs, economic evaluation; ethical analysis; organizational aspects; social aspects; and legal aspects" [11]. The included studies assessed the efficacy of mHealth interventions in health problems and description of the app, clinical effectiveness, partial costs, and economic evaluation. Although the outcome evaluation was multifaceted with respect to the general characteristics and intervention specific of mHealth technologies, a major gap was identified as the lack of evidence on safety, organizational aspects, as well as sociocultural, ethical, and legal aspects in the included studies.

Safety is a major concern for adopting mobile technologies into patient care. In health care, patient safety is paramount. When introducing mHealth solutions into a care setting, the potential impact and threats on patient safety when using a mHealth solution should be scrutinized. Technical safety related to the reliability and validity of evaluating assessment should also be taken into consideration. Privacy, security, collaboration, data sharing, traceability, and transparency are essential for the enhancement of health care services. Technical barriers such as infrastructure, connectivity, bandwidth, resolution, and frame rate associated with data transmission may affect access and accessibility, posing a potential threat of unplanned downtime between patients and practitioners [44].

When searching for studies, we intentionally did not limit the target users of mHealth apps. However, it is worth noting that all the mHealth solutions in the included studies in this review aimed at patients as their primary users, and none of them was designed for renal care providers or caregivers. Furthermore, none of the mHealth solutions was integrated into electronic health record (EHR) systems in their respective care settings. Integrating mHealth interventions into the EHR system might be a major obstacle to the wider adoption of mHealth solutions in renal care.

While utilizing mHealth technologies and implementing this new service, how this new service will fit within the existing organizational framework is an important question and plays a significant role in the evaluation [44]. This may involve changes in business structure and process, business culture, management on workflow and staff, causing extensive organizational changes, as well as the degree of interoperability between information systems and the impact on resource allocation [44]. When a patient performs the dialysis treatment at home, particularly the HD, there is a need for real-time monitoring of vital signs and providing prompt feedback. This may require care providers to work on a different schedule and workloads. How an mHealth solution impacts the care provider's service and funding model and the level of satisfaction among clinicians should also be assessed?

The significant disease burden owing to loss of time and income often brings many changes to patients with CKD undergoing dialysis in their social and work life with diminished capacities. When planning an mHealth solution, it is important to consider additional responsibility imposed on patients and caregivers by technologies. When introducing mobile technologies into dialysis care, is a patient's autonomy enhanced or compromised? How will we maintain the equity in access to service among different socioeconomic groups? On a larger scale, the impact on professional accreditation and liability, information governance, and patient privacy in terms of consent and access control will also need to be evaluated [44].

Of note, long-term effects on morbidity, including physical and mental health, were not assessed in any of the included studies. It suggested a need for rigorous RCTs with longer follow-up period to capture relevant data and evaluate these effects. In addition, the evaluation of the cost-effectiveness of mHealth technologies in patients undergoing dialysis is not done comprehensively. For example, the IT resources used and the associated cost for supporting an mHealth app were not discussed in any of the included studies. Regarding the business scenario, the funding models, direct costing, annual expenditures related to the resources, annual revenue associated with the number of patients or services on activity and reimbursement, and the return on investment were not evaluated from the institutional level.

Although the study results revealed positive outcomes in reducing the utilization of health resources and saving costs, it contrasted with the results of a large study conducted by Henderson et al [45] in the United Kingdom, evaluating the cost-effectiveness of telehealth technology in patients with chronic conditions (heart failure, chronic obstructive pulmonary disease, or diabetes), compared with the standard practice. The findings suggested that the quality-adjusted life year gained by patients in the intervention group was similar to that of the control arm; however, the telehealth intervention was associated with higher total costs. Thus, they concluded, "Telehealth does not seem to be a cost-effective addition to standard support and treatment." [45]. Although their research neither targeted mHealth interventions nor recruited patients undergoing dialysis, as the largest study evaluating telehealth technology from the cost-effectiveness perspective, it did highlight the need for a

comprehensive economic evaluation framework when assessing mHealth technologies in health care.

One cohort study indicated a slight decrease in perceived QoL in the intervention group from baseline to the midpoint of the study, whereas scores remained the same as their peers receiving standard care [28]. The 21% dropout rate (5/24 patients in the intervention group) may have contributed to such a result because these dropout patients were representative of a sicker population with potential lower perceived QoL.

Implications for Future Research

This scoping review revealed the scarcity of evidence or the gap in evaluating the impact of mHealth solutions on care provider's organizational and legal aspects, as well as patients' sociocultural and safety aspects for adult patients on chronic dialysis. Importantly, the economic value of mHealth technologies is an integral component in the management of this patient population. Future research could include additional data collection to enable the intent-to-treat analysis of efficacy and cost-benefit [28]. Because limited RCTs have been completed on the topic, this review provides a necessary baseline from which additional trials with larger sample sizes in a variety of patient population analysis is warranted [46].

In addition, this review could inform a co-design process and the development of a comprehensive service model utilizing mHealth technologies for supporting home monitoring and self-management of adult patients undergoing chronic dialysis.

Limitations

Publication bias was one of the limitations of this study design, for example, non-English publications had been excluded. Consequently, relevant studies initiated in native non-English speaking countries were not captured in this review. Most of the studies included were conducted in developed countries (20/22), which may lead to limited global generalizability. Of the 22 included studies, 15 (68%) originated in North America. Regional discrepancies, clinical and social practices, as well as health care systems and policies, may not be generalizable to other regions.

Conclusions

The mHealth technologies have been used for adult patients undergoing chronic dialysis, with the main functions being nutrition or dietary self-monitoring, RBM, web-based portal, self-monitoring of in-session dialysis-specific information, and self-monitoring of lifestyle or behavioral change. On the basis of this scoping review, mHealth technologies showed neutral to positive results on patient satisfaction and acceptance, clinical effectiveness, economic assessment, HRQoL, and impact on lifestyle or behavioral change in this cohort.

Despite the potential benefits of utilizing mHealth technologies in patients undergoing dialysis, mHealth solutions have not been widely adopted and integrated into standard renal care. This review highlighted the lack of a comprehensive evaluation that includes a patient's safety and their sociocultural status, as well as a care provider's organizational, ethical, and legal aspects when assessing mHealth technologies in care of patients on dialysis. Due to sparse evidence in the literature, the clinical effectiveness and economic effects have not been adequately assessed, especially missing the long-term effects and cost-effectiveness of using mHealth technologies. More rigorous studies in this field continue to be performed, making cost-benefit evaluation a standard process to assist the decision making to an established sustainable business model and organizational support for mHealth implementation in management of patients undergoing dialysis.

Acknowledgments

The authors would like to thank Meng Zhu and Jennifer Shen, the alumni of the University of Waterloo, for their help with search strategy on gray literature and the screening of titles and abstracts.

Authors' Contributions

YY designed the study, developed the search strategy and executed literature search, analyzed data, and was the main writer of the manuscript; HC supervised the entire process, co-developed the search strategy, provided guidance in formulating inclusion and exclusion criteria, discussed findings, and contributed to the editing of the manuscript; HQ analyzed and discussed findings and edited the manuscript; PPM provided guidance on reporting findings, contributed to the analyses and Methods and Discussion section, and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keyword search strategy.
[DOCX File , 17 KB - mhealth_v8i4e15549_app1.docx]

Multimedia Appendix 2
Gray literature search methods.
[DOCX File , 13 KB - mhealth_v8i4e15549_app2.docx]

Multimedia Appendix 3
Characteristics of the included papers.
[DOCX File , 35 KB - mhealth_v8i4e15549_app3.docx]

References

1. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
2. World Health Organization. Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016.
3. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
4. Rojahn K, Laplante S, Sloand J, Main C, Ibrahim A, Wild J, et al. Remote monitoring of chronic diseases: a landscape assessment of policies in four European countries. PLoS One 2016;11(5):e0155738 [FREE Full text] [doi: 10.1371/journal.pone.0155738] [Medline: 27195764]
5. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 9;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
6. Jones KR, Lekhak N, Kaewluang N. Using mobile phones and short message service to deliver self-management interventions for chronic conditions: a meta-review. Worldviews Evid Based Nurs 2014 Apr;11(2):81-88. [doi: 10.1111/wvn.12030] [Medline: 24597522]
7. Lee Y, Cui Y, Tu M, Chen Y, Chang P. Mobile health to maintain continuity of patient-centered care for chronic kidney disease: Content analysis of apps. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e10173 [FREE Full text] [doi: 10.2196/10173] [Medline: 29678805]
8. 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]
9. Murray AM. Cognitive impairment in the aging dialysis and chronic kidney disease populations: an occult burden. Adv Chronic Kidney Dis 2008 Apr;15(2):123-132 [FREE Full text] [doi: 10.1053/j.ackd.2008.01.010] [Medline: 18334236]
10. Bieber SD, Weiner DE. Telehealth and home dialysis: a new option for patients in the United States. Clin J Am Soc Nephrol 2018 Aug 7;13(8):1288-1290 [FREE Full text] [doi: 10.2215/CJN.03010318] [Medline: 30042227]
11. Lampe K, Mäkelä M, Garrido MV, Anttila H, Autti-Rämö I, Hicks NJ, European network for Health Technology Assessment (EUnetHTA). The HTA core model: a novel method for producing and reporting health technology assessments. Int J Technol Assess Health Care 2009 Dec;25(Suppl 2):9-20. [doi: 10.1017/S0266462309990638] [Medline: 20030886]
12. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: Systematic search in app stores and evaluation. JMIR mHealth and uHealth 2019;7(9). [doi: 10.2196/12604] [Medline: 31486408]
13. 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]
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
15. Connelly K, Siek KA, Chaudry B, Jones J, Astroth K, Welch JL. An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability. J Am Med Inform Assoc 2012;19(5):705-712 [FREE Full text] [doi: 10.1136/amiajnl-2011-000732] [Medline: 22582206]
16. Welch JL, Astroth KS, Perkins SM, Johnson CS, Connelly K, Siek KA, et al. Using a mobile application to self-monitor diet and fluid intake among adults receiving hemodialysis. Res Nurs Health 2013 Jun;36(3):284-298 [FREE Full text] [doi: 10.1002/nur.21539] [Medline: 23512869]
17. Chaudhry BM, Schaefbauer C, Jelen B, Siek KA, Connelly K. Evaluation of a Food Portion Size Estimation Interface for a Varying Literacy Population. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7 - 12, 2016; California, San Jose, USA p. 5645-5657 URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85015010909&doi=10.1145%2f2858036.2858554&partnerID=40&md5=7b7fb9e78cb31f372db63c6e18d5949c [doi: 10.1145/2858036.2858554]
18. Hand RK, Leon JB, Steiber AL. Feasibility test of an online nutrition algorithm on a tablet computer versus additional patient care time in improving patient outcomes. Top Clin Nutr 2014;29(3):250-256. [doi: 10.1097/tin.0000000000000006]
19. Stark S, Snetselaar L, Piraino B, Stone RA, Kim S, Hall B, et al. Personal digital assistant-based self-monitoring adherence rates in 2 dialysis dietary intervention pilot studies: BalanceWise-HD and BalanceWise-PD. J Ren Nutr 2011 Nov;21(6):492-498 [FREE Full text] [doi: 10.1053/j.jrn.2010.10.026] [Medline: 21420316]

20. Sevick MA, Stone RA, Novak M, Piraino B, Snetselaar L, Marsh RM, et al. A PDA-based dietary self-monitoring intervention to reduce sodium intake in an in-center hemodialysis patient. Patient Prefer Adherence 2008 Feb 2;2:177-184 [FREE Full text] [Medline: 19920960]
21. Imtiaz R, Atkinson K, Guerinet J, Wilson K, Leidecker J, Zimmerman D. A pilot study of okkidney, a phosphate counting application in patients on peritoneal dialysis. Perit Dial Int 2017;37(6):613-618. [doi: 10.3747/pdi.2017.00050] [Medline: 28970367]
22. Dey V, Jones A, Spalding EM. Telehealth: acceptability, clinical interventions and quality of life in peritoneal dialysis. SAGE Open Med 2016;4:2050312116670188 [FREE Full text] [doi: 10.1177/2050312116670188] [Medline: 27757228]
23. Weinhandl ED, Collins AJ. Relative risk of home hemodialysis attrition in patients using a telehealth platform. Hemodial Int 2018 Jul;22(3):318-327. [doi: 10.1111/hdi.12621] [Medline: 29210164]
24. Lew SQ, Sikka N, Thompson C, Magnus M. Impact of remote biometric monitoring on cost and hospitalization outcomes in peritoneal dialysis. J Telemed Telecare 2019 Dec;25(10):581-586. [doi: 10.1177/1357633X18784417] [Medline: 30001668]
25. Magnus M, Sikka N, Cherian T, Lew S. Satisfaction and improvements in peritoneal dialysis outcomes associated with telehealth. Appl Clin Inform 2017 Mar 1;8(1):214-225 [FREE Full text] [doi: 10.4338/ACI-2016-09-RA-0154] [Medline: 28246673]
26. Neumann CL, Wagner F, Menne J, Brockes C, Schmidt-Weitmann S, Rieken EM, et al. Body weight telemetry is useful to reduce interdialytic weight gain in patients with end-stage renal failure on hemodialysis. Telemed J E Health 2013 Jun;19(6):480-486. [doi: 10.1089/tmj.2012.0188] [Medline: 23614336]
27. Minatodani DE, Chao PJ, Berman SJ. Home telehealth: facilitators, barriers, and impact of nurse support among high-risk dialysis patients. Telemed J E Health 2013 Aug;19(8):573-578. [doi: 10.1089/tmj.2012.0201] [Medline: 23742629]
28. Berman SJ, Wada C, Minatodani D, Halliday T, Miyamoto R, Lindo J, et al. Home-based preventative care in high-risk dialysis patients: a pilot study. Telemed J E Health 2011 May;17(4):283-287. [doi: 10.1089/tmj.2010.0169] [Medline: 21480788]
29. Karopadi AN, Antony S, Subhramanyam S, Nayak K. Remote monitoring of peritoneal dialysis: why? where? how? Hong Kong J Nephrol 2013;15(1):6-13. [doi: 10.1016/j.hkjn.2013.03.002]
30. Nayak A, Karopadi A, Antony S, Sreepada S, Nayak KS. Use of a peritoneal dialysis remote monitoring system in India. Perit Dial Int 2012;32(2):200-204 [FREE Full text] [doi: 10.3747/pdi.2011.00124] [Medline: 22383718]
31. Kiberd J, Khan U, Stockman C, Radhakrishnan A, Phillips M, Kiberd BA, et al. Effectiveness of a web-based eHealth portal for delivery of care to home dialysis patients: a single-arm pilot study. Can J Kidney Health Dis 2018;5:2054358118794415 [FREE Full text] [doi: 10.1177/2054358118794415] [Medline: 30210802]
32. Nicdao MA, Kim J, Baldacchino T, Jung YH, Mikaheal M, Patekar A, et al. 'My home Hemo' app - a new telehealth tool for Remote Monitoring of patents on home haemodialysis. Renal Soc Australas J 2016;12(2):41 [FREE Full text]
33. Hayashi A, Yamaguchi S, Waki K, Fujiu K, Hanafusa N, Nishi T, et al. Testing the feasibility and usability of a novel smartphone-based self-management support system for dialysis patients: a pilot study. JMIR Res Protoc 2017 Apr 20;6(4):e63 [FREE Full text] [doi: 10.2196/resprot.7105] [Medline: 28428168]
34. Liu N, Kim J, Jung Y, Arisy A, Nicdao MA, Mikaheal M, et al. Remote monitoring systems for chronic patients on home hemodialysis: field test of a copresence-enhanced design. JMIR Hum Factors 2017 Aug 29;4(3):e21 [FREE Full text] [doi: 10.2196/humanfactors.7078] [Medline: 28851680]
35. Harrington DM, Myers L, Eisenman K, Bhise V, Nayak K, Rosner MH. The use of a tablet computer platform to optimize the care of patients receiving peritoneal dialysis: a pilot study. Blood Purif 2014;37(4):311-315. [doi: 10.1159/000365440] [Medline: 25170838]
36. Han M, Williams S, Mendoza M, Ye X, Zhang H, Calice-Silva V, et al. Quantifying physical activity levels and sleep in hemodialysis patients using a commercially available activity tracker. Blood Purif 2016;41(1-3):194-204 [FREE Full text] [doi: 10.1159/000441314] [Medline: 26765515]
37. Havas K, Bonner A, Douglas C. Self-management support for people with chronic kidney disease: patient perspectives. J Ren Care 2016 Mar;42(1):7-14. [doi: 10.1111/jorc.12140] [Medline: 26404772]
38. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: A proof-of-concept randomized controlled trial. JMIR Res Protoc 2013 Sep 4;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
39. Donald M, Kahlon BK, Beanlands H, Straus S, Ronksley P, Herrington G, et al. Self-management interventions for adults with chronic kidney disease: a scoping review. BMJ Open 2018 Mar 22;8(3):e019814 [FREE Full text] [doi: 10.1136/bmjopen-2017-019814] [Medline: 29567848]
40. Nayak KS, Ronco C, Karopadi AN, Rosner MH. Telemedicine and remote monitoring: supporting the patient on peritoneal dialysis. Perit Dial Int 2016;36(4):362-366 [FREE Full text] [doi: 10.3747/pdi.2015.00021] [Medline: 27385806]
41. Jeddi FR, Nabovati E, Amirazodi S. Features and effects of information technology-based interventions to improve self-management in chronic kidney disease patients: A systematic review of the literature. J Med Syst 2017 Sep 18;41(11):170. [doi: 10.1007/s10916-017-0820-6] [Medline: 28920174]

42. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
43. Chen JF, Wu LL, Chou SC. Effects of Journaling Dietary Intake App on the Health Outcomes of Chronic Kidney Disease Stage 3B-5. In: Proceedings of the 2016 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS'16; January 5-8, 2016; Koloa, HI, USA p. 3379-3388. [doi: 10.1109/hicss.2016.421]
44. Lew S. Annual Dialysis Conference. 2018 Mar 3. Telehealth in Peritoneal Dialysis Patient Management URL: http://annualdialysisconference.org/wordpress/wp-content/themes/adc/2018_Handouts/Lew_Telehealth_in_PD.pdf [accessed 2020-01-22]
45. Henderson C, Knapp M, Fernández JL, Beecham J, Cartwright M, Beynon M, et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. Bmj 2013:346.
46. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]

Abbreviations

BP: blood pressure
CADTH: Canadian Agency for Drugs and Technologies in Health
CINAHL: The Cumulative Index to Nursing and Allied Health Literature
CKD: chronic kidney disease
EHR: electronic health record
ER: emergency room
HD: hemodialysis
HHD: home hemodialysis
HRQoL: health-related quality of life
HTA: health technology assessment
IWG: interdialytic weight gain
KDQOL: Kidney Disease Quality of Life
mHealth: mobile health
PDA: personal digital assistant
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PTH: parathyroid hormone
QoL: quality of life
RBM: remote biometric monitoring
RCTs: randomized controlled trials
SF-36: 36-Item Short Form Health Survey

Edited by G Eysenbach; submitted 14.11.19; peer-reviewed by D Juzwishin, D John; comments to author 23.12.19; revised version received 28.12.19; accepted 31.12.19; published 03.04.20.

Please cite as:
Yang Y, Chen H, Qazi H, Morita PP
Intervention and Evaluation of Mobile Health Technologies in Management of Patients Undergoing Chronic Dialysis: Scoping Review
JMIR Mhealth Uhealth 2020;8(4):e15549
URL: https://mhealth.jmir.org/2020/4/e15549
doi:10.2196/15549
PMID:32242823

Review

Improving Engagement Among Patients With Chronic Cardiometabolic Conditions Using mHealth: Critical Review of Reviews

Kamila Cheikh-Moussa[1*], PhD; Jose Joaquin Mira[2*], PhD; Domingo Orozco-Beltran[3*], PhD

[1]University Hospital San Juan de Alicante, The Foundation for the Promotion of Health and Biomedical Research, San Juan de Alicante, Alicante, Spain

[2]Health Psychology Departament, Miguel Hernández University, Elche, Spain

[3]Clinical Medicine Department, Miguel Hernández University, San Juan de Alicante, Alicante, Spain

[*]all authors contributed equally

Corresponding Author:
Jose Joaquin Mira, PhD
Health Psychology Departament
Miguel Hernández University
University Avenue
Edificio Altamira
Elche, 03202
Spain
Phone: 34 966658984
Email: jose.mira@umh.es

Abstract

Background: The burden imposed by cardiometabolic diseases remains a principal health care system concern. Integration of mobile health (mHealth) interventions is helpful for telemonitoring of these patients, which enables patients to be more active and take part in their treatment, while being more conscious and gaining more control over the outcomes. However, little is known about the degree to which users engage, and the extent to which this interaction matches the usage pattern for which mHealth interventions were designed.

Objective: The aim of this study was to describe the characteristics and results of studies on mHealth solutions that measured the effects of interventions with patient engagement in the context of chronic cardiometabolic diseases.

Methods: A critical review of systematic reviews was conducted to recover data on interventions focused on the engagement of patients with chronic cardiometabolic diseases using mHealth technologies. Articles (from January 1, 2010) were searched in the Medlars Online International Literature Medline (Medline/Pubmed), Embase, Cochrane Library, PsycINFO, and Scielo databases. Only studies that quantified a measure of engagement by patients with cardiometabolic disease were included for analysis. The Critical Appraisal Skills Programme (CASP) was used to determine included studies considering the quality of the data provided. The Scottish Intercollegiate Guidelines Network (SIGN) checklist was used to assess the quality of the evidence according to the methodology used in the studies reviewed. Engagement was defined as the level of patient implication or participation in self-care interventions. Engagement measures included number of logs to the website or platform, frequency of usage, number of messages exchanged, and number of tasks completed.

Results: Initially, 638 papers were retrieved after applying the inclusion and exclusion criteria. Finally, only three systematic reviews measuring engagement were included in the analysis. No reviews applying a meta-analysis approach were found. The three review articles described the results of 10 clinical trials and feasibility studies that quantified engagement and met the inclusion criteria assessed through CASP. The sample size varied between 6 and 270 individuals, who were predominantly men. Cardiac disease was the principal target in the comparison of traditional and mHealth interventions for engagement improvement. The level of patient engagement with mHealth technologies varied between 50% and 97%, and technologies incorporating smartphones with a reminder function resulted in the highest level of engagement.

Conclusions: mHealth interventions are an effective solution for improving engagement of patients with chronic cardiometabolic diseases. However, there is a need for advanced analysis and higher-quality studies focused on long-term engagement with specific

interventions. The use of smartphones with a single app that includes a reminder function appears to result in better improvement in active participation, leading to higher engagement among patients with cardiometabolic diseases.

(JMIR Mhealth Uhealth 2020;8(4):e15446) doi:10.2196/15446

KEYWORDS

mHealth; patients; telemedicine; engagement; chronic disease; cardiovascular disease; diabetes; obesity

Introduction

Background

Historically, patient engagement has been an essential factor to obtain better results and outcomes in health care interventions. A medical or health care service focuses on active participation of the patient during treatment to improve the outcome and enhance the patient's health [1-4]. Such interventions further increase the patient's awareness of taking more control over their health status, whereas traditional health care usually places the patient in a passive role during treatment. Recent studies have shown that higher patient engagement levels can be achieved using novel technological solutions such as mobile apps and e-devices [5-10]. Patient engagement is particularly relevant in cases of chronic diseases in which the outcome of the intervention largely depends on lifestyle choices and self-care capacity, along with the manner in which that patients cope with their daily lives.

Mobile Health in Chronic Cardiometabolic Diseases

Mobile health (mHealth) technologies such as the use of mobile phones and other wireless technologies in medical care can empower patients, with promising possibilities for optimizing health systems, enhancing health and care outcomes, and reducing resource consumption [11]. The use of mHealth technologies is starting to become more widespread in the case of cardiometabolic diseases (diabetes, coronary diseases, and obesity), which remain top priorities and principal concerns of all health care systems [12-14]. Indeed, recent studies have shown better clinical indicators, more healthy behaviors, and greater use of preventive behaviors among patients that are motivated by participation in their own health care [15,16].

The integration of virtual telemonitoring interventions [3,4] meets the challenge of ensuring the adequate use of health services by controlling expenditures on medication and diagnostic tests, and in the delivery of effective monitoring of patients at the same time, offering patients greater control of their conditions. These telemonitoring technologies include medical and public health practices supported by mobile devices such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [2]. These devices permit patients to make informed decisions on their care options and to have direct interaction with health providers. In addition, health providers can have exhaustive control of the symptoms, adherence to treatment, and engagement of the patients with treatment indications [6-8].

Patient Engagement

Although recent reviews have shown that mHealth technologies are both effective and growing in popularity [7,9], their effectiveness heavily relies on patient engagement, as a lack of such engagement can result in treatment failure. Patient engagement refers to "the process of building the capacity of patients, families, carriers, as well as health care providers, to ease and support the active participation of patients in their care, to enhance safety, quality and people-centeredness of health care service delivery" [17]. Despite the many definitions of patient engagement, they all share an underlying theme: the facilitation and strengthening of the role of those using services as coproducers of health, and health care policy and practice. Overall, patient engagement involves active partnership at various levels across the health care system, including direct care, organizational design and governance, and policy making, to improve health and health care.

Accordingly, the factors that may positively influence this relationship include increases in the patients' perceived benefit of interacting with the agent (by providing useful information or entertainment) [5], decreases in their perceived costs [6], increases in their perceived investment in the system, and decreases in their perceptions of viable alternatives to using the system [7]. These factors all tend to increase user commitment in continuing to participate with the agent and thereby ensures their long-term engagement [3,18,19].

Engagement has been related to the coproduction concept, which describes how patients may individually or collectively engage in the delivery of their treatments and services in partnership with doctors and other health professionals [20]. Thereby, the engagement concept considers the inclusion of patients and family members as active members of the health care team and collaborative partnerships with providers and provider organizations.

The World Health Organization declared engagement as a main factor indicating patient safety in health care, which is considered a global challenge. However, the majority of studies and policies are focused on issues related to hospital care. Therefore, such policies also need to be adapted for primary care because most health care is now offered in this setting. Recently, mHealth interventions have been proposed as a solution to improve patient use of health services, thereby increasing their participation in their own care and the safety of treatments. Studies of the Valcrònic research group demonstrated higher engagement and safer usage of treatments, along with better knowledge of the disease for patients using daily mHealth devices [8-10,18-25]. However, little is known about the degree to which patients engage and if this matches the usage pattern of mHealth interventions.

Based on this background, the engagement concept has been addressed from different viewpoints, and thus it is important to clarify its primary purpose in the management of health results.

Several review articles on the topic have led to renewed interest in the concept of engagement. Therefore, from a practical point of view, a comprehensive review or meta-analysis could offer more information about engagement usage in patients with cardiometabolic diseases considering their prevalence, high consumption of health care resources, and recent findings showing that a high level of patient engagement was related to better health care outcomes [16,26].

Toward this end, the aim of this study was to describe the characteristics and results of mHealth solutions based on systematic reviews and meta-analyses of studies that investigated and measured the effects of interventions according to patient engagement in the context of chronic cardiometabolic diseases.

Methods

Study Design

We conducted a critical review of systematic reviews and meta-analyses on studies of mHealth interventions focused on the engagement of patients with cardiometabolic diseases. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) protocol was applied [27].

Data Sources

Search Strategy

Studies were searched from the Medlars Online International Literature (Medline) database, via PubMed, Embase, The Cochrane Library, American Psychology Association (PsycINFO), and Scientific Electronic Library Online (Scielo) databases, using Medical Subject Headings (MeSH) and key terms searched in the title and abstracts using the Boolean connectors in Textbox 1.

Textbox 1. Boolean connectors used for database search.

("TREATMENT ADHERENCE AND COMPLIANCE" [title/abstract] OR "HEALTH BEHAVIOR" [title/abstract] OR "ENGAGEMENT" [title/abstract]) AND ("TELEMEDICINE" [title/abstract] OR "TELECARE" [title/abstract] OR "telehealth" [title/abstract] OR "homecare" [title/abstract] OR "telemonitoring" [title/abstract] OR "home monitoring" [title/abstract] OR "remote monitoring" [title/abstract] OR "ehealth" [title/abstract] OR "telerehabilitation" [title/abstract] OR "mobile health" [title/abstract] OR "mhealth" [title/abstract] OR "assisted living" [title/abstract] OR "technology-based" [title/abstract] OR "information technology" OR "health communication" [title/abstract] OR "internet-based" [title/abstract] OR "web-based" [title/abstract] OR "on-line" [title/abstract] OR "smartphones" [title/abstract] OR "mobile apps" [title/abstract] OR "mobile phone" OR "monitoring devices" [title/abstract])

The filters "Humans", "Meta-Analysis", "Review", and "10 years" were used to identify all relevant, peer-reviewed systematic reviews and meta-analyses of interventional studies published as of January 1, 2010.

The studies included among the retrieved reviews were also checked for compliance with inclusion criteria using the Critical Appraisal Skills Programme (CASP). A manual search of the references was performed to reduce possible publication bias and to identify undetected studies. Both systematic reviews and meta-analyses were considered regardless of the country, institution, author, or language.

Inclusion and Exclusion Criteria

Articles were eligible for inclusion in the analysis when they met the following criteria: systematic review or meta-analysis of randomized or controlled clinical trials, or a feasibility, usability, and utility (FUU) design measuring the effects of mHealth interventions on engagement of adult (>18 years old) patients with chronic cardiometabolic conditions. Studies were excluded if they assessed nonchronic diseases, cancer, respiratory disease, mental health, substance abuse, did not involve engagement measurement or telemonitoring, or assessed a pediatric population.

Type of Intervention

Reviews or meta-analyses that investigated the effectiveness of interventions applying text messages, smartphones or phones with internet, mobile apps, videos, and websites were all considered.

Outcome Measures

The main outcome assessed was patient engagement, which was measured as the number of logs to the website or platform, frequency of usage, amount of messages exchanged, and task completion.

Information Extraction

Quantitative and qualitative information on patient engagement suffering from chronic cardiometabolic conditions was extracted from systematic reviews or meta-analyses. An electronic form was developed to group the papers by the following items: review, author information, year, study design, participants, intervention, outcomes assessed and comparisons performed, pooled sizes of outcomes meta-analyzed, and the main conclusion. The availability of meta-analysis studies and the quality of evidence provided by the studies included in the systematic reviews were considered in the selection of studies for subsequent review. From this initial selection, a reconceptualization of the findings provided by systematic review studies was conducted.

Two authors assessed the relevance and adequacy of the studies (CM and JM). The selection was valid when the concordance between the two authors (kappa index) was higher than 0.80 (representing a high or very high strength of concordance). The third author was available for arbitration in case of persistent disagreements, followed by consensus among all authors.

Quality Assessment

The CASP tool was used for appraisement by two authors. Reviews that did not meet at least 6 of 10 screening items were

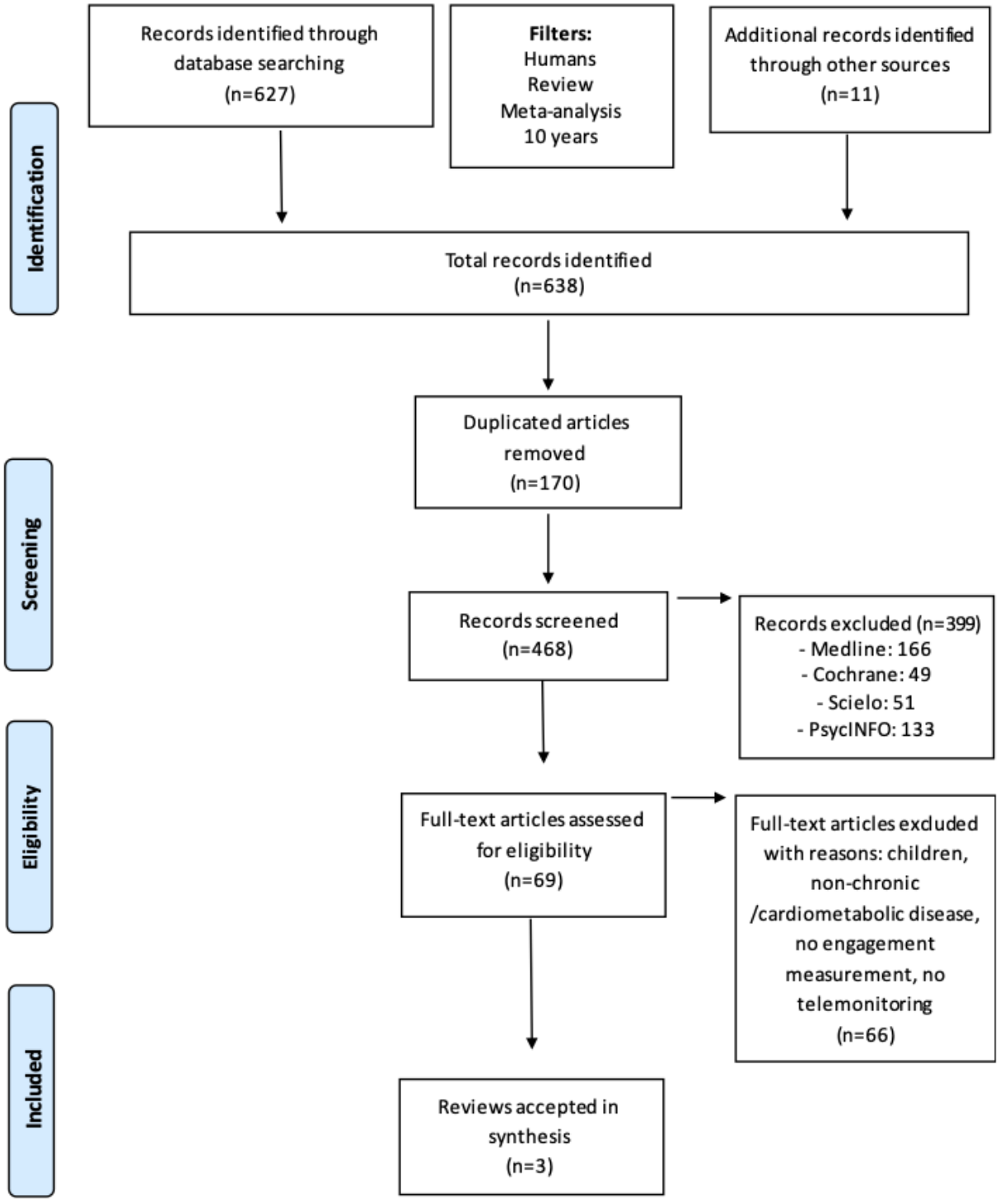
excluded. The Scottish Intercollegiate Guidelines Network (SIGN) criteria were applied to classify the quality of the evidence provided by the retrieved reviews.

Results

Retrieved Papers

The initial search returned 627 records (186 Medline, 50 Cochrane, 51 Scielo, 170 PsycINFO, and 170 Embase); no meta-analysis studies were detected. Eleven additional articles were identified from the reference lists of the studies included in the initial screen. After applying the inclusion and exclusion criteria and removal of duplicates, three systematic reviews with engagement evaluation in cardiometabolic patients that met all quality criteria were extracted and included in the final analysis [28-30]. Figure 1 summarizes the overall search strategy and article selection process.

Figure 1. Flowchart of article selection process.



The three reviews ultimately included for analysis only described the main results without providing quantitative data and with no statistical pooling analysis performed. These reviews included evidence from a total of 150 papers (93 clinical trials and 57 FUU or qualitative studies), but only 10 of these studies quantified engagement specifically. Therefore, the results of these 10 studies were extracted from the reviews and their findings were also summarized. The remaining 140 empirical studies focused on the fulfillment of tasks, rather than engagement, and did not include specific measurements of engagement.

Concordance Between Reviewers

The concordance (kappa index) between the authors identifying and extracting information from these studies was 90%, demonstrating very good agreement to reach consensus about the inclusion of articles without requiring intervention of a third author. The assessment of quality was performed using the CASP tool and returned a score of 8-10 for the accepted systematic reviews. All three of the selected systematic reviews studies were published within the last four years (2016-2018).

Study Characteristics

Table 1 summarizes the measures and interventions used in the empirical studies included in the systematic reviews analyzed. Cardiac disease was the main chronic disease of focus [28], followed by diabetes mellitus type 2 [29,30] and obesity [29]. These studies described the characteristics of different engagement tools, collected information of mHealth interventions that examined the usage of smartphones or computer self-assessment on health outcomes [28,29], and described how text messages were used as reminders to improve engagement in patients with type 2 diabetes [30].

Table 1. Characteristics of systematic reviews and patient engagement.

Review	Quality/source	Disease	mHealth type/source	Engagement measure (%)	Description
Hamilton et al [28]	SIGN[a]: ++ N[b]: 7 of 9 studies 2 RCT[c] 2 CT[d] 3 FUU[e]	Cardiac disease and heart failure	Smartphone	Outdoor walking-based exercise program Follow-up > 180 days Adherence to protocol Engagement with technology Adherence Visits (0 to 6 weeks or 4 months) Response to blood glucose reminder measure Access to core sessions Access to optional sessions	Step counter Blood pressure self-measurement and data entry Daily readings Lifestyle counseling and usual care Daily messages and tasks Educational material and videos Medication reminders Physical activity prompts Screenings and surveys Wellness diary Relaxation audio files Community care team Internet Web portal for viewing of patient data Text messages Video and telephone mentoring
Perski et al [29]	SIGN: + N: 2 of 117 studies 1 RCT 1 FUU	Diabetes Obesity	Computer-assisted self-management program with or without expert support	Visited (0 to 6 weeks) Visited (6 weeks to 4 months) Access to core sessions Access to optional sessions	Internet-based website “My Path” (“*Mi Camino*”) for self-management with or without the addition of social support from the health care team and peer group meetings. 12 weekly online sessions to control and review weight goals.
Nelson et al [30]	SIGN: + N: 1 of 24 studies 1 FUU	Type 2 diabetes	Text message reminders for blood glucose measurement	Response to blood glucose reminder measure	Three text messages per week requesting blood glucose readings and three text messages with appointment reminders before each scheduled appointment.

[a]SIGN: (-) low, (+) acceptable, (++) high quality.

[b]N: number of eligible studies.

[c]RCT: randomized controlled clinical trial.

[d]CT: controlled clinical trial.

[e]FUU: feasibility, usability, and utility.

There was substantial variation in how intervention engagement was reported across the three studies [28-30]. The FUU-based studies of the mHealth interventions delivered for cardiovascular diseases are typically analyzed based on significant improvement in the quality of life with higher rates of patient engagement for patients unable to attend traditional center-based programs [28].

Engagement with mHealth interventions rates decreased over time, and longer interventions, patients of older age, more familiarity with the use of these technologies, and lower health literacy showed progressively poor engagement [29,30]. The impact of these mHealth interventions to reduce hospital utilization was inconsistent [29].

Since all three reviews were descriptive in nature, no quantitative results were discussed, and there was no statistical pooling. These studies included only four randomized trials [31-34], one controlled trial [35], and five FUU designs [36-40] in which engagement outcomes were quantified. One of the systematic reviews [28] included six empirical studies that assessed engagement with mHealth strategies in which outcomes were compared with those in patients receiving traditional care [31,32,34,35,37,38].

Participants and Study Design

The sample size of studies included in the reviews varied between 6 and 270 individuals [33,36]. The age of the patients was between 50 and 66 years [31,37]. Men were in higher proportion in general, although one of the studies included equal numbers of men and women [33]. Most of the studies included two groups for assessment: the intervention group (mHealth, S) and the traditional control group (TC); only one study included a three-group design (Tables 2 and 3). The follow-up period varied from 3 months to 4 months. The studies used different measures of engagement, and the global percentage of adherence/engagement or daily readings was used.

Table 2. Characteristics and design quality of included reviews.

Review and included studies	Participants, n (% male)	S[a] (% male); TC[b], (% male)	Age (years); median (IQR), mean (SD), or range	Design	Evidence quality (SIGN[c])
Hamilton et al [28]					
Scherr D, 2006 [37]	20	S1, 14 (93); S2, 6 (83)	50 (14)	FUU[d]	+
Scherr D, 2009 [31]	120	S, 66 (69.5); TC, 54 (72)	66 (64-74)	RCT[e]	++
Worringham C, 2011 [36]	6	—[f]	53.6 (42-67)	FUU	+
Blasco A, 2012 [32]	203 (80)	S, 102 (81); TC, 101 (79)	60.6 (23.8)	RCT	++
Seto E, 2012 [35]	100	S, 50 (82); TC, 50 (76)	53.5 (14)	CT[g]	+
Varnfield M, 2014 [34]	94	S, 53 (91); TC, 41 (83)	52.13 (9.2)	RCT	+
Forman D, 2014 [38]	26 (77)	—	59 (43-76); 33%>65	FUU	+
Nelson et al [30]					
Fischer H, 2012 [39]	47	—	50-59	FUU	+
Perski et al [29]					
Glasgow R, 2011 [33]	270	S1, 137 (54.7); S2, 133 (48.9)	57.8 (9.3)	RCT	++
Arden-Close E, 2015 [40]	132	S1, 137 (54.7); S2, 133 (48.9)	57.8 (9.3)	FUU	+

[a]S: mHealth study group.

[b]TC: traditional control group.

[c]SIGN: (-) low, (+) acceptable, (++) high quality.

[d]FUU: feasibility, usability, utility.

[e]RCT: randomized clinical trial.

[f]Not available.

[g]CT: controlled clinical trial.

Table 3. Outcomes of engagement using mHealth interventions.

Review and included studies	Disease	Study period (months)	mHealth type	Non-mHealth	Result
Hamilton et al [28]					
Scherr D, 2006 [37]	S[a]1: CHF[b]; S2: HT[c]	3	Clinical app	TC[d] Blood pressure automatic monitor	S1: 94%; S2: 84%; follow-up >90 days
Scherr D, 2009 [31]	S/TC: CHF	6	Clinical app, email	TC	S: 95% adherence
Worringham C, 2011 [36]	S: ACS[e] TC: none	1.5	Real-time monitoring post exercise sessions, emergency mobile phone contact	none	87% of sessions (outdoor walking-based exercise program) completed
Blasco A, 2012 [32]	S/TC: CVD[f] and risk factors[g]	12	Wireless app protocol, web portal	TC	98% completed >50% of sessions; 83% completed >75% of sessions
Seto E, 2012 [35]	S: CHF	6	Clinical app, email, text messages, website	TC Telephone contact	S: 50% adherence in 80% of patients; 80% adherence in 66% of patients; 95% adherence in 32% of patients; follow-up >180 days
Varnfield M, 2014 [34]	S/TC: myocardial infarction	6	Smartphone	TC Qualitative patient and clinician feedback	S: 94% adherence; TC: 68% adherence (P<.05)
Forman D, 2014 [38]	S: CHF TC: none	1	Heart coach app	TC Qualitative patient and clinician feedback	90% daily engagement with technology
Nelson et al [30]					
Fischer H, 2012 [39]	S: diabetes II	3	Text message reminder for blood glucose measurements	None	S: 79% of users responded regularly to >50% of blood glucose reminder message prompts
Perski et al [29]					
Glasgow R, 2011 [33]	S1: diabetes II; S2: diabetes II	6	S1/S2: computer-assisted self-management program with human support	None	S1: 66% visits 0-6 weeks, 74% visits 6 weeks to 4 months (P=.14); S2: 44% visits 0-6 weeks, 51% visits 6 weeks to 4 months (P=.22)
Arden-Close E, 2015 [40]	S: BMI >30 with HT or diabetes	6-12	Web weight management intervention	None	47% access to core sessions; 47% access to optional sessions; 3% no access

[a]S: mHealth study group.

[b]CHF: congestive heart failure.

[c]HT: hypertension.

[d]TC: traditional control group.

[e]ACS: acute coronary syndrome.

[f]CVD: cardiovascular disease.

[g]Risk factors include tobacco smoking, low-density lipoprotein cholesterol≤100 mg/dL (2.6 mmol/L), hypertension, or diabetes mellitus.

Mobile Health Technologies

Internet-based technologies were predominant within studies, including the use of smartphones with health-integrated applications [21,28,32,34-37], audio-visual devices (iPod and iPod touch) [28,38], and computers [29,30,39,40] focused on patient self-management. The features of the apps and programs are described in Table 1.

The randomized clinical trial included in the review by Perski et al [29] assessed improvement of engagement with or without instructions provided for use of the mHealth tool (computer) [33], whereas Nelson et al [30] reported the results of a feasibility study in patients with type 2 diabetes who were reinforced with reminder tools such as text messages or email to regulate the measurement of glucose blood levels [39]. Hamilton et al [28] included studies assessing a principal mHealth tool that was combined with other features such as

email reminders or Web portals [31,32,35-38]. However, we categorized the study types based on the main technology used in such cases.

When the mHealth intervention was compared with a traditional clinical practice [31,32,34,35,37,38], this was based on face-to-face interventions during regular consultations and diagnosis evaluating equipment, with qualitative information on patients and clinical feedback, as described in Tables 2 and 3.

Measurement of Engagement

Studies included in the Perski et al [29] review registered the number of visits to a Web-based program or platform [34,40]. Papers extracted from Hamilton et al [28] registered task completion, whereas those included in the Nelson et al [30] review measured the interaction level with agent messages. All empirical studies included in the reviews used self-management systems; however, two of them also included support features as reminders along with a human support system [39,40]. The engagement was measured by the frequency of responses, and two studies considered the number of follow-up days [35,37].

Findings of the Studies

Overall, mHealth interventions demonstrated high engagement, with the level of patient participation varying between 50% and 97%. Clinical trials from the Hamilton et al [28] review [21,24,35] showed the highest level of engagement (95%) when the interventions were limited to one or two tools using smartphones, whereas a higher number of tools resulted in lower patient interaction. Websites and surveys resulted in less engagement with the intervention (47%-74%) [32,39,40].

Quality of Evidence

The SIGN scale determined that the three selected systematic reviews met 10 of the 12 items of evidence quality, and the 10 selected studies included in the identified reviews presented a high evidence quality level (Tables 2 and 3).

Discussion

Principal Findings

Emerging evidence suggests that mHealth promotes more engagement than traditional interventions. However, few reviews were identified that included a measure of quantitative engagement, and there was poor quality evidence owing to the lack of advanced study designs such as meta-analysis. For the included studies, the duration was not sufficiently long to compensate for the attraction potential that the innovation of these technological solutions implies. In addition, a long-term study (24 months or longer) could decrease participation. Despite the important role that participants should have in mHealth studies, few studies used a specific intervention or measurement of their engagement.

The engagement approach differed among studies, with most focused on the promotion of quality of life and patient behavior with the internet-based intervention [28,31,38]. The feasibility of these interventions was mainly examined as a solution to overcome barriers in health systems in attracting the attention to patients that are compromised by the current high costs [28,34-36]. Analysis of the causes of early dropouts has not been carried out due to incompatibility with these technological solutions (eg, in the case of patients that are technologically illiterate). Consequently, the reviews could not discuss potential solutions for enhancing the usability of mHealth interventions.

A high number of interventions focused on evaluating engagement in patients with cardiac diseases, suggesting that there is a need to approach cardiovascular disease as the leading cause of global mortality (85% due to heart attack and stroke) [41]. High heterogeneity was found for almost all variables, including sample size and sex, and only one study included proportional representation of men and women [29].

Quality of Evidence

Although the overall internal quality of the selected papers was high, a large number of papers were discarded owing to the low quality of current evidence. The clinical trials retrieved in the detected reviews indicated improvement of engagement with the use of these new strategies. In addition, the FUU studies demonstrated high participation of patients using mHealth interventions; this was reinforced by the high quality of evidence detected with the SIGN scale.

No long-term follow-up period appeared to jeopardize the aim of interventions. This is related to the extensive use of pilot projects in the mHealth sector, which may be due to limited government ownership, multiple barriers around prioritizing funding, overall cost, and acceptance by health authorities and populations [42].

Reliability and Applicability of the Evidence

There were also differences between the type of mHealth interventions, with higher engagement observed with the use of smartphones. This finding is in line with the results of a survey of the Global Observatory for eHealth in which 62% of the participating countries used mHealth tools for patient monitoring, 69% of which used mobile devices (text, voice, or multimedia reminders) [42].

The characteristics of the engagement assessment tools varied, with some focused on self-measurement and others using tasks or survey complements, with lack of support from physicians or experienced agents as a common feature. These differences reflect the complexity of engagement measurement, which depends of two distinguishable components: the directional component determined by the hedonic quality of the object, and other sources of force intensity (eg, opposition to interfering forces, regulatory fit, overcoming personal resistance) that are included in the construct labeled as “strength of engagement” [43].

Overall, we found that the use of mobile devices was related to an improvement of engagement to mHealth interventions, with a higher level of user commitment observed when a single tool was used with limited functions and with expert support. There is a need for studies with homogeneity in design, engagement measurement, and the type of mHealth tool. Further clinical trials can provide more information about the effectivity and efficiency of mobile technologies in patients with chronic

conditions. This could permit comprehensive investigations of outcomes using meta-analysis techniques.

Limitations

The few retrieved reviews is one of the principal limitations of this analysis, which was due to the lack of studies with engagement measurement. More homogeneity in the design of studies and participant characteristics could facilitate advanced statistical analyses such as systematic reviews and meta-analyses. In addition, these studies could help in the development of new tools focused on patient participation. Moreover, studies on patient social networks could provide more information on engagement.

Future Research

It is expected that active engagement has a direct impact on the reduction of health costs, but this aspect has rarely been analyzed to date. The number of dropouts or poor engagement with mHealth interventions requires more attention in further studies. Moreover, it could be of interest to study the characteristics of patients who have taken better advantage of these tools. There is also a need to detect patients that obtained relatively less benefits using mHealth interventions, which could help to better personalize the technologies or increase availability for this population.

Conclusions

Smartphones and other mobile devices with a single tool that includes reminders can improve participation and induce higher engagement with mHealth interventions in patients with chronic conditions.

Acknowledgments

This study was supported by the project Prometeo/2017/173, Valencian Ministry of Education, Research, Culture and Sports.

Authors' Contributions

All authors contributed to the development of the study. KM, JM, and DB participated equally in the study design, concept, literature search, and the review of papers. Analysis was performed by KM and supervised by JM and DB; these three authors participated equally in the writing and manuscript review.

Conflicts of Interest

None declared.

References

1. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. Patient Educ Counsel 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
2. WHO Global Observatory for eHealth. World Health Organization. Switzerland: World Health Organization; 2011. mHealth: new horizons for health through mobile technologies: second global survey on eHealth URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2019-02-18]
3. World Health Organization. World Health Organization. Geneva, Switzerland: World Health Organization; 2010. Global status report on noncommunicable diseases 2010 URL: https://www.who.int/nmh/publications/ncd_report2010/en/ [accessed 2019-02-19]
4. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
5. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov 28;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
6. Paré G, Poba-Nzaou P, Sicotte C. Home telemonitoring for chronic disease management: an economic assessment. Int J Technol Assess Health Care 2013 Apr;29(2):155-161. [doi: 10.1017/S0266462313000111] [Medline: 23514722]
7. Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]
8. Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in Primary Care for Patients With Chronic Conditions: The ValCrònic Quasi-Experimental Study. J Med Internet Res 2017 Dec 15;19(12):e400 [FREE Full text] [doi: 10.2196/jmir.7677] [Medline: 29246881]
9. 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 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
10. Sabate E. World Health Organization. Geneva, Switzerland: World Health Organization; 2003. Adherence to long-term therapies: evidence for action URL: https://www.who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2019-02-19]

11. Vo V, Auroy L, Sarradon-Eck A. Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
12. Srivastava A. Challenges in the treatment of cardiometabolic syndrome. Indian J Pharmacol 2012 Mar;44(2):155-156 [FREE Full text] [doi: 10.4103/0253-7613.93579] [Medline: 22529466]
13. Fisher M. Cardiometabolic disease: the new challenge? Pract Diab Int 2006 Apr;23(3):95-97. [doi: 10.1002/pdi.909]
14. Saljoughian M. Cardiometabolic syndrome: a global health issue. US Pharm 2016;41(2):HS19-HS21.
15. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
16. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2012 Nov 20;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
17. World Health Organization. Patient Engagement: Technical Series on Safer Primary Care. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252269/9789241511629-eng.pdf;jsessionid=2A0A830D9896E76E1437A65664BABF0C?sequence=1 [accessed 2019-02-19]
18. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
19. Rockville M, Maurer M, Dardess P, Carman K, Frazier K, Smeeding L. Agency for Healthcare Research and Quality. Rockville, MD: Prepared by American Institutes for Research under contract HHSA 290-200-600019; 2012 May. Guide to patient and family engagement: environmental scan report 2012; AHRQ Publication No. 12-0042-EF URL: https://www.ahrq.gov/sites/default/files/publications/files/ptfamilyscan.pdf [accessed 2019-02-19]
20. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul 16;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
21. Dent M, Pahor M. Patient involvement in Europe--a comparative framework. J Health Organ Manag 2015;29(5):546-555. [doi: 10.1108/JHOM-05-2015-0078] [Medline: 26222875]
22. Coulter A. Patient engagement--what works? J Ambul Care Manage 2012;35(2):80-89. [doi: 10.1097/JAC.0b013e318249e0fd] [Medline: 22415281]
23. Davis RE, Jacklin R, Sevdalis N, Vincent CA. Patient involvement in patient safety: what factors influence patient participation and engagement? Health Expect 2007 Sep;10(3):259-267 [FREE Full text] [doi: 10.1111/j.1369-7625.2007.00450.x] [Medline: 17678514]
24. Sandy LG, Tuckson RV, Stevens SL. UnitedHealthcare experience illustrates how payers can enable patient engagement. Health Aff (Millwood) 2013 Aug;32(8):1440-1445. [doi: 10.1377/hlthaff.2012.1082] [Medline: 23918489]
25. Higgins T, Larson E, Schnall R. Unraveling the meaning of patient engagement: A concept analysis. Patient Educ Couns 2017 Jan;100(1):30-36. [doi: 10.1016/j.pec.2016.09.002] [Medline: 27665500]
26. Forhan M, Risdon C, Solomon P. Contributors to patient engagement in primary health care: perceptions of patients with obesity. Prim Health Care Res Dev 2013 Oct;14(4):367-372. [doi: 10.1017/S1463423612000643] [Medline: 23237022]
27. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
28. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 07;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
29. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
30. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of User Engagement with Mobile- and Web-Delivered Self-Care Interventions for Adults with T2DM: A Review of the Literature. Curr Diab Rep 2016 Jul 2;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
31. 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]
32. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
33. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]

34. Varnfield M, Karunanithi M, Lee C, 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 27;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
35. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
36. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS One 2011 Feb 09;6(2):e14669 [FREE Full text] [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]
37. Scherr D, Zweiker R, Kollmann A, Kastner P, Schreier G, Fruhwald FM. Mobile phone-based surveillance of cardiac patients at home. J Telemed Telecare 2006 Jun 22;12(5):255-261. [doi: 10.1258/135763306777889046] [Medline: 16848939]
38. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
39. Fischer H, Moore S, Ginosar D, Davidson A, Rice-Peterson C, Durfee M, et al. Care by cell phone: text messaging for chronic disease management. Am J Manag Care 2012 Feb 01;18(2):e42-e47. [Medline: 22435883]
40. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A Visualization Tool to Analyse Usage of Web-Based Interventions: The Example of Positive Online Weight Reduction (POWeR). JMIR Hum Factors 2015 May 19;2(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
41. Mendis S, Puska P, Norrving B. Global atlas on cardiovascular disease prevention and control. Policies, strategies and interventions. Geneva: World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization; 2011. URL: https://www.who.int/cardiovascular_diseases/publications/atlas_cvd/en/ [accessed 2019-02-19]
42. World Health Organization. Global diffusion of eHealth: making universal health coverage achievable: report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2018-09-01]
43. Pham MT, Avnet T. Rethinking Regulatory Engagement Theory. J Consum Psychol 2009 Apr;19(2):115-123. [doi: 10.1016/j.jcps.2009.02.003]

Abbreviations

CASP: Critical Appraisal Skills Programme
FUU: feasibility, usability and utility
MeSH: Medical Subject Headings
Medline: Medlars Online International Literature
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Scielo: Scientific Electronic Library Online
SIGN: Scottish Intercollegiate Guidelines Network
S: mHealth study group
TC: traditional control group

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by M Anderson, GE Iyawa, F Fatehi, T Irizarry; comments to author 30.09.19; revised version received 24.10.19; accepted 26.01.20; published 08.04.20.

Please cite as:
Cheikh-Moussa K, Mira JJ, Orozco-Beltran D
Improving Engagement Among Patients With Chronic Cardiometabolic Conditions Using mHealth: Critical Review of Reviews
JMIR Mhealth Uhealth 2020;8(4):e15446
URL: https://mhealth.jmir.org/2020/4/e15446
doi:10.2196/15446
PMID:32267239

Review

Smart Glasses for Caring Situations in Complex Care Environments: Scoping Review

Charlotte Romare[1,2], CCRN, MSc; Lisa Skär[2], RN, PhD

[1]Region Blekinge, Karlskrona, Sweden

[2]Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden

Corresponding Author:
Charlotte Romare, CCRN, MSc
Region Blekinge
Intensive Care Unit
Karlskrona, 371 85
Sweden
Phone: 46 455731000
Email: charlotte.romare@regionblekinge.se

Abstract

Background: Anesthesia departments and intensive care units represent two advanced, high-tech, and complex care environments. Health care in those environments involves different types of technology to provide safe, high-quality care. Smart glasses have previously been used in different health care settings and have been suggested to assist health care professionals in numerous areas. However, smart glasses in the complex contexts of anesthesia care and intensive care are new and innovative. An overview of existing research related to these contexts is needed before implementing smart glasses into complex care environments.

Objective: The aim of this study was to highlight potential benefits and limitations with health care professionals' use of smart glasses in situations occurring in complex care environments.

Methods: A scoping review with six steps was conducted to fulfill the objective. Database searches were conducted in PubMed and Scopus; original articles about health care professionals' use of smart glasses in complex care environments and/or situations occurring in those environments were included. The searches yielded a total of 20 articles that were included in the review.

Results: Three categories were created during the qualitative content analysis: (1) smart glasses as a versatile tool that offers opportunities and challenges, (2) smart glasses entail positive and negative impacts on health care professionals, and (3) smart glasses' quality of use provides facilities and leaves room for improvement. Smart glasses were found to be both a helpful tool and a hindrance in caring situations that might occur in complex care environments. This review provides an increased understanding about different situations where smart glasses might be used by health care professionals in clinical practice in anesthesia care and intensive care; however, research about smart glasses in clinical complex care environments is limited.

Conclusions: Thoughtful implementation and improved hardware are needed to meet health care professionals' needs. New technology brings challenges; more research is required to elucidate how smart glasses affect patient safety, health care professionals, and quality of care in complex care environments.

(JMIR Mhealth Uhealth 2020;8(4):e16055) doi:10.2196/16055

KEYWORDS

anesthesia department; critical care; intensive care units; scoping review; smart glasses

Introduction

Complex Care Environments

Improvements in medical skills and technology have made health care increasingly complex [1]. Anesthesia departments and intensive care units (ICUs) represent two advanced, high-tech, and complex care environments [2,3]. In the anesthesia department, patients undergo planned or acute surgeries, treatments, or examinations. The patients are often under sedation or anesthesia, which affects vital organ functions. Specialized health care professionals are responsible for maintaining the patient's ventilation and handling changes in the homeostatic balance caused by sedation or anesthesia. Advanced technology, such as ventilators, physiological monitoring, and the anesthesia station, make this possible [2]. The most critically ill patients are admitted to the ICU. These

patients can have failure in one or more vital organ systems, such as the cardiovascular, respiratory, or renal system. Numerous examinations and treatments are performed and used, such as mechanical ventilation, bronchoscopy, dialysis, and multiple potent drugs [3]. In both ICUs and anesthesia departments, changes in the patient's condition can occur rapidly and may demand an immediate response from health care professionals to save the patient's life, hence, close surveillance is vital. Health care in these complex care environments is based on well-trained and dedicated health care professionals, teamwork, and the use of technology to provide high-quality care and ensure patient safety [2,3]. Caring situations in complex care environments include, for example, advanced medical, technological, and caring components and the surrounding specific environment. In this study, we use the expression *complex care environment* to describe all these aspects in the contexts above.

Patient Safety

Patients being cared for in complex care environments are in a vulnerable state, due to their conditions and the treatments they need. According to the World Health Organization, patient safety work aims to prevent avoidable patient harm and provide a safe health care environment. They also state that delivering safe complex care is a challenge [4]. The use of advanced technology, such as ventilators and physiological monitoring, is a prerequisite for care in anesthesia departments and ICUs. Technology is known to increase patient safety and to enhance patient care [5], but technology also imposes risks. In 2019, the Emergency Care Research Institute (ECRI) included both ventilators and physiological monitoring on their annual top-10 list of health technology hazards [6]. This imposes continuous work for patient safety in complex care environments. Patient safety work is not only related to the use of technology. In complex situations, several factors interact; patient safety work is also related to other aspects, for example, working conditions and routines [4,7]. Health care professionals in complex care environments incorporate several factors into their surveillance during patient care in order to provide safe care [8]. Through close surveillance, health care professionals can support both the physical and emotional needs of the patient, to protect the patent from suffering and harm. This promotes a patient-safe way of working [9], as do proper implementation and use of new technology [7]. It is also important for new technology to add value to patient care and to bring desired consequences [10].

Smart Glasses

Smart glasses are a product suggested to aid health care professionals in numerous areas, such as surgery, accessing electronic health records, remote instructions, and education [11-13]. They are a computing device worn as a pair of glasses, which presents information within the user's field of view through a prism. Smart glasses are a platform for apps and can display text and images, use a camera, and communicate via Bluetooth and Wi-Fi. The user interacts with the smart glasses through physical input or voice commands [14]. Smart glasses can send and receive information online, or through local area networks, and the information can be displayed in the prism. Smart glasses can also be used to communicate by voice or video and to capture pictures or video. The uses for smart glasses depend on the apps in the device; tailored apps provide the possibility for multiple purposes. The most well-known brand of smart glasses today is Google Glass, which was introduced to the market in 2013.

Smart Glasses in Complex Care Environments

According to our literature search, a few reviews have been published about smart glasses in surgical and nonsurgical settings. Different areas of use are described, such as to provide visualization during laparoscopy, to broadcast live surgery to medical students, to take pictures and record videos to facilitate medical documentation, to record encounters with patients, and to use as a navigational tool to maintain attention to the operative field [11-13,15-19]. The idea of head-mounted and hands-free equipment as an aid in anesthesia departments is not new [20]; health care professionals have shown interest in, and have seen the potential for, smart glasses in intensive care [21]. Since there is a growing interest in smart glasses and since technology might have an effect on patient safety, it is important to conduct a scoping review on smart glasses used by health care professionals in complex care environments in order to identify the knowledge and experiences in this field. In our study we use the term *health care professionals* to describe physicians with different levels of experience and training, registered nurses, specialized nurses, and other professionals working closely with patients, such as assistants. To our knowledge, only a few studies have been performed within our area of interest [22,23]. This indicates that the use of smart glasses in complex care environments is a new and evolving area, making it even more important to investigate. This field is novel, innovative, and has been found to have potential to improve both patient care and patient safety in other health care settings [16,19]. The aim of this study was, therefore, to highlight potential benefits and limitations with health care professionals' use of smart glasses in situations occurring in complex care environments.

Methods

Design

A scoping review was chosen as the methodology of this study since it addresses broad research questions and is advocated for new areas [24-26]. This review followed all six stages suggested by Arksey and O'Malley [25] and the methodological development by Levac et al [27]. Results are reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews (PRISMA-ScR) in order to increase methodological transparency [26].

Data Search and Selection

Step 1 was to identify the research question; a scoping review approach has been suggested [25,27] and our path toward the final aim was described in the Introduction. Step 2 was to identify relevant articles. Before we began the database searches, we consulted experienced librarians who assisted in choosing the most appropriate databases and search terms as suggested by Arksey and O'Malley [25]. PubMed and Scopus were chosen

to cover research within both health care and engineering. Search terms were also discussed among the authors and with other researchers within the fields of nursing, medicine, and engineering. New search terms were added several times during the process. The final search terms are presented in Multimedia Appendix 1. As stated initially, only two articles were identified during our initial searches [22,23], which made us broaden our searches to include vital signs monitoring, alarm management, and patient safety, since these are important aspects of care in complex care environments. This resulted in one additional article from a clinical setting [28] and two from simulated settings [29,30]. We also found articles about isolated events occurring in complex care environments, such as electrocardiogram (ECG) reading and cardiopulmonary resuscitation (CPR). We did not actively search for specific isolated events using those words as search terms in the database searches, but we did include articles found during our searches.

The database searches are presented in Multimedia Appendix 2.

Additional articles were identified using reference lists and research networks (ie, ResearchGate and Academia). The deadline for searching databases was set to April 2018, and the deadline for searching other sources was set to December 2018. Step 3 was the study selection of the scoping process [25,27]. The inclusion and exclusion criteria are presented in Textbox 1. We made no limitations on the publication date, since smart glasses are a new product.

Titles in the search results list were screened first, followed by the abstracts, if needed, in order to identify relevant articles. The full-text articles were obtained and read if they seemed eligible for this review. Screening was performed by the first author (CR) and discussed among the authors. A flowchart of the search process, similar to a PRISMA 2009 flow diagram [31], is presented in Figure 1.

Textbox 1. Inclusion and exclusion criteria for publications.

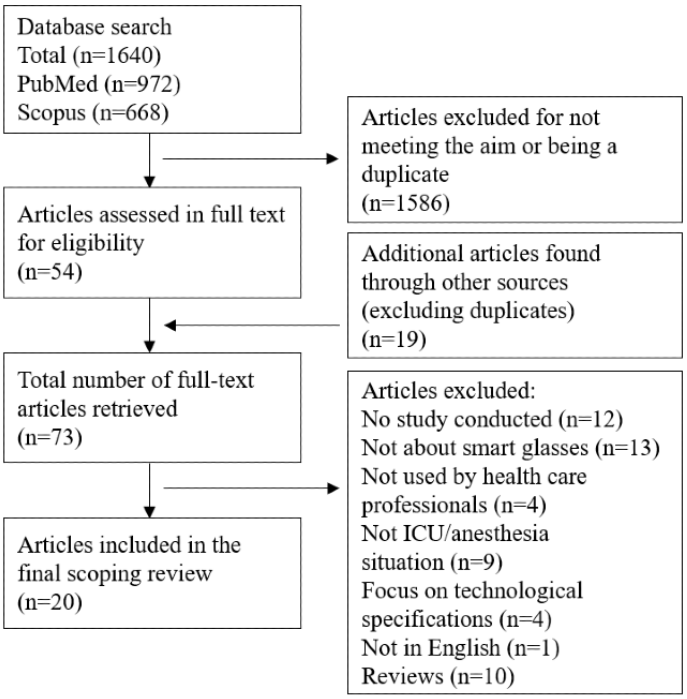
Inclusion criteria

- Smart glasses used in complex care environment
- Smart glasses used in situation occurring in complex care environment
- Smart glasses used by health care professionals
- Written in English

Exclusion criteria

- Smart glasses used by students
- Smart glasses used by patients
- Review article

Figure 1. Flowchart of search.



Data Summary and Analysis

Step 4 involved charting the data to gain an overview. This charting is presented in Multimedia Appendix 3. During step 5, the results were collated, summarized, and reported [25]. This can be a challenging process and it is recommended to divide step 5 into three parts: analysis, reporting results, and considering the overall implications of the results. A qualitative content analysis is recommended [27], hence, we chose to follow Polit and Beck's [24] description of this process. Meaning units meeting the aim of this study were marked in the included articles and condensed while still retaining the core content. Notes about context were added to the condensed units. The condensed units were then continuously numbered, labelled with a code, sorted into subcategories, and then sorted into categories; hence, analysis was on a manifest level [24]. Neither the analysis nor the scoping process occurred in a one-way direction but went back and forth between the steps as more knowledge was obtained. Step 6—the last step—in the scoping process was to enable practitioners and consumers to contribute to the work. The results of this scoping review have been presented to, and discussed with, engineers, a physician, registered nurses, and nurses specialized in intensive care and anesthesiology.

Results

Overview

The aim of this study was to highlight potential benefits and limitations of health care professionals' use of smart glasses in situations occurring in complex care environments. A total of 20 articles [22,23,28-30,32-46] were found eligible for our scoping review and were included in the content analysis (see Multimedia Appendix 3). These included research articles (16/20, 80%), conference articles (2/20, 10%), a case report (1/20, 5%), and a correspondence (1/20, 5%). The included articles originated from the United States (12/20, 60%), European countries (7/20, 35%), and Australia (1/20, 5%) and were published in a variety of scientific journals and conference proceedings. One article was published in 2012; the rest were published between 2014 and 2018. A majority of the articles were from simulated or laboratory settings (11/20, 55%) and 1 article out of 20 (5%) was conducted in both a simulated and clinical setting; both qualitative and quantitative designs were used. During analysis, three categories were created: (1) *Smart glasses as a versatile tool that offers opportunities and challenges*, (2) *Smart glasses entail positive and negative impacts on health care professionals*, and (3) *Smart glasses' quality of use provides facilities and leaves room for improvement*.

Smart Glasses as a Versatile Tool That Offers Opportunities and Challenges

Smart glasses were found to be used in several situations occurring in complex care environments, including in daily practice [22,32], for vital signs monitoring [28-30], for consultation and assessment [33,34], for CPR evaluation [35,36], for documentation (ie, verbal, photo, and video) [37-40], and for viewing medical images [41-46].

Smart glasses were found to be easy to use in procedural settings [30]. Procedures were performed correctly [45,46] and with equivalent technique, both with and without smart glasses [29,35]. Increased time for completing tasks was noted when using smart glasses [37,46]. When using smart glasses for vital signs monitoring, abnormal signs were noted earlier than with traditional monitoring [29,30]; smart glasses were found suitable for this purpose [28,30] and increased awareness of vital signs [29,30]. Even though smart glasses made it easier to monitor vital signs, especially if working alone, health care professionals did not feel that smart glasses could replace the traditional monitor [30]. Presenting vital signs in smart glasses made uninterrupted monitoring possible, even when engaged in other activities [23,32] or at a remote location [23].

Smart glasses provided the possibility to share information with colleagues [22,32,37,40]. In some cases, visual media from smart glasses caused the remote consultants to change the management plan for patients. The remote consultants mostly gained confidence in the management plans and found the visual media helpful [33]. Assessing patients remotely through smart glasses showed high agreement with on-site investigators, with assessment of pupil size as the least correlating parameter [34]. Gaining expert help through smart glasses' audio-video link during CPR was found to be helpful and reassuring. Technique and management improved, but CPR was sometimes interrupted, both because of the instructions given and by discussions with the remote expert [35]. Research found that smart glasses were eligible for educational purposes as well as for accessing patient medical records [32] and databases [22]. Furthermore, smart glasses were used to read patient barcodes for patient identification in order to increase patient safety [38]. The possibility of increasing patient safety through smart glasses was also mentioned by others [30,32].

Smart glasses could facilitate documentation, although text from voice recognition in a medical context needed improvement [22,38]; poor audio quality was seen as a contributing factor. Text was usually recognized when health care professionals talked clearly and slowly [22]. A context-specific vocabulary was suggested and the ability to review and edit the text was seen as necessary [38]. The default setting for both audio [38] and video recordings needed to be longer in order to be useful in clinical practice [22,38]. Smart glasses were found to be easy to use for video recordings [36] and provided good quality [22,39]. Video from smart glasses was rated better than video from a standard video camera. Both visibility and audibility were equivalent to that of an on-site observer. Health care professionals stated that they would be uncomfortable recording an actual event [36], and some were worried that they would be filmed unknowingly [22]. There was some discrepancy between what the user saw and what was recorded. In order to capture the right area of interest, the user had to angle the head [22,35] to an uncomfortable position [22]. Difficulties in capturing the correct area of interest were also noted when taking photos with smart glasses; this and a decrease in sharpness were the main differences between smart glasses and an ordinary digital camera [37]. Others found no difference between photographing with smart glasses and an ordinary digital camera, but they preferred to preview photos on a larger

screen than in smart glasses [38]. Some aspects of complex care were seen to be best documented by a photo [39] that could easily be captured by smart glasses, hands-free and without assistance [37-39]. Since smart glasses lack the ability to zoom, health care professionals sometimes had to come closer to the photo object than they preferred [22,37].

The small size of the smart glasses' display caused dissatisfaction when working with medical images, as did the lack of zoom [41]. Health care professionals found it difficult to notice subtle findings in medical images [22]. To improve the concept, high-quality images were requested [41] and the provided ability to zoom and pan was appreciated by users [44]. Health care professionals were not confident in their interpretations of medical images in smart glasses [41], and interpretations were less correct when performed in smart glasses than when performed traditionally [41,43,44]. When interpreting streamed ECG in smart glasses, no difference was noted from standard conditions regarding noticing different rhythms [42].

Smart Glasses Entail Positive and Negative Impacts on Health Care Professionals

Smart glasses were described as new tools for health care professionals in the included articles. Health care professionals felt unfamiliar with smart glasses [23,36] and noted that there was a learning curve [30,44,45]. If the smart glasses' camera was used improperly, the quality of images was affected [38]; practice [22] or training courses [40] were suggested. Health care professionals' general impressions of smart glasses were positive [22,23,30,40] and they stated that they would like to use smart glasses again [23,29,46]. Health care professionals did not feel interrupted or disturbed by smart glasses during procedures or patient management [22,23,30,33,37,39]. No objective or subjective nervousness or anxiety were found [29], although some health care professionals did feel distracted by the smart glasses [29,35,36], which affected their performance negatively [29,35]. Increased focus on, and quality of, the task performed using smart glasses were noted as a positive aspect; however, on the negative side, it was difficult to talk to the patient and to the smart glasses at the same time [40]. When using smart glasses during procedures, health care professionals gained increased focus on the procedural field, and ergonomics improved since they did not have to turn their heads to view monitors [30,45,46]. Health care professionals did, however, spend more time looking at the smart glasses display than they did at a traditional ultrasound screen [46]. Smart glasses were found to be comfortable to wear [23,29,34,35,37,46]. Some users who wore prescription glasses found it difficult to combine these with smart glasses [23,29,35], while others did not have this issue [36]. On smart glasses where the prism was fixed to the right eye, left-handed users reported discomfort [29]. Health care professionals reported eye strain and fatigue after using smart glasses [23,36,42]. Some health care professionals did not find it problematic to use smart glasses the whole day, while others found it infeasible [40].

Smart Glasses' Quality of Use Provides Facilities and Leaves Room for Improvement

This category involves aspects of technical performance, navigation, and hardware. The quality of photos and videos captured by smart glasses was positively evaluated [22,33,35,37,38,43-45], although photos from an ordinary digital camera received higher ratings [37]. With adequate lighting, no photos were over- or underexposed [38]. Smart glasses had no flash, which led to decreased photo quality in low-light environments, and overexposure occurred with overhead operating lamps [22]. Furthermore, the absence of the ability to zoom affected the possibility of getting the correct area in focus for the photos [22,37]. The smart glasses display was easily seen [22,23] and the contrast improved if the background was dark. During videoconferencing, small letters were not legible [22] and the display was considered too small to provide all details on medical images, such as radiographs [43]. When communicating with others using smart glasses, the room needed to be quiet for good audibility [22].

Wi-Fi and/or Bluetooth were used for data transmission. Smart glasses were able to connect to Wi-Fi and Bluetooth without problems [22], but issues with Wi-Fi coverage were noted [23,34,38]. During data transmission, stuttering, cutoffs, and delays occurred [22,23,34,46]. Data saved in smart glasses were automatically uploaded to a cloud server when smart glasses were charged and connected to Wi-Fi. This could be avoided by connecting smart glasses to a computer prior to charging in order to transfer and delete data without uploading it to the server [22].

Smart glasses could be controlled by voice or physical input, such as using a touch pad, eyeblinks, or head movements. Controlling smart glasses through a temple touch pad was found to be easy and intuitive. In sterile environments [22] and when hands were busy or contaminated [38], hands-free handling was found useful. Voice control worked well in both silent and busy environments [37], but problems with voice control for video recordings were reported [33,38]. Features for controlling smart glasses by gestures were tested and appreciated by users [22,37,38] but did not always work well in practice, due to unintentional input [22,38].

Issues were raised about the limited battery life of smart glasses [22,29,36,46]. When recording video or using teleconferencing, the battery lasted 30-40 minutes and otherwise up to 10 hours [22]. Smart glasses were also found to produce a noticeable amount of heat [28,36,46]. When used clinically, smart glasses can be equipped with splatter eye protection [22] and were disinfected using disinfecting wipes [37] or by wiping with 70% isopropanol [22].

Discussion

Overview

This scoping review shows that smart glasses have both benefits and limitations in complex care environments. Increased understanding is provided about different situations where smart glasses might be used by health care professionals in clinical practice in anesthesia departments and ICUs. Research about

smart glasses in clinical complex care environments is limited; several of the included studies were conducted in simulated settings or were minor clinical studies. The results also show that smart glasses could affect health care professionals and their performance, both positively (eg, through increased focus on procedural fields) and negatively (eg, causing discomfort during use). The quality of use of smart glasses is highlighted and there are some concerns that need attention before implementing the use of smart glasses in clinical complex care. This is all useful knowledge in the process of implementing smart glasses in anesthesia departments and ICUs.

Principal Results and Comparison With Prior Work

In complex care environments, technology is a prerequisite for the advanced care conducted. When health care professionals feel confident with equipment, complex care can be carried out in a safe way [47]. This review shows that patient management deteriorated if health care professionals became disturbed by the smart glasses. The results further highlight that there was a learning curve associated with the use of smart glasses. This indicates that user training is crucial when introducing smart glasses into complex care environments in order to maintain high-quality care and patient safety. The same is true for prudent implementation of any new technology. Both of these aspects have been discussed in relation to ICUs [21] and anesthesia departments seem to adhere to this as well. Generally, when implementing new technology, health care professionals need to see a clear benefit with the new device [48], and implementation in complex care environments does not seem to be an exception. To ensure patient privacy and patient safety, this review shows that ethical issues also need to be taken into consideration before implementing the use of smart glasses into complex care environments, as well as in other contexts [13,15,17,49]. Information security and privacy are well-known issues when implementing eHealth solutions in health care [50], and cybersecurity is at the top of ECRI's annual list of patient safety risks for 2019 [6]. Research regarding cybersecurity in health care has increased over the last 20 years, but there are still gaps to fill [51]. An extensive review about ethical issues related to smart glasses states that data security and privacy were the most frequently highlighted features found in the research [52]. Smart glasses as a new platform also imposes new ethical challenges related to privacy (eg, it is impossible for patients and health care professionals to know if or when they are being recorded or photographed by smart glasses). This makes context-specific development, implementation, and user routines important from an ethical view in order to provide patient safety [52]. Both intended and unintended consequences of new technology, such as smart glasses, need to be taken into account in the process of implementation [10], for example, in complex care environments.

The results show that smart glasses are versatile tools that could be used for several situations occurring in complex care environments. Patient vital signs are one important part of surveillance in complex care environments that are used to detect early changes in patients' conditions that might need urgent and immediate attention. This review shows that smart glasses presenting vital signs made health care professionals detect abnormal vital signs faster. The results also reveal that health care professionals did not have to turn their heads away from the patients in order to view monitors. This has been shown earlier in complex care environments with other types of more cumbersome head-mounted displays [53-55] and with smart glasses in surgical settings [17]. Not having to turn one's head away from the procedural field has been suggested to increase patient safety [53]. Further, the results show that smart glasses provided the possibility for uninterrupted monitoring when health care professionals needed to leave the traditional monitor out of sight. This is in line with earlier research conducted in surgical settings [49] and has been seen as a valuable asset for increased patient safety [21].

This review indicates that infrastructure, smart glasses' performance, and health care professionals will affect the usability of smart glasses in complex care environments. Infrastructure, such as Wi-Fi and streaming, is a prerequisite for clinical use of most new technology [50], including smart glasses, and has been found to be a limitation in both surgical and nonsurgical settings [15,18]. This review found that complex care environments are no exception. Other technical limitations, including battery life and heat generation, were found in this review and are well known [13,15,17,49]. Technical improvements have been made recently [56], but no research was found using new, improved smart glasses. This review shows that the quality of photos and video captured by smart glasses seems to be sufficient for most clinical uses in complex care environments, but not for interpreting medical images with subtle findings. This has been concluded in the past for surgical settings [19], although other reviews have found photo and video quality to be a clinical limitation in various settings [13,15]. In complex care environments, monitoring vital signs in real time is one area of use for smart glasses, and this review found no negative results regarding image quality or the ability to detect abnormalities when smart glasses were used for viewing this kind of information.

This review shows that research about smart glasses in clinical complex care environments is limited. The results from this review can provide valuable knowledge to meet the growing interest from health care professionals, product developers, and researchers concerning smart glasses and their possible implementation in complex care environments.

Methodological Considerations and Limitations

Since a scoping review aims to conduct a wide rather than in-depth synthesis of research [25], PubMed and Scopus were chosen to search for articles from both health care and engineering. Perhaps more articles would have been found if more databases had been used, but PubMed and Scopus are big databases with wide coverage and were found sufficient. The search and screening processes were performed by the first author (CR) with support from experienced librarians. It is possible that relevant articles were missed and that these would have been found if this process had been performed by more than one researcher [27]. After reading the obtained full-text articles, inclusions and exclusions were discussed among the authors until consensus was reached. In the initial searches, "Google Glass" was found as a keyword, hence we chose to add this phrase as a search term. If we had added other brand

names, we might have found additional articles, but Google Glass is the most well-known brand of smart glasses. Articles focusing on surgeons' use of smart glasses were not actively searched for; however, those found using our search terms were included if the inclusion criteria were met. The surgeons' focus was assumed to be mainly on the surgical field, but they often work closely with an anesthesiologist or a nurse anesthetist. It is possible that issues such as team communication or other applicable information were addressed in articles with a surgical focus, which could have added to our study as well. Reviews can be accepted in a scoping review, but we chose to exclude them since there is a risk of bias when interpreting other researchers' interpretations. *Gray literature*, such as dissertations and books, can also be included in a scoping review [25]. We found some gray literature during our searches (eg, correspondence and nonscientific articles) and they were included if inclusion criteria were met. Furthermore, a scoping review does not seek to assess the quality or impact of the results from the articles included [24,25]. This is why no quality assessment was made during the inclusion process. Articles with both qualitative and quantitative designs were included in our study, and text from results and tables were included in the analysis. The analysis process was discussed among the authors to increase credibility. The sixth optional stage in the scoping process (ie, step 6), where practitioners and consumers were included [25], added value to the study through creative discussions and input. After conducting this study, the authors conclude that a scoping review was suitable to fulfill the research objective.

Conclusions

Smart glasses were found to be both a helpful tool and a hindrance in caring situations that might occur in complex care environments. Thoughtful implementation and improved hardware are needed to meet health care professionals' needs. It has been stated earlier that all new technology brings new errors, and that new technologies should be tested before widespread implementation [7]. New technology might also bring ethical challenges [52]. Therefore, we conclude that more research is required to elucidate how smart glasses affect patient safety, health care professionals, and quality of care in complex care environments.

Acknowledgments

We would like to thank the librarians at Blekinge Institute of Technology for their valuable knowledge regarding database searches. We would also like to thank the PhD students and senior lecturers at Blekinge Institute of Technology for fruitful discussions and valuable input. This research was funded by the Scientific Council at Region Blekinge. The funder did not influence any part of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final search terms.
[DOCX File , 14 KB - mhealth_v8i4e16055_app1.docx]

Multimedia Appendix 2
Presentation of database searches.
[DOCX File , 39 KB - mhealth_v8i4e16055_app2.docx]

Multimedia Appendix 3
Charting of articles included in the qualitative content analysis.
[DOCX File , 32 KB - mhealth_v8i4e16055_app3.docx]

References

1. Aroch R, Berggren L, Edberg K, Kock Redfors M, Lifbom A, Malmberg A, et al. SFAI. Malmö, Sweden: Swedish Society for Anaesthesia and Intensive Care (SFAI); 2012 Nov 16. Treatment strategies for life support in intensive care URL: https://sfai.se/wp-content/uploads/files/22-2%20Treatment_strategy.pdf [accessed 2020-03-04]
2. Nagelhout JJ, Elisha S. Nurse Anesthesia. 6th edition. St. Louis, MO: Elsevier; 2017.
3. Morton PG, Fontaine DK. Critical Care Nursing: A Holistic Approach. 11th edition. Philadelphia, PA: Wolters Kluwer; 2018.
4. Patient Safety: Making Health Care Safer. Geneva, Switzerland: World Health Organization; 2017. URL: https://tinyurl.com/uj7lyfy [accessed 2020-03-04]
5. Laila SA, Ahmed NTM, Mogahed MMA. Nurses' Perception Regarding the Use of Technological Equipment in the Critical Care Units. Journal of American Science 2011;7(10):545-552 [FREE Full text] [doi: 10.7537/marsjas071011.67]
6. 2019 Top 10 Health Technology Hazards: Executive Brief. A Report from Health Devices. Plymouth Meeting, PA: ECRI Institute; 2018. URL: https://www.ecri.org/Resources/Whitepapers_and_reports/Haz_19.pdf [accessed 2020-03-10]

7. Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
8. Pfrimmer DM, Johnson MR, Guthmiller ML, Lehman JL, Ernste VK, Rhudy LM. Surveillance: A Nursing Intervention for Improving Patient Safety in Critical Care Environment. Dimens Crit Care Nurs 2017;36(1):45-52. [doi: 10.1097/dcc.0000000000000217]
9. Nilsson U, Jaensson M. Anesthetic nursing: Keep in touch, watch over, and be one step ahead. J Perianesth Nurs 2016 Dec;31(6):550-551. [doi: 10.1016/j.jopan.2016.09.005] [Medline: 27931710]
10. Wüller H, Garthaus M, Remmers H. Augmented reality in nursing: Designing a framework for a technology assessment. Stud Health Technol Inform 2017;245:823-827. [Medline: 29295213]
11. 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 [FREE Full text] [doi: 10.1177/0141076816663560] [Medline: 27729595]
12. Klein GO, Singh K, von Heideken J. Smart glasses: A new tool in medicine. Stud Health Technol Inform 2015;216:901. [Medline: 26262203]
13. Mitrasinovic S, Camacho E, Trivedi N, Logan J, Campbell C, Zilinyi R, et al. Clinical and surgical applications of smart glasses. Technol Health Care 2015;23(4):381-401. [doi: 10.3233/THC-150910] [Medline: 26409906]
14. Google Help. Tech specs URL: https://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354 [accessed 2018-09-13]
15. Dougherty B, Badawy SM. Using Google Glass in nonsurgical medical settings: Systematic review. JMIR Mhealth Uhealth 2017 Oct 19;5(10):e159 [FREE Full text] [doi: 10.2196/mhealth.8671] [Medline: 29051136]
16. Wrzesinska N. MEDtube. 2015 Dec. The use of smart glasses in healthcare − review URL: https://medtube.net/science/wp-content/uploads/2017/03/The-use-of-smart-glasses-in-healthcare-%E2%80%93-review.pdf [accessed 2020-03-10]
17. Kolodzey L, Grantcharov PD, Rivas H, Schijven MP, Grantcharov TP. Wearable technology in the operating room: A systematic review. BMJ Innov 2016 Dec 21;3(1):55-63. [doi: 10.1136/bmjinnov-2016-000133]
18. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google Glass in surgical settings: Systematic review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
19. 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;135(3):918-928. [doi: 10.1097/prs.0000000000001056]
20. Block FE, Yablok DO, McDonald JS. Clinical evaluation of the 'head-up' display of anesthesia data. Preliminary communication. Int J Clin Monit Comput 1995 Feb;12(1):21-24. [doi: 10.1007/bf01142442] [Medline: 7782663]
21. Romare C, Hass U, Skär L. Healthcare professionals' views of smart glasses in intensive care: A qualitative study. Intensive Crit Care Nurs 2018 Apr;45:66-71 [FREE Full text] [doi: 10.1016/j.iccn.2017.11.006] [Medline: 29291922]
22. 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]
23. Drake-Brockman TF, Datta A, von Ungern-Sternberg BS. 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]
24. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. 10th edition. Philadelphia, PA: Wolters Kluwer; 2016.
25. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
27. Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
28. Fernando S, Wenjin W, Kirenko I, de Haan G, Bambang-Oetomo S, Corporaal H, et al. Feasibility of contactless pulse rate monitoring of neonates using Google Glass. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH'15). Brussels, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering); 2015 Presented at: 5th EAI International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH'15); October 14-16, 2015; London, UK p. 198-201 URL: https://doi.org/10.4108/eai.14-10-2015.2261589 [doi: 10.4108/eai.14-10-2015.2261589]
29. Iqbal MH, Aydin A, Lowdon A, Ahmed HI, Muir GH, Khan MS, et al. The effectiveness of Google Glass as a vital signs monitor in surgery: A simulation study. Int J Surg 2016 Dec;36(Pt A):293-297 [FREE Full text] [doi: 10.1016/j.ijsu.2016.11.013] [Medline: 27833004]
30. Liebert CA, Zayed MA, Aalami O, Tran J, Lau JN. Novel use of Google Glass for procedural wireless vital sign monitoring. Surg Innov 2016 Aug;23(4):366-373. [doi: 10.1177/1553350616630142] [Medline: 26848138]
31. PRISMA. URL: http://www.prisma-statement.org [accessed 2019-10-12]
32. García-Cruz E, Bretonnet A, Alcaraz A. Testing smart glasses in urology: Clinical and surgical potential applications. Actas Urol Esp 2018 Apr;42(3):207-211. [doi: 10.1016/j.acuro.2017.06.007] [Medline: 29037757]

33. Gupta S, Boehme J, Manser K, Dewar J, Miller A, Siddiqui G, et al. Does wearable medical technology with video recording capability add value to on-call surgical evaluations? Surg Innov 2016 Oct;23(5):498-504. [doi: 10.1177/1553350616656278] [Medline: 27335083]
34. Skolnik AB, Chai PR, Dameff C, Gerkin R, Monas J, Padilla-Jones A, et al. Teletoxicology: Patient assessment using wearable audiovisual streaming technology. J Med Toxicol 2016 Dec;12(4):358-364 [FREE Full text] [doi: 10.1007/s13181-016-0567-3] [Medline: 27381429]
35. Drummond D, Arnaud C, Guedj R, Duguet A, de Suremain N, Petit A. Google Glass for residents dealing with pediatric cardiopulmonary arrest: A randomized, controlled, simulation-based study. Pediatr Crit Care Med 2017 Feb;18(2):120-127. [doi: 10.1097/PCC.0000000000000977] [Medline: 28165347]
36. Kassutto SM, Kayser JB, Kerlin MP, Upton M, Lipschik G, Epstein AJ, et al. Google Glass video capture of cardiopulmonary resuscitation events: A pilot simulation study. J Grad Med Educ 2017 Dec;9(6):748-754 [FREE Full text] [doi: 10.4300/JGME-D-17-00155.1] [Medline: 29270266]
37. Albrecht U, von Jan U, 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]
38. 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]
39. 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]
40. Wüller H, Behrens J, Klinker K, Wiesche M, Krcmar H, Remmers H. Smart glasses in nursing − Situation change and further usages exemplified on a wound care application. In: Studies in Health Technology and Informatics Series. Volume 253: German Medical Data Sciences: A Learning Healthcare System. Amsterdam, the Netherlands: IOS Press; 2018:191-195.
41. 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]
42. Schaer R, Salamin F, Jiménez del Toro OA, Atzori M, Müller H, Widmer A. Live ECG readings using Google Glass in emergency situations. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy p. 315-318. [doi: 10.1109/EMBC.2015.7318363]
43. Spaedy E, Christakopoulos GE, Tarar MN, 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]
44. Stetler J, Resendes E, Martinez-Parachini JR, Patel K, Amsavelu S, Tarar MN, 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]
45. Udani AD, Harrison TK, Howard SK, Kim TE, Brock-Utne JG, Gaba DM, et al. Preliminary study of ergonomic behavior during simulated ultrasound-guided regional anesthesia using a head-mounted display. J Ultrasound Med 2012 Aug;31(8):1277-1280. [doi: 10.7863/jum.2012.31.8.1277] [Medline: 22837293]
46. 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]
47. Tunlind A, Granström J, Engström Å. Nursing care in a high-technological environment: Experiences of critical care nurses. Intensive Crit Care Nurs 2015 Apr;31(2):116-123. [doi: 10.1016/j.iccn.2014.07.005] [Medline: 25442241]
48. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
49. Vorraber W, Voessner S, Stark G, Neubacher D, DeMello S, Bair A. Medical applications of near-eye display devices: An exploratory study. Int J Surg 2014 Dec;12(12):1266-1272 [FREE Full text] [doi: 10.1016/j.ijsu.2014.09.014] [Medline: 25462704]
50. Ahmed B, Dannhauser T, Philip N. A systematic review of reviews to identify key research opportunities within the field of eHealth implementation. J Telemed Telecare 2019 Jun;25(5):276-285. [doi: 10.1177/1357633X18768601] [Medline: 29703127]
51. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. Health care and cybersecurity: Bibliometric analysis of the literature. J Med Internet Res 2019 Feb 15;21(2):e12644 [FREE Full text] [doi: 10.2196/12644] [Medline: 30767908]
52. Hofmann B, Haustein D, Landeweerd L. Smart glasses: Exposing and elucidating the ethical issues. Sci Eng Ethics 2017 Jun;23(3):701-721. [doi: 10.1007/s11948-016-9792-z] [Medline: 27432401]
53. Ormerod DF, Ross B, Naluai-Cecchini A. Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. In: Proceedings of the 6th International Symposium on Wearable

Computers (ISWC 2002). New York, NY: IEEE; 2002 Presented at: 6th International Symposium on Wearable Computers (ISWC 2002); October 7-10, 2002; Seattle, WA p. 131-132. [doi: 10.1109/ISWC.2002.1167229]
54. Liu D, Jenkins SA, Sanderson PM, Fabian P, Russell WJ. Monitoring with head-mounted displays in general anesthesia. Anesth Analg 2010;110(4):1032-1038. [doi: 10.1213/ane.0b013e3181d3e647]
55. Liu D, Sanderson PM, Jenkins SA, Watson MO, Russell WJ. Patient monitoring in anesthesia with a head-mounted display: Simulator studies and a clinical trial. In: Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society. 2009 Presented at: 53rd Annual Meeting of the Human Factors and Ergonomics Society; October 19-23, 2009; San Antonio, TX p. 733-737. [doi: 10.1177/154193120905301121]
56. Stein S. CNET. 2017 Jul 18. Google Glass returns: This time, it's professional URL: https://www.cnet.com/news/google-glass-2-goes-for-enterprise/ [accessed 2019-03-27]

Abbreviations

CPR: cardiopulmonary resuscitation
ECG: electrocardiogram
ECRI: Emergency Care Research Institute
ICU: intensive care unit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews

Edited by G Eysenbach; submitted 31.08.19; peer-reviewed by S Almerud, S Badawy; comments to author 03.10.19; revised version received 27.11.19; accepted 06.02.20; published 20.04.20.

Please cite as:
Romare C, Skär L
Smart Glasses for Caring Situations in Complex Care Environments: Scoping Review
JMIR Mhealth Uhealth 2020;8(4):e16055
URL: http://mhealth.jmir.org/2020/4/e16055/
doi:10.2196/16055
PMID:32310144

Review

The Acceptability and Effectiveness of Web-Based Developmental Surveillance Programs: Rapid Review

Jess Baker[1], BSc, MSc, PhD; Jane Kohlhoff[2], BA, DPhil, PhD; Se-Inyenede Onobrakpor[3], MBBS, MHM; Sue Woolfenden[4], PhD; Rebecca Smith[5], BSc; Constanze Knebel[6], MD; Valsamma Eapen[1], PhD, MBBS, FRCP

[1]The University of New South Wales, Liverpool, Australia
[2]The University of New South Wales, Carramarr, Australia
[3]The Ingham Institute, Liverpool, Australia
[4]The University of New South Wales, Sydney, Australia
[5]South Eastern Sydney Local Health District, Sydney, Australia
[6]Westmead Children's Hospital, Westmead, Australia

Corresponding Author:
Jess Baker, BSc, MSc, PhD
The University of New South Wales
L1 Mental Health Centre
Liverpool Hospital
Liverpool, NSW, 2170
Australia
Phone: 61 2 9616 4205
Email: jessica.baker@unsw.edu.au

Abstract

Background: Web-based developmental surveillance programs may be an innovative solution to improving the early detection of childhood developmental difficulties, especially within disadvantaged populations.

Objective: This review aimed to identify the acceptability and effectiveness of web-based developmental surveillance programs for children aged 0 to 6 years.

Methods: A total of 6 databases and gray literature were searched using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses–informed protocol. Data extraction included variables related to health equity.

Results: In total, 20 studies were identified. Most papers implemented web-based versions of the Modified Checklist for Autism in Toddlers, Revised with Follow-Up screener for autism spectrum disorder or Parent Evaluation of Developmental Status screeners for broad developmental delay. Caregivers and practitioners indicated a preference for web-based screeners, primarily for user-friendliness, improved follow-up accuracy, time, and training efficiencies.

Conclusions: Although evidence is limited as to the necessity of web- versus face-to-face–based developmental screening, there are clear efficiencies in its use.

Trial Registration: PROSPERO CRD42019127894; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=127894

(JMIR Mhealth Uhealth 2020;8(4):e16085) doi:10.2196/16085

KEYWORDS

public health surveillance; mass screening; developmental disabilities; neurodevelopmental disorders; review literature as topic; health care disparities

Introduction

Background

Healthy early child development is an important predictor of emotional and physical well-being and school attainment, whereas (neuro)developmental difficulties such as communication or motor skills disorders and autism spectrum disorders (ASD) can hinder the fulfillment of optimal health and schooling [1]. The earlier such difficulties can be detected as an instigator of timely intervention, the better the outcome [2-6]. However, many developmental difficulties are not detected until school age [7]. Globally, it is estimated that approximately 43% or 250 million children younger than 5

years are at risk of not reaching their developmental potential [8]. Furthermore, there is evidence to suggest that an *inverse care law* is in operation, with children from disadvantaged backgrounds being least likely to engage with early detection initiatives, despite being at the highest risk of developmental delay [9-13].

Population health is largely defined by social determinants. Facilitators of disadvantage include living in a medically underserved area, being of an ethnic minority background, and having a lower socioeconomic status and fewer educational opportunities [14]. Policy bodies recommend universal developmental surveillance for all children younger than 5 years and targeted developmental surveillance for children in disadvantaged populations as best practice in early detection and breaking health inequities, respectively [8].

Developmental surveillance is defined as a flexible, longitudinal, and cumulative process whereby health professionals identify children who may be at risk of developmental delays via clinical interviews and observations. It also often involves the use of a standardized screening instrument at routine times in the first several years of a child's life [15]. The use of standardized instruments to identify and refine a developmental risk or potential issue that emerges from surveillance is termed developmental screening [16]. Screening provides greater sensitivity and specificity than surveillance. It does not yield a diagnosis per se but rather is a process by which a child's development may be identified as atypical compared with similar-aged children. This is different to developmental trackers or records. Child development apps or smart diaries, such as babyTRACKS (babyTRACKS team) and BabySteps (Beyond Blue & Queensland University of Technology Institute of Health and Biomedical Innovation), assist with caregiver monitoring but do not guarantee action on the part of the caregiver or the pivotal connection with a health professional should any developmental issue be identified [17,18]. Moreover, they typically only involve the motivated few who download the app. For example, Australian studies reported inconsistent use or follow-up of *The Blue Book*, a government resource distributed to all parents at their child's birth to assist with monitoring their child's development [19,20].

Developmental surveillance and screening are costly endeavors limited by the existing service infrastructure and resources. The increasing popularity and ownership of mobile technologies offer potential economies of scale, as it may be easier to perform surveillance for a larger number of people [21]. This improved reach, accompanied by possible clinical efficiencies in reduced human error and time, and improved cross-disciplinary communication and referral uptake via automated systems may be especially beneficial among disadvantaged communities where health knowledge and access to and engagement with services may be lowest [22-25].

Distinct from cellular telephony, mobile technologies with internet access include smartphones and ultraportable computers or tablet PCs, such as iPhones or iPads. Evidence regarding the implementation of web-based health interventions in maternal and child health is mixed [26]. Successful implementation depends on the usability of the intervention to both the deliverer (ie, the health professional) and the recipient (ie, the patient). If an intervention, or in this context a screening tool, is considered usable by the patient, then they are more likely to adhere to treatment recommendation and benefit from improved clinical outcomes. If a screening tool is considered usable by the health practitioner, it is more likely to be implemented as intended [27]. Usability can be thought of as encompassing 2 components [28]. The subjective component refers to users' satisfaction with or acceptability of the tool, such as the perceived usefulness and ease of use. This is almost a necessary but not sufficient condition for the objective component, which pertains to the effectiveness of a tool.

Objectives

This review aimed to (1) systematically search for and identify existing web-based programs that implement a developmental surveillance or screening tool and (2) appraise them for their usability (ie, acceptability and effectiveness) in the early detection of developmental delay in infants and preschool children, with particular consideration given to sample demographics implicated in health disadvantage.

Methods

The review was registered with PROSPERO (CRD42019127894). The search strategy and inclusion and exclusion criteria were based on the following Participants; Intervention; Comparators; Outcome definition.

Participants

The review considered studies that included children aged 0 to 6 years (ie, infants aged 0-2 years and preschool children aged 2-6 years).

Intervention

The review included developmental surveillance or screening tools and methods that had the aim of assisting a health professional in identifying children at risk of developmental delay. The review focused on such practices delivered via the internet. It excluded cellular or telehealth, electronic health records, data management systems, or general computer or video use—unless they were used as part of a web-based system. Similarly, interactive computer systems designed to assist health professionals in making a diagnostic decision were only included if they had reached the application (ie, not design/algorithm) phase and were being implemented over a web-based system. Smart or robotic assessment toys were also excluded unless they were part of a web-based screening assessment.

Comparators

Studies need not have had a control group for inclusion in the review. Comparators, if present, included paper versions of the web-based tool.

Outcome

The review considered outcomes that reported the following:

- Information on the existence of a web-based developmental surveillance/screening tool or program, inclusive of study protocols, abstract proceedings, and dissertations.

- Original data pertaining to the user satisfaction with or acceptability of a web-based developmental surveillance/screening program, as a standalone tool or in comparison to a paper version of the cited tool.
- Original data pertaining to the effectiveness or ability of a given intervention to detect the risk of early (neuro)developmental delay. This included predictive specificity/sensitivity analyses, relative to the target diagnosis or as compared with a paper-based version of the tool. This included the risk of ASDs, communication disorders, developmental disabilities, and motor skills disorders. It excluded tools focused exclusively on detecting mental health disorders, such as attention deficit hyperactivity disorder, conduct or oppositional defiance disorders, or Tourette syndrome, or parent-child relationship issues, such as separation anxiety or reactive attachment disorder. The review excluded outcome variables related to medical disorders or syndromes such as low birthweight or preterm cardiac issues and single gene disorders such as cystic fibrosis and Down syndrome.

Information Sources

Studies were identified through the following methods:

- Electronic databases were systematically searched: MEDLINE (Medical Literature Analysis and Retrieval System Online; Ovid platform), EMBASE (Excerpta Medica database; Ovid platform), EmCare (Ovid platform), CINAHL (Cumulative Index of Nursing and Allied Health Literature; EBSCO platform), PsycINFO (Ovid platform), and Cochrane Library (Wiley platform).
- Reference lists of included studies were checked.
- Expert consultation: experts in the developmental surveillance field were consulted to identify other articles for possible inclusion in the review.

Search Strategy

The MEDLINE search strategy was developed using an iterative process of preliminary searches testing MeSH (Medical Subject Headings) and keyword search terms under the guidance of an academic librarian. New search terms were incorporated, as relevant papers were identified. The final MEDLINE search strategy is provided in Textbox 1. Once this strategy was finalized, it was adjusted to the subject headings, syntax, and operating systems of the other databases. The search was conducted in February 2019. No date or country limits were imposed upon the search (the date was set *organically* by the search for web-based programs), and only studies published in the English were included. Nonoriginal data, such as reviews, letters, opinions, or narratives, were excluded. A gray literature search for unpublished studies also included a Google search and a Google Scholar search with the key terms "(mobile OR electronic OR smart) in various combinations with ("developmental surveillance" OR developmental screening)." The first 100 sources of each search were reviewed.

Textbox 1. MEDLINE search strategy.

(exp Neurodevelopmental disorders/ or exp language development disorders/ or Developmental disabilit*.tw. or Neurodevelopmental Disorder*.tw. or developmental status.tw. or developmental milestone*.tw. or developmental delay*.tw. or autism.tw. or autistic.tw. or delayed development.tw. or development disorder*.tw. or language delay*.tw. or language disorder*.tw. or speech disorder*.tw. or communication disorder*.tw. or Aspergers.tw. or language development.tw. or Developmental coordination disorder.tw. or Developmental dyspraxia.tw. or Delayed speech.tw. or child development.tw. or language impairment.tw. or abnormal development.tw. or developmental disorder*.tw. or speech delay.tw. or language disabilit*.tw. or delayed speech.tw. or learning disabiliti*.tw. or intellectual disabilit*.tw. or learning problem*.tw. or learning difficult*.tw. or learning deficit.tw. or learning impairment*.tw. or psychomotor-delay.tw or psychomotor-disorder.tw or psychomotor-impairment.tw or delayed-psychomotor.tw or motor-skills-disorder*.tw or motor-skills-delay.tw)

AND

(exp mobile applications/ or exp Decision Making, Computer-Assisted/ or mhealth.tw. or mobile.tw. or ehealth.tw. or smart.tw. or web-based.tw. or multimedia.tw. or computer-assisted.tw)

AND

(exp public health surveillance/ or exp Mass screening/ or exp early diagnosis/ or screening.tw. or surveillance.tw. or diagnosis.tw. or early identification.tw or early detection.tw)

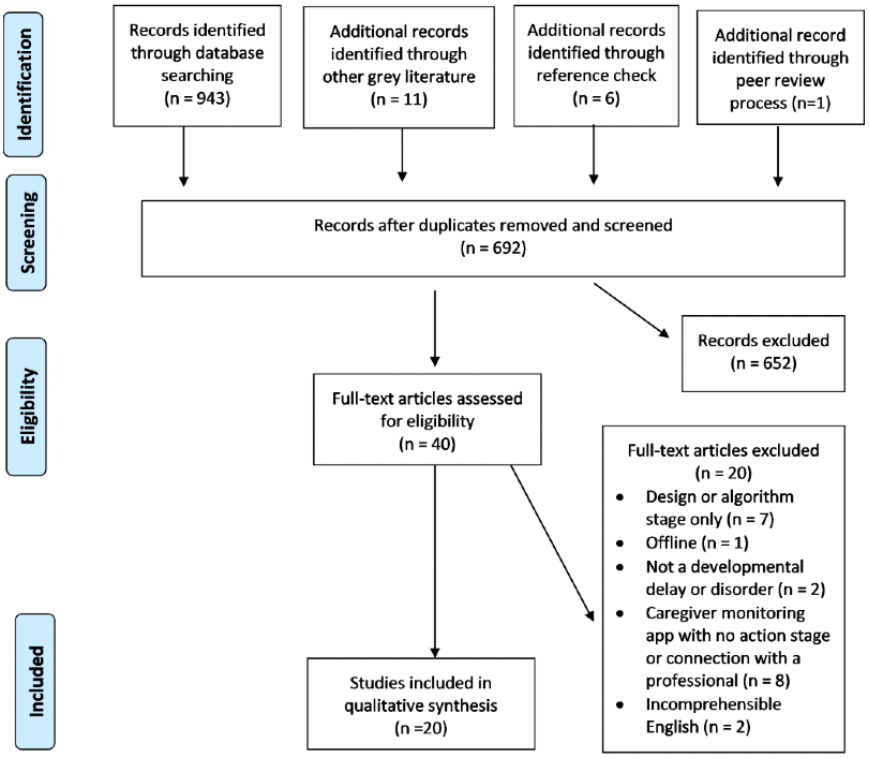
AND

(exp infant/ or exp child, preschool/ or exp child/ or neonatal.tw .or child.tw. or infant.tw. or toddler.tw or preschool*.tw. or paediatric*.tw or pediatric*.tw. or nursery.tw. or kindergarten.tw.)

Study Selection

Literature search citations were collated in a reference management software Endnote X8, and duplicate citations were removed. Authors JB and JK independently screened the titles and abstracts to determine whether a study met the general inclusion criteria. Each article was rated as include, exclude, or unclear. The full text of all articles classified as include or unclear was retrieved for formal review. Next, authors JB and JK independently assessed the full text of each study according to the predetermined inclusion criteria. Disagreements were resolved by discussion between the 2 reviewers and third author adjudication. The reasons for excluding studies were recorded. Review authors were not blind to the journal titles or authorship information of the studies. The search results are presented in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Assessment of Methodological Quality

Author JB assessed the methodological quality of included data-generating studies (ie, not abstracts or conference proceedings describing a device) using the Effective Public Health Practice Project (EPHPP) tool, which is recommended for quality assessment in systematic reviews inclusive of uncontrolled studies [29]. Paper quality was rated as *strong*, *moderate*, or *weak*, based on an overall assessment of study selection bias, design, blinding, control of confounders, data collection tools, and withdrawals or dropouts [30,31]. Author SO double coded the paper quality, with rater differences reconciled through discussion.

Data Extraction

Study data were extracted using standard forms based on the Cochrane Effective Practice and Organization of Care Review Group. Data items extracted included study design, setting, study population (including number, age, and gender of participants), the device, disorder or type of delay being detected, and study outcomes relevant to the review question. Extraction also included social determinants identified by the PROGRESS-Plus acronym that impact disadvantage [14]. These were Place of residence, Race/ethnicity/language, Occupation, Gender, Religion, Education, Socioeconomic status, and Social capital. The *Plus* refers to contextual factors such as (1) personal characteristics associated with discrimination (eg, age and disability), (2) features of relationships (eg, smoking parents), and (3) time-dependent relationships (eg, leaving the hospital or other instances where a person may be temporarily at a disadvantage) [14]. Data extraction forms were first piloted and amended as necessary. Author JB performed the initial extraction on the included studies, which was then reviewed and refined by the author SO.

Data Synthesis

A small selection of studies was anticipated; thus, a narrative approach was used to synthesize the results in relation to what web-based developmental screening programs exist and how acceptable and effective they are. The synthesis also considered social determinants related to disadvantage [32].

Results

Identified Studies

The search identified 20 eligible studies (Figure 1). Most web-based developmental surveillance programs were trials of online versions of established questionnaires, which based on automated risk scoring linked the child to a health practitioner for assessment either in person or by remote video. The gray literature and reference list searches identified several caregiver monitoring apps [17,18]. Most did not meet surveillance/screening criteria of embodying action to connect the user with a health professional should the developmental risk be apparent. For this reason, they were excluded at the full-text stage. The gray literature and reference list search also identified the prototype designs of smart toy screeners for ASD. However, except for 1 study, these were at the design stage or not yet web based [33]. All studies, except 1, were of adequate quality to be included in the review. A summary of the identified web-based measures is provided in Table 1. A total of 4 studies were at the design or protocol stage, and acceptability or effectiveness data were available in the remaining 15 studies.

The data extraction table is available in Multimedia Appendix 1.

Table 1. A summary of identified web-based developmental surveillance/screening tools, including the targeted developmental delay and the availability of acceptability or effectiveness data.

Web-based screener	Delay/disorder	Number of papers	Design/protocol stage only	Acceptability data	Effectiveness data
M-CHAT-R/F[a] (including Austim Barta)	ASD[b]	9[c]	No	Yes	Yes
Parent Evaluation of Developmental Status tools	Broad	3[c]	No	Yes	Yes
Child Health and Development Interactive System (including M-CHAT and Ages and Stages Questionnaire)	ASD/broad	1	No	Yes	No
Taipei II Pre-schooler Developmental Checklist	Broad	2	No	Yes	Yes
Cognoa	ASD	1	No	Yes	Yes
Mobile Autism Risk Assessment	ASD	1	No	No	Yes
Naturalistic Observation Diagnostic Assessment	ASD	1	No	Yes	No
Gades System (Denver Developmental Screening Test)	Language	1	Yes	No	No
Pirate Adventure Autism Assessment app	ASD	1	Yes	No	No
Smart Autism/Autism Express (including M-CHAT-R/F and Childhood Autism Spectrum Test)	ASD	1	Yes	No	No
Smart Toy Tower	Psychomotor	1	Yes	No	No

[a]M-CHAT-R/F: Modified Checklist for Autism in Toddlers, Revised with Follow-Up.

[b]ASD: Autism spectrum disorders.

[c]Glascoe [34] reported on both measures.

Data Quality of Identified Studies

Accommodating for design limitations of the 20 studies, the overall modal quality was *moderate*, as based on the EPHPP tool. The study samples typically appeared representative of their target population, in that most studies were conducted during routine visits across public health clinics. However, some selection bias was introduced in studies conducted in academic clinics. Overall, there was an absence of data across studies pertaining to the percentage of selected individuals who agreed to participate. With the exception of 2 studies that implemented a controlled cross-over design [35,36], most studies were of a one-time assessment or cohort nature, which limited overall study design quality. Furthermore, caregivers or health professionals performing the assessment were necessarily unblinded to the *intervention* status because of the salient web versus paper format, or information about blinding was not reported. Assessment of components related to attrition rates or control of confounders was largely rated as *not applicable* because of most studies being of a one-time assessment or cohort design.

What Web-Based Developmental Surveillance Programs Exist?

Web-Based Modified Checklist for Autism in Toddlers, Revised With Follow-Up

The most consistently used web-based developmental surveillance measure was the digital Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F) [37]. The M-CHAT-R/F is a two-stage screening tool designed to identify children aged 16 to 30 months who should receive an assessment for possible early signs of ASD or developmental delay. Parents score an initial set of 20 yes/no questions. Depending on the score, their child is deemed as *low*, *medium*, or *high* risk. Medium-risk children complete the second follow-up questionnaire, and the responses to this determine whether the child is referred for diagnostic evaluation or no immediate action is required. High-risk children bypass the follow-up questionnaire and are referred immediately for diagnostic evaluation. The web-based M-CHAT-R/F automatically directs parents to the appropriate follow-up questions, and health practitioners can access the results immediately.

The review also identified a paper on the web-based Child Health and Development Interactive System (CHADIS), a *package* of several screening instruments assessing a range of developmental, behavioral, and socioemotional issues [38]. The CHADIS administered the M-CHAT to children aged 0 to 3 years. A web-based M-CHAT-R/F was also used in a stepped developmental screening approach—in that it was implemented after the completion of a web-based broadband developmental screen [34].

A total of 4 web-based M-CHAT-R/F studies were conducted in America, with non-English–speaking participants typically excluded [39-42]. These studies did, however, report diverse cultural samples, including a significant proportion of Medicaid-insured families (a government program for persons whose income and resources are insufficient to pay for health care), across different socioeconomic statuses and urban and rural practices. Two identified conference proceedings reported trialing a smart M-CHAT-R/F app in Bangladesh with pictorial representations [43,44], whereas another trialed a web-based M-CHAT-R/F in a *low-risk* population in Sicily [45].

The Parents' Evaluation of Developmental Status and Parents' Evaluation of Developmental Status: Developmental Milestones

Three identified studies used the broad-screen developmental Parents' Evaluation of Developmental Status (PEDS) tools for children aged 0 to 8 years [34,46,47]. The PEDS elicits parents' concerns about their child with 10 open questions, such as Do you have any concerns about how your child understands what you say? and Do you have any concerns about how your child behaves? The PEDS: Developmental Milestones is a 6- to 8-item screener that assesses children's skills across expressive language, receptive language, fine motor, gross motor, social emotional, self-help, and academic domains. The tools are available through a web-based portal, *PEDStestonlin*e. Automated scoring generates referral letters and take-home parent summary reports and identifies appropriate billing codes. A portal is also available through which parents can complete measures before their child's health visit, and findings are sent to each clinic or provider without parents first seeing the results. The *PEDStestonline* also includes an option to deliver the MCHAT-R electronically.

The review identified 1 large-scale American implementation study of the *PEDStest online*, in which participating families were described as having *elevated psychosocial risk factors* [34]. The study reported that compared with census data, the sample was disproportionately poor, of an ethnic minority, non-English speaking, and had lower than average high school graduation rates. The review also identified the PEDS tools as a smartphone app for trial in South Africa. The trial was set in government Baby Wellness Clinics, and non-English–speaking caregivers were excluded [46]. Another South African trial of the application targeted vulnerable families who had children infected or affected by HIV/AIDS. The net monthly income of the majority of these families was less than US $155, and most households included more than 3 children [47].

Other Autism Spectrum Disorder Developmental Screening Tools

Most other identified web-based developmental screening tools were for ASD. One study described the development and evaluation of a *Naturalistic Observation Diagnostic Assessment (NODA)*. NODA is a web-based smartphone app that allows parents to collect in-home videos of their 2- to 6-year-old children's behavior to support practitioners in completing a diagnostic assessment of ASD [48]. The paper described an associated web portal that permits practitioners to direct the in-home video collection, access the child's developmental history, and conduct a remote diagnostic assessment by linking behaviors tagged in the videos to diagnostic criteria. Another study compared the effectiveness of a web-based screening tool called *Cognoa* for children aged 18 to 72 months, which integrated a 15-item parent report with the clinical ratings of brief video segments uploaded via parent's smartphones to calculate ASD risk based on an automated score [49].

Bardhan described the framework of a cloud-based mobile app for virtual ASD screening, named *Smart Autism* in one paper and *Autism E*xpress in another [44,50]. It incorporated the questions of 3 different ASD screening tools for children aged 0 to 17 years. The tools relevant to the review target age were the M-CHAT-R/F and the Childhood Autism Spectrum Test, a 39-item, yes or no evaluation aimed at parents of children aged 3 to 11 years [51]. The caregiver enters their child's birthdate, and the app selects the appropriate questionnaire. If ASD is suspected based on the automated scoring of the caregiver's responses, a video stream is sent to the user's device to play in front of the child. The camera within the mobile device records the child's reactions and uploads the video to the cloud for an expert to observe. If the expert suspects that the child might have ASD, the nearest autism resource center (ARC) is notified, and the ARC address is sent to the user. The expert observes the child in person at the ARC and provides their diagnostic decision in the cloud.

Two other studies described web-based ASD screeners for an age range inclusive of, but broader than, the study focus. One tested the sensitivity and specificity of a new Mobile Autism Risk Assessment (MARA) for children aged 16 months to 17 years. The MARA was a 7-item parent report of a child's communication, social skills, and behaviors, to triage those at the highest risk of ASD. It is completed electronically on an iPad, computer, or any other device connected to the internet and is automatically scored [52]. In the trial of MARA, non-English–speaking caregivers were excluded. The design of a Pirate Adventure Autism Assessment app was also identified in the review. The app adapted well-established affect recognition and Theory of Mind tests (ie, Smarties Tube and Sally Ann tests) to diagnose ASD via an engaging pirate adventure story line for children aged 6 to 7 years [53].

A Developmental Language Delay Screening Tool

One identified paper detailed the development of a web-based clinical decision support solution, termed the Gades System, to support the efficient detection of language disorders among children aged 0 to 6 years in routine visits to pediatricians in primary care. It evaluated sensory reception, speech perception,

speech production, and pragmatic language and had foundations in the established *Denver Developmental Screening Test* [54]. Negative responses to questions termed *Alert Milestones* recommend a follow-up visit to the pediatrician within 3 months to reevaluate the level of language acquisition, whereas negative responses to questions categorized as *Alarm Milestones* suggest a direct referral to a specialist in language disorders.

A Psychomotor Development Delay Screening Tool

The review also identified a *smart toy* screener for detecting children's psychomotor delays through natural interaction with toys. Specifically, the paper described the design, development, and validation of a tower with 5 stackable cubes embedded with a sensor data collector module [55]. As toddlers aged 23 to 37 months made the tower, sensors in the cubes sent data to a collector module through a wireless connection for the automatic detection of psychomotor developmental delays.

General Developmental Screening Tools

Two identified studies described the development and validation of a web-based version of the Taiwanese government–established *Taipei Pre-schooler Developmental Checklist second version* (Taipei II) for early detection of developmental delay [35,36]. It provides 11- to 13-item checklists for 13 age groups from 4 months to 6 years, related to easily observable behaviors or skills in the domains of motor, cognition, language/communication, and emotion/social development. The Taipei II specifies the use of pictures instead of text to avert literacy barriers.

The web-based CHADIS screening assessment for developmental, behavioral, and socioemotional issues also included the *Ages and Stages Questionnaire* (*ASQ)* for children aged 0 to 3 years [38]. The parent-reported ASQ assesses a child's personal, social, gross motor, fine motor, problem-solving, and communication skills [56].

How Acceptable Are Existing Web-Based Developmental Surveillance Programs?

Implementation Preferences

Most trials of the web-based screeners for ASD or broad developmental delay were conducted out of pediatric or general practitioner community clinics, public health centers, or university specialty clinics and incorporated into routine baby wellness or child checks [34-42,45,46,49,52,53]. One of the South African PEDS tool studies was incorporated into a home-based health visitor service [47]. The Smart Toy Tower or Gades System screeners for psychomotor or language delay were based in nursery schools [54,55].

For studies beyond the protocol stage, namely, the digital M-CHAT-R/F and PEDStestonline papers, implementation typically took the format of parents completing the screener on a provided netbook, computer, or iPad in the clinic waiting room before meeting a health practitioner for a routine child checkup. The results were immediately provided to the practitioner to guide their action during the family's appointment [39-42]. An additional implementation method for the PEDStestonline was to give the parents an appointment reminder to complete the online screening assessment before the next scheduled visit using a link provided [34].

In an evaluation of the implementation of the PEDS tools across 79 clinics that frequently used the screeners (online and offline), 24 clinics used the web-based parent portal version and collectively screened 2086 of the total 20,941 (9.96%) eligible children [34]. Many clinics were private practices, and parents accessing the portal were significantly more likely to be English speaking. The clinics with the highest web-based portal uptake provided computers in the waiting room and had attendants assist parents with using the computers, read the questions aloud for families with limited literacy, or entertain the children so that parents could complete the screens undisturbed [34]. The clinic with the lower rates of portal use utilized the appointment reminder approach. The review also identified a capacity to train community health workers in South Africa in the use of a smartphone app of the PEDS tools [46,47].

The CHADIS assessment incorporating the M-CHAT and ASQ was completed by parents 2 weeks in advance of a child welfare visit. The assessment results were communicated to the practitioner and used to determine what type of visit the family pursued. Healthy development scores determined an *electronic visit (e-visit)* by email exchange, with or without a brief face-to-face encounter with a practitioner for a physical health checkup [38]. Scores that indicated a concern necessitated an *extended encounter* with a health professional.

Uptake Efficiency

In trials of the M-CHAT-R/F, web-based screening was reported to be an efficient and acceptable method over paper-based screening. For example, there was a 59% increase in the number of toddlers screened per month when web-based screening was introduced at an urban pediatric clinic [39]. The web-based method also resulted in only 3% of missing data at the follow-up stage, compared with 35% of missing data utilizing the paper-based method [39]. Similarly, the implementation of a smart M-CHAT-R/F at a primary care clinic occasioned a 38% increase in the accurate documentation of screening results (from 54% to 92%), and a 60% increase in appropriate action for children screening positive for a delay (from 25% to 85%) [42]. When continued use of the paper form did occur, it was reported to be primarily when multiple patients arrived and there were insufficient computers to screen simultaneously [41].

Practitioner Preferences

More than 90% of the participating physicians indicated that they preferred the web-based method of the M-CHAT-R/F over paper forms [41]. They agreed that it improved their clinical assessment of ASD risk and that the automatically generated score made the M-CHAT-R/F easier and faster to use. Furthermore, primary care practitioners were shown to successfully administer the web-based M-CHAT/RF after a 10-min interactive multimedia demonstration [42]. For the CHADIS, of which the M-CHAT and ASQ were a part, the sample of 7 providers believed that the online format helped them to focus on the family visits and that they would continue to use the web-based system [38]. For Cognoa, the app that integrated parent report and video ratings for ASD screening,

clinicians reported the video component to be the most helpful. In addition, 88% of physicians found the automated summary report helpful, and 76% of physicians found that the report and videos alerted them to look for particular behaviors [49]. Clinicians also rated the NODA as clinically useful in performing an ASD diagnosis [48].

Caregiver Satisfaction

Higher rates of parental satisfaction were also reported with the iPad versus paper administration of the M-CHAT [40]. Most parents did not require help to complete the computerized M-CHAT, and most parents did not experience anxiety when they viewed the result of their M-CHAT on the iPad screen [40]. Caregiver satisfaction with, and preference over its paper version, was also reported for the Taipei II [35,36]. Similarly, parents were reportedly easily able to use the NODA app without prior training [48]. Regarding the CHADIS, which was inclusive of the M-CHAT and ASQ, three-fourths of parents reported that the online previsit assessment improved their child welfare visit. However, nearly one-fourth of parents found the web-based assessment difficult to use [38].

How Effective Are Web-Based Screeners in Detecting Developmental Delay?

One identified study found that the web-based format of the earlier M-CHAT lowered both (1) false-positive screens (through the use of the M-CHAT follow-up if the child was deemed at risk after the first-level screen) and (2) false-negative screens (through the elimination of human scoring errors) [40]. However, in a later study comparing a paper-based version with the web-based version of the updated M-CHAT-R/F, no significant differences were observed in screen-positive rates or total scores [39]. Similarly, high positive and negative correspondence were demonstrated between the paper-based tools and smartphone PEDS tools in the South African trial [46].

Excellent agreement was also observed between the text and multimedia versions of the Taipei II [35], and the MARA reported 90% and 80% sensitivity and specificity in detecting ASD, respectively [52]. The Cognoa was compared with other often-used paper ASD screeners, including the M-CHAT-R/F. It was shown to accurately identify ASD in 71% of children aged 18 to 72 months. The overall specificity in detecting ASD (0.62) was significantly higher than the other measures [49]. The Gades System based on the Denver Developmental Screening Test demonstrated acceptable reliability for identifying language delay in children aged 0 to 3 years (97% agreement), and 67% agreement for children aged 4 to 6 years [54].

Discussion

Principal Findings

This was the first paper to identify existing web-based developmental surveillance or screening tools and summarize their acceptability and effectiveness in the early detection of developmental delay in infants and preschool children. It is timely, given the increased focus on electronic health and applications in health care [22-25]. The search identified 20 eligible studies. The majority (n=9) of studies used web-based versions of the established M-CHAT-R/F screener for ASD or PEDS tools for detecting broad developmental delay. Only 2 studies screened for language or psychomotor delay [54,55]. This indicates a lag in the web-based movement regarding these developmental domains. Implementation typically took the format of parents completing the screener on an iPad in the waiting room before a routine baby wellness checkup at a community health clinic, with the automatically scored results immediately provided to the treating practitioner to guide their action during the visit.

Web-based versions of developmental screeners demonstrated improved acceptability in primary care relative to their paper version. Caregivers preferred web-based screeners primarily for their ease of use and time efficiency [40,41]. However, families were less positive about the ease of use of the CHADIS. This could be because it involved an assortment of questionnaires, rather than 1 individual questionnaire [38].

Clinicians also found web-based versions of screeners easier and quicker to use. The clinical implication of this is readily apparent when one considers that a key factor for not using the follow-up screen of the M-CHAT/RF is administration time [41]. Specifically, studies reported improved screening rates, accurate documentation, and appropriate follow-up action for children screened positive when the web-based versions were used [39,41,42]. The M-CHAT follow-up screen is most helpful in reducing high initial screen false-positive rates [57]. In this way, it could have an indirect positive impact on the effectiveness of the screener. One identified study found that the web-based M-CHAT lowered both false-positive and false-negative screen rates [40]. However, another study found no significant difference in screen rates for the paper-based version versus the web-based version of the M-CHAT-R/F [39].

Certainly, evidence for the effectiveness of the web-based developmental screeners was minimal. Satisfactory sensitivity and specificity rates were reported for the MARA [52], and a marginally satisfactory specificity rate was reported for the Cognoa [49]. The paper-based M-CHAT-R/F and PEDS tools endorse strong psychometric properties and a wealth of effectiveness data [37,58]. However, it cannot be assumed that this translates over to web-based versions. Future research would benefit from examining whether the improved acceptability of web-based screeners reported in this review can translate into quicker identification of a potential developmental delay, earlier referrals, and ultimately earlier age of diagnosis.

Implications for Disadvantaged Communities

One of the biggest barriers to screening access is not speaking English. The screening tools developed in Bangladesh and Taiwan used pictures to avert literacy barriers [35,36,43,44]. The M-CHAT-R/F and PEDS tools are also well placed to address this barrier in that they are available in numerous different languages. However, where non-English speakers were included in a study, non-English–speaking caregivers were still less likely to complete the web-based version versus paper-based version of the screener. Furthermore, it was in clinics that employed attendants to read questions aloud for families with limited literacy that reported the highest web-based portal uptake [34].

The reported improvement in follow-up rates for web-based versus paper-based screeners may help reduce sociodemographic barriers relevant to developmental screening, in that higher rates of incomplete follow-ups are associated with lower maternal education [59]. The finding that community health workers could be trained in a smartphone app of the PEDS [46,47] also holds promise in increasing engagement rates in disadvantaged communities by removing some of the logistical barriers associated with screening access. The CHADIS was innovative in trying to capitalize on web-based screeners to minimize the resource demands involved in developmental screening via triaging the intensiveness of the routine child welfare checkup [38]. Although there is intuitive validity in assigning children with scores indicative of healthy or delayed development to an email-based or face-to-face welfare visit, respectively, most parents reported not wanting to see the e-visit replaced by a regular visit [38]. Certainly, experts reported that it was the video (ie, face-to-face) component of the Cognoa app that was most helpful in screening for ASD as compared with the parent report [49].

Indeed, the review identified several emerging smart applications that integrated caregiver report with videos for remote viewing by experts, such as the Cognoa and SmartAutism or AutismExpress [44,49,50]. Although such designs put an onus on the parent to initiate the developmental checkup, which may be less likely in disadvantaged communities [19,60], they do present as promising opportunities for accessing rural and under-resourced communities.

Several studies purposefully targeted culturally diverse or socially disadvantaged communities or families [34,46,47]. This is in keeping with policy body recommendations for targeted developmental surveillance. However, it challenges generalizability. Future research explicitly comparing the implementation of web-based screeners across disadvantaged and nondisadvantaged groups would facilitate a better understanding of the needs of disadvantaged communities relevant to screening inequities. Moreover, most identified studies were based in high-income countries; yet, it is in low- and middle-income countries that estimates of children younger than 5 years not reaching their developmental potential are the highest [8].

Limitations

The review was limited by its stringent exclusion criteria, meaning that some promising web-based tools in production were beyond the scope of the study. Furthermore, the inclusion of some studies that targeted a wider age range than the study focus may have added noise to the data [52,53], and raises the debate of where developmental screening ends and broader general assessment begins.

Conclusions

In summary, although the research is limited as to whether a web-based system is necessary for developmental screening, this review clearly highlights the important time and follow-up efficiencies that can facilitate policy body recommendations for universal developmental surveillance. Societal reliance on smart technology is increasing. It is hoped that increasing traction in web-based developmental screeners will continue as a possible means to promoting the valuable earlier detection of developmental delay in the infant and preschool years.

Acknowledgments

The authors would like to sincerely thank librarian Colleen Hutchison for her invaluable and enthusiastic support in guiding the search strategy. The authors would also like to thank the anonymous reviewers for their feedback. This study was funded by the Early Life Determinants of Health Clinical Academic Group of the Sydney Partnership for Health, Education, Research, and Enterprise (2017).

Authors' Contributions

JB, VE, JK, SW, CK, and RS participated in the design of the study. JB created the search strategy with Colleen Hutchinson. JB conducted the review, screening, data extraction, quality appraisal, and drafted the manuscript. JK acted as a second reviewer. VE, RS, and SW acted as third reviewers. SO performed the double data extraction and quality appraisal.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data extraction table.
[DOCX File , 55 KB - mhealth_v8i4e16085_app1.docx]

References

1. Tonge BJ, Einfeld SL. Psychopathology and intellectual disability: the Australian child to adult longitudinal study. Int Rev Res Ment Retard 2003;26:61-91. [doi: 10.1016/s0074-7750(03)01002-4]
2. Thurlbeck. Growing up in Britain: ensuring a healthy future for our children. Br Med J 2000 Mar 18;320(7237):809A [FREE Full text] [Medline: 10720390]

3. García JL, Heckman JJ, Leaf DE, Prados MJ. National Bureau of Economic Research. 2016 Dec. The Life-Cycle Benefits of an Influential Early Childhood Program URL: https://www.nber.org/papers/w22993.pdf [accessed 2020-04-07]
4. Einfeld SL, Tonge BJ, Clarke KS. Prevention and early intervention for behaviour problems in children with developmental disabilities. Curr Opin Psychiatry 2013 May;26(3):263-269. [doi: 10.1097/YCO.0b013e32835fd760] [Medline: 23493133]
5. Oberklaid F, Baird G, Blair M, Melhuish E, Hall D. Children's health and development: approaches to early identification and intervention. Arch Dis Child 2013 Dec;98(12):1008-1011. [doi: 10.1136/archdischild-2013-304091] [Medline: 23968776]
6. Adams RC, Tapia C, Council on children with disabilities. Early intervention, IDEA Part C services, and the medical home: collaboration for best practice and best outcomes. Pediatrics 2013 Oct;132(4):e1073-e1088. [doi: 10.1542/peds.2013-2305] [Medline: 24082001]
7. Glascoe FP. Parents' evaluation of developmental status: how well do parents' concerns identify children with behavioral and emotional problems? Clin Pediatr (Phila) 2003 Mar;42(2):133-138. [doi: 10.1177/000992280304200206] [Medline: 12659386]
8. Black MM, Walker SP, Fernald LC, Andersen CT, DiGirolamo AM, Lu C, Lancet Early Childhood Development Series Steering Committee. Early childhood development coming of age: science through the life course. Lancet 2017 Jan 7;389(10064):77-90 [FREE Full text] [doi: 10.1016/S0140-6736(16)31389-7] [Medline: 27717614]
9. Eapen V, Woolfenden S, Williams K, Jalaludin B, Dissanayake C, Axelsson EL, et al. 'Are you available for the next 18 months?' - methods and aims of a longitudinal birth cohort study investigating a universal developmental surveillance program: the 'Watch Me Grow' study. BMC Pediatr 2014 Sep 22;14:234 [FREE Full text] [doi: 10.1186/1471-2431-14-234] [Medline: 25241772]
10. Overs BJ, Woolfenden S, Williams K, Jalaludin B, Axelsson EL, Dissanayake C, 'Watch Me Grow' Study Group. Predictors of developmental surveillance completion at six months of age in south western Sydney. Child Care Health Dev 2017 Mar;43(2):307-315 [FREE Full text] [doi: 10.1111/cch.12425] [Medline: 27910128]
11. Garg P, Ha MT, Eastwood J, Harvey S, Woolfenden S, Murphy E, et al. Explaining culturally and linguistically diverse (CALD) parents' access of healthcare services for developmental surveillance and anticipatory guidance: qualitative findings from the 'Watch Me Grow' study. BMC Health Serv Res 2017 Mar 22;17(1):228 [FREE Full text] [doi: 10.1186/s12913-017-2143-1] [Medline: 28330490]
12. Eapen V, Walter A, Guan J, Descallar J, Axelsson E, Einfeld S, The 'Watch Me Grow' Study Group. Maternal help-seeking for child developmental concerns: associations with socio-demographic factors. J Paediatr Child Health 2017 Oct;53(10):963-969. [doi: 10.1111/jpc.13607] [Medline: 28661061]
13. Woolfenden S, Eapen V, Axelsson E, Hendry A, Jalaludin B, Dissanayake C, 'Watch Me Grow' study group. Who is our cohort: recruitment, representativeness, baseline risk and retention in the 'Watch Me Grow' study? BMC Pediatr 2016 Mar 24;16:46 [FREE Full text] [doi: 10.1186/s12887-016-0582-1] [Medline: 27013220]
14. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol 2014 Jan;67(1):56-64. [doi: 10.1016/j.jclinepi.2013.08.005] [Medline: 24189091]
15. Carey W, Coleman WL, Crocker AC, Feldman HM, Elias ER. Developmental Behavioral Pediatrics. Fourth Edition. Philadelphia: Elsevier Health Sciences; 2009.
16. -. Developmental surveillance and screening of infants and young children. Pediatrics 2001 Jul;108(1):192-196. [doi: 10.1542/peds.108.1.192] [Medline: 11433077]
17. Suh H, Porter JR, Hiniker A, Kientz JA. @BabySteps: Design and Evaluation of a System for Using Twitter for Tracking Children's Developmental Milestones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI'14; April 26 - May 1, 2014; Ontario, Toronto, Canada p. 2279-2288. [doi: 10.1145/2556288.2557386]
18. Ben-Sasson A, Ben-Sasson E, Jacobs K, Saig E. Baby CROINC: An Online, Crowd-Based, Expert-Curated System for Monitoring Child Development. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth'17; May 23 - 26, 2017; Barcelona, Spain p. 110-119. [doi: 10.1145/3154862.3154887]
19. Garg P, Ha MT, Eastwood J, Harvey S, Woolfenden S, Murphy E, et al. Health professional perceptions regarding screening tools for developmental surveillance for children in a multicultural part of Sydney, Australia. BMC Fam Pract 2018 Apr 2;19(1):42 [FREE Full text] [doi: 10.1186/s12875-018-0728-3] [Medline: 29609547]
20. Woolfenden S, Eapen V, Jalaludin B, Hayen A, Kemp L, Dissanyake C, 'Watch Me Grow' study group. Prevalence and factors associated with parental concerns about development detected by the Parents' Evaluation of Developmental Status (PEDS) at 6-month, 12-month and 18-month well-child checks in a birth cohort. BMJ Open 2016 Sep 8;6(9):e012144 [FREE Full text] [doi: 10.1136/bmjopen-2016-012144] [Medline: 27609853]
21. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
22. Peek ME. Can mHealth Interventions Reduce Health Disparities among Vulnerable Populations? Divers Equal Health Care 2017;14(2). [doi: 10.21767/2049-5471.100091]

23. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Sep 1;8(9):54429 [FREE Full text] [doi: 10.5539/gjhs.v8n9p183] [Medline: 27157176]
24. Sondaal SFV, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
25. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and Evaluating mHealth Interventions for Vulnerable Populations: A Systematic Review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21 - 26, 2018; Montreal QC, Canada p. 1-17. [doi: 10.1145/3173574.3173589]
26. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and appropriateness of mHealth interventions for maternal and child health: systematic review. JMIR Mhealth Uhealth 2018 Jan 9;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
27. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
28. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
29. Jackson N, Waters E, Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. Health Promot Int 2005 Dec;20(4):367-374. [doi: 10.1093/heapro/dai022] [Medline: 16169885]
30. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
31. Thomas B, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
32. Oxman AD, Lavis JN, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 10: Taking equity into consideration when assessing the findings of a systematic review. Health Res Policy Syst 2009 Dec 16;7(Suppl 1):S10. [doi: 10.1186/1478-4505-7-S1-S10] [Medline: 20018100]
33. Dehkordi PS, Moradi H, Mahmoudi M, Pouretemad HR. The design, development, and deployment of RoboParrot for screening autistic children. Int J Soc Robot 2015 Jul 5;7(4):513-522. [doi: 10.1007/s12369-015-0309-8]
34. Glascoe FP. Evidence-based early detection of developmental-behavioral problems in primary care: what to expect and how to do it. J Pediatr Health Care 2015;29(1):46-53. [doi: 10.1016/j.pedhc.2014.06.005] [Medline: 25088278]
35. Cheng HK, Chang H, Huang P, Ju Y, Chen L, Tseng KC. The design and validation of a child developmental e-screening system. J Med Syst 2017 Apr;41(4):67. [doi: 10.1007/s10916-017-0701-z] [Medline: 28283996]
36. Cheng HK, Chen L, Cheng C, Ju Y, Chen C, Tseng KC. A multimedia child developmental screening checklist: design and validation. J Med Internet Res 2016 Oct 24;18(10):e277 [FREE Full text] [doi: 10.2196/jmir.6249] [Medline: 27777218]
37. Robins DL, Casagrande K, Barton M, Chen CM, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics 2014 Jan;133(1):37-45 [FREE Full text] [doi: 10.1542/peds.2013-1813] [Medline: 24366990]
38. Bergman DA, Beck A, Rahm AK. The use of internet-based technology to tailor well-child care encounters. Pediatrics 2009 Jul;124(1):e37-e43. [doi: 10.1542/peds.2008-3385] [Medline: 19564267]
39. Brooks BA, Haynes K, Smith J, McFadden T, Robins DL. Implementation of web-based autism screening in an urban clinic. Clin Pediatr (Phila) 2016 Sep;55(10):927-934 [FREE Full text] [doi: 10.1177/0009922815616887] [Medline: 26581361]
40. Harrington JW, Bai R, Perkins AM. Screening children for autism in an urban clinic using an electronic M-CHAT. Clin Pediatr (Phila) 2013 Jan;52(1):35-41. [doi: 10.1177/0009922812463957] [Medline: 23117237]
41. Campbell K, Carpenter KL, Espinosa S, Hashemi J, Qiu Q, Tepper M, et al. Use of a digital modified checklist for autism in toddlers - revised with follow-up to improve quality of screening for autism. J Pediatr 2017 Apr;183:133-9.e1 [FREE Full text] [doi: 10.1016/j.jpeds.2017.01.021] [Medline: 28161199]
42. Sturner R, Howard B, Bergmann P, Morrel T, Andon L, Marks D, et al. Autism screening with online decision support by primary care pediatricians aided by M-CHAT/F. Pediatrics 2016 Sep;138(3) [FREE Full text] [doi: 10.1542/peds.2015-3036] [Medline: 27542847]
43. Bardhan S, Mridha MM, Ahmed E, Ullah MA, Ahmed HU, Akhter S, et al. Autism Barta - A Smart Device Based Automated Autism Screening Tool for Bangladesh. In: Proceedings of the 2016 5th International Conference on Informatics, Electronics and Vision. 2016 Presented at: ICIEV'16; May 13-14, 2016; Dhaka, Bangladesh. [doi: 10.1109/ICIEV.2016.7760073]

44. Bardhan S, Ullah A, Ahmed HU, Rabbani MG, Al Mamun KA. Autism Express - A Cloud Based Framework for Autism Screening, Confirmation and Intervention. In: Proceedings of the 2016 IEEE Region 10 Conference. 2016 Presented at: TENCON'16; November 22-25, 2016; Singapore. [doi: 10.1109/tencon.2016.7848032]
45. Gagliano A, Fama FI. Prima Pietra: A Research and Clinical Program on Early Detection and Intervention in Italy. In: IRIS Università degli Studi di Messina. 2015 Presented at: International Congress of ESCAP European Child and Adolescent Psychiatry; 2015; Madrid URL: https://iris.unime.it/handle/11570/3103689#.XoxkoMgzaM8
46. Maleka BK, van der Linde J, Glascoe FP, Swanepoel DW. Developmental screening-evaluation of an m-Health version of the parents evaluation developmental status tools. Telemed J E Health 2016 Dec;22(12):1013-1018. [doi: 10.1089/tmj.2016.0007] [Medline: 27286191]
47. van der Merwe MN, Mosca R, Swanepoel DW, Glascoe FP, van der Linde J. Early detection of developmental delays in vulnerable children by community care workers using an mHealth tool. Early Child Dev Care 2019;189(5):855-866. [doi: 10.1080/03004430.2018.1480481]
48. Nazneen N, Rozga A, Smith CJ, Oberleitner R, Abowd GD, Arriaga RI. A novel system for supporting autism diagnosis using home videos: iterative development and evaluation of system design. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e68 [FREE Full text] [doi: 10.2196/mhealth.4393] [Medline: 26085230]
49. Kanne SM, Carpenter LA, Warren Z. Screening in toddlers and preschoolers at risk for autism spectrum disorder: evaluating a novel mobile-health screening tool. Autism Res 2018 Jul;11(7):1038-1049. [doi: 10.1002/aur.1959] [Medline: 29734507]
50. Al Mamun KA, Bardhan S, Ullah MA, Anagnostou E, Brian J, Akhter S, et al. Smart Autism - A Mobile, Interactive and Integrated Framework for Screening and Confirmation of Autism. In: Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2016 Presented at: EMBC'16; August 16-20, 2016; Orlando, FL, USA p. 5989-5992. [doi: 10.1109/EMBC.2016.7592093]
51. Scott FJ, Baron-Cohen S, Bolton P, Brayne C. The CAST (Childhood Asperger Syndrome Test): preliminary development of a UK screen for mainstream primary-school-age children. Autism 2002 Mar;6(1):9-31. [doi: 10.1177/1362361302006001003] [Medline: 11918111]
52. Duda M, Daniels J, Wall DP. Clinical evaluation of a novel and mobile autism risk assessment. J Autism Dev Disord 2016 Jun;46(6):1953-1961 [FREE Full text] [doi: 10.1007/s10803-016-2718-4] [Medline: 26873142]
53. Jordan E. Pirate adventure autism assessment app: a new tool to aid clinical assessment of children with possible autistic spectrum disorder. J Altern Med Res 2018;10(1):17-23 [FREE Full text]
54. Ruiz ML, Duboy M, Loriente CT, de la Cruz IP. Evaluating a web-based clinical decision support system for language disorders screening in a nursery school. J Med Internet Res 2014 May 28;16(5):e139 [FREE Full text] [doi: 10.2196/jmir.3263] [Medline: 24870413]
55. García MA, Ruiz ML, Rivera D, Vadillo L, Duboy MA. A smart toy to enhance the decision-making process at children's psychomotor delay screenings: a pilot study. J Med Internet Res 2017 May 19;19(5):e171 [FREE Full text] [doi: 10.2196/jmir.7533] [Medline: 28526666]
56. Velikonja T, Edbrooke-Childs J, Calderon A, Sleed M, Brown A, Deighton J. The psychometric properties of the Ages & Stages Questionnaires for ages 2-2.5: a systematic review. Child Care Health Dev 2017 Jan;43(1):1-17. [doi: 10.1111/cch.12397] [Medline: 27554865]
57. Khowaja M, Robins DL, Adamson LB. Utilizing two-tiered screening for early detection of autism spectrum disorder. Autism 2018 Oct;22(7):881-890 [FREE Full text] [doi: 10.1177/1362361317712649] [Medline: 28905644]
58. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, et al. Screening for autism spectrum disorder in young children: US preventive services task force recommendation statement. J Am Med Assoc 2016 Feb 16;315(7):691-696. [doi: 10.1001/jama.2016.0018] [Medline: 26881372]
59. Khowaja MK, Hazzard AP, Robins DL. Sociodemographic barriers to early detection of autism: screening and evaluation using the M-CHAT, M-CHAT-R, and follow-up. J Autism Dev Disord 2015 Jun;45(6):1797-1808 [FREE Full text] [doi: 10.1007/s10803-014-2339-8] [Medline: 25488122]
60. Woolfenden S, Posada N, Krchnakova R, Crawford J, Gilbert J, Jursik B, et al. Equitable access to developmental surveillance and early intervention--understanding the barriers for children from culturally and linguistically diverse (CALD) backgrounds. Health Expect 2015 Dec;18(6):3286-3301 [FREE Full text] [doi: 10.1111/hex.12318] [Medline: 25510182]

Abbreviations

ARC: autism resource center
ASD: autism spectrum disorder
ASQ: Ages and Stages Questionnaire
CHADIS: Child Health and Development Interactive System
EPHPP: Effective Public Health Practice Project
e-visit: electronic visit
MARA: Mobile Autism Risk Assessment
M-CHAT-R/F: Modified Checklist for Autism in Toddlers, Revised with Follow-Up

NODA: Naturalistic Observation Diagnostic Assessment
PEDS: Parents' Evaluation of Developmental Status

Edited by G Eysenbach; submitted 02.09.19; peer-reviewed by F Glascoe, J VanDerLinde, A Ben-Sasson; comments to author 08.10.19; revised version received 05.01.20; accepted 10.02.20; published 23.04.20.

Please cite as:
Baker J, Kohlhoff J, Onobrakpor SI, Woolfenden S, Smith R, Knebel C, Eapen V
The Acceptability and Effectiveness of Web-Based Developmental Surveillance Programs: Rapid Review
JMIR Mhealth Uhealth 2020;8(4):e16085
URL: http://mhealth.jmir.org/2020/4/e16085/
doi:10.2196/16085
PMID:32324149

Review

Effectiveness of Mobile Health Interventions on Diabetes and Obesity Treatment and Management: Systematic Review of Systematic Reviews

Youfa Wang[1,2*], MS, MD, PhD; Jungwon Min[1,3*], MS, PhD; Jacob Khuri[4], MPH; Hong Xue[5], MS, PhD; Bo Xie[6,7], MS, PhD; Leonard A Kaminsky[1], PhD; Lawrence J Cheskin[8,9], MD

[1]Fisher Institute of Health and Well-Being, College of Health, Ball State University, Muncie, IN, United States
[2]Department of Nutrition and Health Sciences, College of Health, Ball State University, Muncie, IN, United States
[3]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, United States
[4]School of Public Health, Imperial College London, London, United Kingdom
[5]Department of Health Behavior and Policy, Virginia Commonwealth University, Richmond, VA, United States
[6]School of Nursing, The University of Texas at Austin, Austin, TX, United States
[7]School of Information, The University of Texas at Austin, Austin, TX, United States
[8]Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States
[9]Department of Health, Behavior and Society, Johns Hopkins University, Baltimore, MD, United States
[*]these authors contributed equally

Corresponding Author:
Youfa Wang, MS, MD, PhD
Fisher Institute of Health and Well-Being
College of Health
Ball State University
2000 W University Ave
Muncie, IN, 47306
United States
Phone: 1 765 285 8259
Email: youfawang@gmail.com

Abstract

Background: Diabetes and obesity have become epidemics and costly chronic diseases. The impact of mobile health (mHealth) interventions on diabetes and obesity management is promising; however, studies showed varied results in the efficacy of mHealth interventions.

Objective: This review aimed to evaluate the effectiveness of mHealth interventions for diabetes and obesity treatment and management on the basis of evidence reported in reviews and meta-analyses and to provide recommendations for future interventions and research.

Methods: We systematically searched the PubMed, IEEE Xplore Digital Library, and Cochrane databases for systematic reviews published between January 1, 2005, and October 1, 2019. We analyzed 17 reviews, which assessed 55,604 original intervention studies, that met the inclusion criteria. Of those, 6 reviews were included in our meta-analysis.

Results: The reviews primarily focused on the use of mobile apps and text messaging and the self-monitoring and management function of mHealth programs in patients with diabetes and obesity. All reviews examined changes in biomarkers, and some reviews assessed treatment adherence (n=7) and health behaviors (n=9). Although the effectiveness of mHealth interventions varied widely by study, all reviews concluded that mHealth was a feasible option and had the potential for improving patient health when compared with standard care, especially for glycemic control (−0.3% to −0.5% greater reduction in hemoglobin A_{1c}) and weight reduction (−1.0 kg to −2.4 kg body weight). Overall, the existing 6 meta-analysis studies showed pooled favorable effects of these mHealth interventions (−0.79, 95% CI −1.17 to −0.42; I2=90.5).

Conclusions: mHealth interventions are promising, but there is limited evidence about their effectiveness in glycemic control and weight reduction. Future research to develop evidence-based mHealth strategies should use valid measures and rigorous study designs. To enhance the effectiveness of mHealth interventions, future studies are warranted for the optimal formats and

the frequency of contacting patients, better tailoring of messages, and enhancing usability, which places a greater emphasis on maintaining effectiveness over time.

(JMIR Mhealth Uhealth 2020;8(4):e15400) doi:10.2196/15400

KEYWORDS

diabetes mellitus; obesity; overweight; mHealth; mobile app; telemedicine

Introduction

Background

Diabetes and obesity have become global epidemics [1,2]. The global prevalence of type 2 diabetes mellitus (T2DM) and overweight/obesity among adults had increased from 9% in 2014 [3] to 40% in 2016 [4]. They both have significant and overlapping health and economic consequences, such as excess morbidity and mortality and health care resource utilization [5,6]. The global direct cost of diabetes and obesity has been estimated at US $825 billion and US $2 trillion per year, respectively [7,8]. However, patients often do not have adequate access to or cannot afford health care. Many others are not able to adhere to their treatment regimen, particularly in low-resource settings [9]. These diseases are difficult to manage effectively, and thus, patients suffer from more complications, in addition to the financial burden on themselves and society [9,10]. Thus, providing adequate health care services that enable patients to manage these chronic diseases is critical.

Self-management practices, such as maintaining a healthy diet and weight, engaging in adequate physical activity (PA), using prescribed medications consistently, frequently checking body weight and blood sugar levels, and maintaining good mental health habits, help patients control diabetes and obesity efficiently [11]. Previous studies have shown that self-management support in diabetes improves hemoglobin A_{1c} (HbA_{1c}) levels, reduces risks of developing life-threatening complications, and positively affects patient psychosocial and behavioral health [6]. However, the lack of individualized and coordinated care, inconvenient and costly education programs, and poor patient-provider communication make self-management practices challenging to adhere to and maintain. Effective services and methods for self-management are needed to reduce health care costs associated with these conditions, while improving the patient's quality of life [12].

Emerging mobile health (mHealth) approaches may help meet these needs. In both developed and developing countries, mobile technology and device usage has been rapidly increasing and plays a vital role in people's daily life [13]. Mobile technology provides mobility, instant access, and direct communication, which allows for faster transfer of health information and efficient health management assistance for patients [14,15]. It can also help provide better and expanded access to more affordable health services in low-income countries and low socioeconomic status groups in middle- and high-income countries [16]. Mobile technologies, specifically mobile apps, present an opportunity to help patients improve their adherence to health care providers' advice, enhance patient-provider communication, and help facilitate and maintain behavioral changes [17,18]. mHealth is increasingly being used to improve the access and delivery of health services, treatment adherence, and management of various diseases and health risk–altering behaviors, such as HIV/AIDS, malaria, tuberculosis, diabetes, asthma, obesity, and smoking [16]. However, research on the applications of mHealth is still at an early stage of development and translation, and many unanswered questions remain.

The availability of commercial chronic disease self-management apps has been increasing rapidly [19,20]. Commercial apps may offer patients many high-quality choices in the self-management of their diseases and conditions. However, the large number of these apps makes it difficult for patients and health care providers to choose among the options wisely. Furthermore, only a small proportion of apps, other mobile devices, and programs have been appropriately tested for effectiveness.

Some previous reviews, including ours, have described the development of app technologies and their utility for patients with obesity, diabetes, and other chronic conditions [19,21-37]. However, their scope did not adequately address the effectiveness of mHealth for diabetes and obesity treatment and management. Many reviews have concluded that mHealth is promising for disease control but reported inconsistent findings on its effectiveness. Furthermore, the methods used in previous reviews have often been flawed for reasons such as not providing quantitative results, conducting a quantitative analysis with clinical/nonclinical trials and other study designs together, not using standardized data extraction, and a limited scope of review [24,35]. A thorough review of the evidence is needed and can help guide future research and interventions [38,39].

Objectives

This study evaluated the effectiveness of mHealth interventions for diabetes and obesity treatment/management by examining published systematic reviews and meta-analyses and provided recommendations for future research and interventions.

Methods

Study Selection

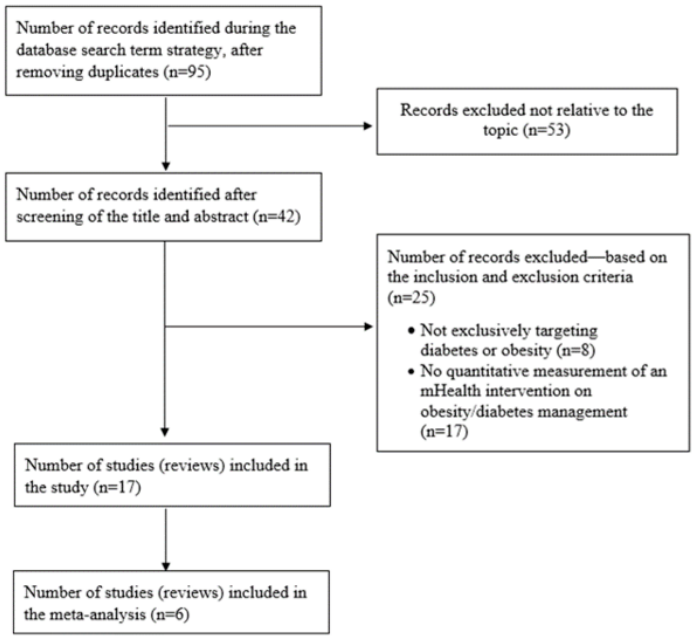
Database and Literature Search Strategy

We searched the PubMed, IEEE Xplore Digital Library, and Cochrane databases to identify systematic reviews and meta-analyses published in English between January 1, 2005, and October 1, 2019, that evaluated the effectiveness of mHealth interventions for obesity and/or diabetes treatment/management. For the search, combinations of key terms were used in the PubMed, for example, "mhealth[Title/Abstract] AND (obesity[Title/Abstract] OR diabetes*[Title/Abstract]) AND review[Title/Abstract]." Search results were further screened

manually by study title, abstract, and full text on the basis of inclusion and exclusion criteria.

The initial search yielded 95 articles. After eliminating duplicates and studies that did not fit the inclusion criteria, 17 reviews meeting the inclusion criteria remained; 6 of the 17 reviews were meta-analyses with randomized controlled trials (RCTs; Figure 1). The 17 reviews assessed 55,604 original intervention studies.

Figure 1. A flow chart of the literature search and study selection procedures. mHealth: mobile health.



Study Inclusion Criteria

Studies were included if they (1) reviewed intervention studies on patients with obesity or/and diabetes; (2) were a systematic review and/or a meta-analysis; (3) tested an mHealth intervention (eg, use of mobile devices, apps, and text message) for managing or treating obesity/diabetes while measuring clinical biomarkers, treatment adherence, or health-related behaviors (eg, healthy eating and exercise); and (4) provided quantitative results examining the effectiveness of the intervention (or use of the mHealth devices/programs).

Study Exclusion Criteria

Studies were excluded if they (1) did not explicitly target diabetes or obesity; (2) were diabetes or obesity prevention studies, not using an mHealth-based intervention program; and (3) did not report quantitative outcomes of mHealth intervention effects in managing obesity or diabetes.

Study Quality Assessment

We used the Assessment of Multiple Systematic Reviews (AMSTAR 2) to assess the quality of selected studies by 16 criteria (eg, study selection, data extraction, assessing risk of bias, study description, and statistical methods) according to the study characteristics [40]. We assigned 1 point to each item that scored *yes* and summed these to calculate a total score (ranging from 0 to 16) for each review. We classified the quality of systematic reviews as high (score range 12-16), moderate (score range 9-11), low (score range 5-8), or critically low (score range 0-4; Multimedia Appendix 1) [41].

Data Extraction and Statistical Analysis

Data were reviewed and extracted by 2 coauthors following the Preferred Reporting Items for Systematic Review and Meta-Analysis guidelines [42]. The information extracted included study year; design; objective; literature search scope and date; the number of articles accessed and included in the systematic review; and nature of the intervention, such as application type and targeted function, outcome measures related to clinical biomarkers, treatment adherence and health-related behaviors, and effectiveness of the mHealth intervention.

Using mixed effect models, we conducted a meta-analysis to evaluate the overall effectiveness of mHealth interventions on the basis of other published meta-analysis results with RCTs. The STATA (StataCorp LLC) metan command was used to calculate pooled estimates of mean differences in changes in clinical outcomes such as HbA_{1c}, body weight, and BMI between intervention and control groups [21-23,32-34].

Results

Main Characteristics of the 17 Included Reviews

Study Topics and Study Design

Multimedia Appendix 2 describes the characteristics of the 17 reviews: 7 studies were conducted in the United States and the others were conducted in Australia (n=1), Canada (n=1), China (n=3), Germany (n=1), New Zealand (n=1), South Korea (n=1), Spain (n=1), and the United Kingdom (n=1). Regarding study design, 15 of the reviews were based on clinical trials, of which 9 included only RCTs, whereas 6 also included quasi-experimental studies. One review was rated as critically low quality (AMSTAR 2 score 0-4), 5 were rated as low quality

(AMSTAR 2 score 5-8), 5 were rated as moderate quality (AMSTAR 2 score 9-11), and 6 were categorized as high quality (AMSTAR 2 score 12-16). Of the 17 reviews, 10 reviewed mHealth interventions in patients with diabetes, 6 in patients with overweight/obesity, and 1 included both conditions. Meta-analyses for various health outcomes were conducted in 6 reviews with RCTs.

We conducted the meta-analysis of the *mean difference* on the basis of the results reported in the 6 reviews using a mixed effect model. The STATA metan command was used to calculate pooled estimates, with confidence limits of mean difference in clinical outcomes. The I^2 statistic quantifies the percentage of variability that can be attributed to between-study differences.

The mean difference was calculated by subtracting the level of clinical outcomes at the end of follow-up from the baseline, comparing the intervention and control groups. This allowed for a comparison of clinical improvement because of the mHealth interventions vs the control group.

Types of Mobile Health Interventions

We categorized the mHealth interventions studied into 5 types (Multimedia Appendix 2): (1) an app, which uses smartphones to deliver educational materials or help patients self-manage their health condition; (2) web-based tools used to provide patient education and/or advice on self-management; (3) text messaging, which uses mobile phone text messages as the primary mode of communication between patients and health care providers; (4) a portable monitoring device/personal digital assistant (PDA), which typically offers patient data collection over a wireless connection and can monitor patients' physiological status; and (5) a pedometer that counts the number of steps taken in a day. These classifications were made on the basis of several considerations, including simplicity, understandability for a nontechnical audience, and the technological complexity involved in the intervention [31].

Mobile apps were the most widely studied intervention type (15 reviews), followed by text messaging (11 reviews) and PDAs (5 reviews). Regarding the major targeted functions of the mHealth interventions reviewed, self-monitoring and management was most common (15 reviews), followed by education or health promotion (8 reviews), reminders or alerts (5 reviews), feedback (3 reviews), social or peer support (2 reviews), and counseling or entertainment (1 review; see Multimedia Appendices 2 and 3).

Targeted Outcomes

All 17 reviews examined changes in clinical biomarkers as outcomes, whereas 9 evaluated health-related behaviors and 7 assessed treatment adherence (Multimedia Appendix 2). HbA_{1c} levels were included as clinical biomarkers in all the reviews of diabetes, whereas body weight/weight status (6 reviews) and BMI (5 reviews) were the main outcomes for obesity intervention reviews. Blood pressure (4 reviews), serum lipid/cholesterol levels (4 reviews), waist circumference, severe hypoglycemia/adverse effects (1 review), and C-reactive protein level (1 review) were also explored. For measuring treatment adherence, medication/treatment adherence and glycemic self-control/monitoring were most common. As indicators of health-related behaviors, PA and diet were frequently measured (n=6), whereas PA alone (n=1), other obesity-related behaviors, or self-care behaviors were less frequently reviewed (n=2).

The Effectiveness of Mobile Health Interventions in Managing Obesity or Diabetes

Clinical Outcomes

We found much heterogeneity in the effectiveness of mHealth interventions for clinical biomarkers (Multimedia Appendix 3). For blood glucose control (including HbA_{1c}), 8 reviews reported statistically significant or large improvements (more than or equal to half of the included studies) [21-23,25,26,29-31], although another 3 reported low improvements (less than half of the included articles) [24,27,28]. Treatment effects on BMI, weight, and waist circumference varied; 5 reviews found large improvements [31-35] and 3 reported small or no effect of the mHealth interventions [23,36,37]. mHealth interventions were found ineffective for improving serum lipids changes in 2 reviews [23,28], whereas 1 review found a few positive changes in cholesterol levels [29]. Blood pressure levels showed small improvements in 2 reviews [24,29], but 2 other reviews found no effect [23,28]. The heterogeneous results may reflect differences in study subjects (eg, T2DM vs type 1 diabetes mellitus [T1DM]) [21,30] and the severity of symptoms (HbA_{1c}<8% vs HbA_{1c}≥8%) [23]. The small number of reviews on serum lipids, cholesterol, and blood pressure (eg, <5 reviews) may be insufficient to examine the effectiveness of mHealth for these parameters.

Regarding the meta-analyses, 3 reviews reported on the effect of mobile apps on HbA_{1c} levels in diabetes [21-23]. A meta-analysis indicated a significant reduction in HbA_{1c} from 0.25% (95% CI −0.41 to −0.09) [21] to 0.48% (95% CI −0.78 to −0.19) [22], presented in Table 1 and Figure 2, but with substantial heterogeneity in the pooled effect (I^2 up to 77%). In particular, differences between the mHealth intervention group and the control group were significant for patients with HbA_{1c} <8% at baseline by −0.33% (−3.61 mmol/mol; I^2= 70%), whereas it was not significant in the patients with HbA_{1c} ≥8% (*P*=.33) [23]. In addition, larger reductions were noticed after app use in HbA_{1c} among patients with T2DM (−0.67%, 95% CI −1.03 to −0.30) [22] compared with patients with T1DM (−0.37%, 95% CI −0.86 to −0.12).

A meta-analysis on RCTs consistently found that app use was associated with significant improvements in body weight and BMI [32-34] from −1.04 kg (95% CI −1.75 to −0.34; I^2=41%) [33] to −2.35 kg (95% CI −2.84 to −1.87; I^2=94%) [32] and from −0.43 kg/m^2 (95% CI −0.74 to −0.13; I^2=50%) [33] to −0.77 kg/m^2 (95% CI −1.01 to −0.52; I^2=0%) [32] than the control group, respectively. When stratified by the application type, only mobile-based interventions showed significant body weight loss (−1.78 kg, 95% CI −2.92 to −0.63; I^2=16%), whereas PDA-based interventions showed nonsignificant changes (−0.23 kg, 95% CI −0.87 to 0.41; I^2=0.0%) [34].

Table 1. Summary of clinical outcomes and behavioral changes from 18 meta-analyses reported in 6 reviews of diabetes and obesity mobile health interventions.

Outcomes	References[a]	Tested interventions/target patient	Intervention vs control groups	Estimated effect of intervention: meta-analysis results of the mean difference between intervention and control groups	Conclusions
HbA_{1c}[b]	Wang et al [21]	Self-management of patients with T1DM[c]	Mobile app or text messaging intervention vs standard care	• −0.25% (95% CI −0.41 to −0.09; I^2=12%) • Subgroup analysis—age: teenagers −0.05% (95% CI −0.43 to 0.33; I^2=0%); adults −0.29% (95% CI −0.47 to −0.11; I^2=48%) • Subgroup analysis—intervention: text message −0.20% (95% CI −0.73 to 0.32; I^2=0%); mobile apps −0.25% (95% CI −0.42 to −0.08; I^2=49%) • Subgroup analysis—duration: ≥6 months −0.29% (95% CI −0.46 to −0.11; I^2=32%); <6 months −0.01% (95% CI −0.44 to 0.41; I^2=0%)	mHealth[d] favors
HbA_{1c}	Wu et al [22]	Self-management of patients with diabetes	Mobile app intervention vs standard care alone	• −0.48% (95% CI −0.78 to −0.19; I^2=76%) • Subgroup analysis: patients with T2DM[e] −0.67% (95% CI −1.03 to −0.30; I^2=47%); patients with T1DM −0.37% (95% CI −0.86 to −0.12; I^2=86%)	mHealth favors
HbA_{1c}	Cui et al [23]	Self-management of patients with T2DM	Smartphone app strategies vs standard diabetes care	• −0.40% (95% CI −0.69 to −0.11; I^2=77%) • Subgroup analysis: baseline HbA_{1c}<8% −0.33% (95% CI −0.59 to −0.06; I^2=70%)	mHealth favors
Body weight	Park et al [32]	Weight loss interventions on patients with OWB[f]	Mobile app/text messaging intervention vs nonmobile device care (standard)	• −2.35 kg (95% CI −2.84 to −1.87; I^2=94%) • Subgroup analysis—duration: at 6 months −2.66 kg (95% CI −3.94 to −1.38; I^2=95%); at ≥12 months −1.23 kg (95% CI −2.25 to −0.21; I^2=0%)	mHealth favors
Body weight	Mateo et al [33]	Weight loss and PA[g] promotion on patients with OWB	Mobile app intervention vs the control diet	• −1.04 kg (95% CI −1.75 to −0.34; I^2=41%)	mHealth favors
Body weight	Khokhar et al [34]	Weight loss interventions on patients with OWB	Mobile electronic device intervention vs the control	• −1.09 kg (95% CI −2.12 to −0.05; I^2=50%) • Subgroup analysis—duration: ≤6 months −0.97 kg (95% CI −2.23 to 0.30; I^2=47%); >6 months −1.20 kg (95% CI −3.34 to 0.94; I^2=62%) • Subgroup analysis—intervention: mobile phone −1.78 kg (95% CI −2.92 to −0.63; I^2=16%); personal digital assistant −0.23 kg (95% CI −0.87 to 0.41; I^2=0.0%)	mHealth favors

Outcomes	References[a]	Tested interventions/target patient	Intervention vs control groups	Estimated effect of intervention: meta-analysis results of the mean difference between intervention and control groups	Conclusions
BMI	Park et al [32]	Weight loss interventions on patients with OWB	Mobile app/text messaging intervention vs nonmobile device care (standard)	• –0.77 kg/m^2 (95% CI −1.01 to −0.52; I^2=0%) • Subgroup analysis—duration: at 3 months −1.10 kg/m^2 (95% CI −2.79 to 0.59; I^2=95%); at 6 months −0.67 kg/m^2 (95% CI −0.71 to −0.63; I^2=0%)	mHealth favors
BMI	Mateo et al [33]	Weight loss and PA promotion on patients with OWB	Mobile app intervention vs the control diet	• −0.43 kg/m^2 (95% CI −0.74 to −0.13; I^2=50%)	mHealth favors
Physical activity	Mateo et al [33]	Weight loss and PA promotion on patients with OWB	Mobile app intervention vs control intervention	• Standardized mean difference in net change 0.40 (95% CI −0.07 to 0.87; I^2=93%)	No significant difference

[a]We selected 6 meta-analyses on randomized controlled trial studies. Please see our pooled meta-analysis presented in Figure 2.

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

[c]T1DM: type 1 diabetes mellitus.

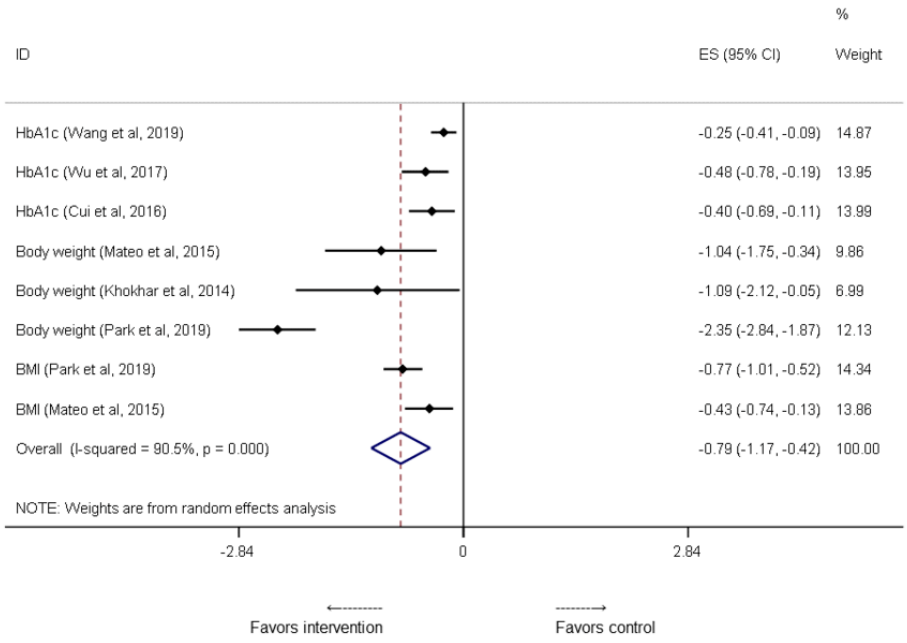
[d]mHealth: mobile health.

[e]T2DM: type 2 diabetes mellitus.

[f]OWB: overweight and obesity.

[g]PA: physical activity.

Figure 2. A meta-analysis of mean differences in changes in clinical outcomes after an intervention, mobile health versus control groups. HbA_{1c}: hemoglobin A_{1c}.



Treatment Adherence

Relatively few reviews examined the treatment effect of mHealth interventions and reported inconsistent results. Out of the 7 reviews that investigated mHealth intervention effects on treatment adherence, 4 reviews found a moderate improvement in glycemic control [23,26], greater adherence in the mHealth intervention group compared with the control group [35], or a reduced attrition rate during an obesity intervention program [37]. However, the other 2 reviews found that fewer than half of the studies improved diabetes management practices [24,28] or medication adherence [28].

Behavioral Changes

Results for behavioral changes were not consistent, and 3 diabetes reviews [24,26,28] and 1 obesity review [35] found a slight improvement in patients' diets, eating habits, and PA behaviors. However, 3 reviews found inconsistent behavioral

changes by disease (diabetes vs obesity) [21] or type of behavior (eating habits/dietary intake vs PA) [36,37], and 1 review found no significant effect of mHealth interventions on the level of PA [33]. A meta-analysis of RCTs found a nonsignificant difference in PA between the intervention and control groups, with a standardized mean difference of 0.40 (95% CI −0.07 to 0.87; I^2=93%; Table 1).

Our meta-analysis (see Figure 2) on results from the existing 6 meta-analysis studies on HbA_{1c}, body weight, and BMI shows an overall effect of −0.79 (95% CI −1.17 to −0.42; I^2=90.5), which shows a pooled favorable effect of these mHealth interventions.

Discussion

Principal Findings

Although there is a strong interest among researchers, health care workers, and patients in mHealth interventions for the treatment of diabetes and obesity, overall, very little is known about its effectiveness. Moreover, at present, the use of mHealth interventions for these conditions is limited. To our knowledge, this is the first study that provides a comprehensive summary of research assessing the effectiveness of mHealth interventions for these conditions.

Published research has yielded mixed results. Examining evidence reported in 17 reviews that assessed a total of 55,604 original studies, this systematic review found that, overall, the impact of mHealth interventions on diabetes and obesity management is promising, especially in the areas of glycemic control and weight management. The majority of the 17 reviews focused on the self-monitoring functions of mHealth. Text messaging and apps were the primary types of mHealth interventions utilized to date. There was heterogeneity in the effectiveness of mHealth as diverse health outcomes (eg, blood pressure, weight, lipids, HbA_{1c}, clinical biomarkers, treatment adherence, and health-related behavior changes) were tested in the original studies, but only a few studies with various study designs and populations (eg, clinical trials, nonclinical trials, and diverse patient subgroups by severity and disease type) and study focus (eg, incentive-driven technology) were available in the review. Nevertheless, all the 17 reviews concluded that mHealth was feasible and potentially can improve health outcomes among patients suffering from diabetes and/or obesity.

Sources of Variations in Existing Research

Clinical biomarkers such as glycemic control and weight change were the primary focus in evaluating the effect of mHealth interventions in the reviews assessed. For example, the change in HbA_{1c} pre- and postintervention was evaluated in 10 reviews. Of these, 7 reviews reported statistically significant/large improvements, but 3 reviews did not; 2 meta-analyses showed 0.25% to 0.48% greater changes in HbA_{1c} following an mHealth intervention compared with standard diabetes care. In contrast, only 4 reviews found some improvement in treatment adherence in all 7 reviews that assessed it. Furthermore, small or insignificant improvements in health-related behaviors were reported in 9 reviews.

Several factors could have caused substantial heterogeneity among the assessment of clinical biomarkers, treatment adherence, and health-related behaviors. First, a small number of original studies examining treatment adherence and behavioral changes might have underpowered the systematic approach in the literature review. Second, the inclusion criteria for the study design (eg, clinical/nonclinical trial and quasi-experimental study), study subjects (eg, mixture of patients with T1DM and T2DM, patients with T1DM only, or poorly controlled patients with diabetes), and application type were not controlled efficiently in the previous reviews. In addition, patient health–related behaviors may require more time to change than was generally allowed in the studies compared with the typically more rapid change in biomarkers, possibly because of the influence of cognitive biases, habits, and social behavioral norms [43].

Implementation and Dissemination of Mobile Health Interventions

In recent years, there has been explosive growth in the number of mobile apps [13,44], including mHealth apps. However, the large number of available mHealth apps may hinder the intended use of these apps [20,27,45]. Limited guidance and the commerce-influenced nature of internet-based searches make it difficult for patients to determine which apps could most effectively help manage their health conditions. The number of app functions has also been negatively correlated with user ratings [46]. Thus, a process of a truly objective app review has the potential to improve the ability of patients to find the appropriate apps that meet their needs and preferences [47]. In addition, this can also help health care providers in making clear recommendations of the best apps to use.

As education and health promotion can favorably influence clinical outcomes, app developers need to fully consider the needs of users in designing features for patients suffering from diabetes/obesity. For example, self-management should be promoted as a key feature in apps targeting patients with T1DM who may need to check their blood glucose level more frequently than those with T2DM. In addition, new mobile messaging services, such as Facebook Messenger, WhatsApp, Snapchat, and Instagram, now exceed the functionality of traditional text messaging. Relevantly, social media features are increasingly popular, particularly among young people. Social networks can help patients achieve behavioral changes by, for instance, providing peer support among patients with similar conditions [48]. Strategies targeting behavioral changes to enhance self-management for patients are not very common among existing apps [49-52]. Thus, mHealth apps could implement social media and network features to more effectively target young users and improve their care [53,54]. Of course, support and adoption of mHealth approaches for the treatment of diabetes and obesity by health care providers will be helpful in maximizing the potential future value of mHealth interventions.

Limitations of Previous Reviews and the Original Mobile Health Intervention Studies

The 17 reviews and the included intervention studies share some limitations. First, some reviews only included a small number

of studies but examined a relatively large number of outcomes [23,27-29]. Second, most of the reviews examined studies conducted in developed countries; few reviews examined those from developing countries [21,28,31]. Third, a heterogeneous study design may cause substantial heterogeneity in meta-analyses or make it difficult to conduct a quantitative analysis across studies [23,28-31,33,37]. Fourth, the reviews included only a small number of RCTs [27,28,31,33,34], and many of the RCTs had short intervention periods [29,31,33,37]. Fifth, only a limited number of original studies reported changes in biomarkers, which may hinder the evaluation of mHealth interventions' clinical impact, especially on blood lipids and blood pressure [23,24]. Finally, the diverse features of the actual mHealth interventions further increase study heterogeneity [24].

Limitations of This Study

First, we examined the results reported in the 17 identified reviews without analyzing the findings from the original studies. Second, there was a high level of heterogeneity in the characteristics and findings of the 17 reviews. Thus, it was challenging to adequately interpret the effectiveness of mHealth interventions across reviews because of different study designs, objectives, and settings. Despite these limitations, this study provided a higher level of analysis and a comprehensive summary of the findings in the growing mHealth field. Compared with previous studies, our study has a number of unique contributions, including the following: (1) our study added quantitative evidence specifically on the applications of mHealth in diabetic and obesity care research and studied objective changes in biomarkers, treatment adherence, and health behaviors after an mHealth intervention, whereas previous studies were general and narratively described mHealth effects on diverse diseases using a small number of articles with low quality; (2) we conducted a meta-analysis on the intervention effects of clinical outcomes, which was lacking in the existing reviews; and (3) our review included newly published reviews that were not included in other studies. This helps identify best practices for fighting the epidemics of diabetes and obesity. In addition, we found a fairly consistent reduction in HbA_{1c} and body weight from mHealth interventions across multiple reviews.

Recommendations for Future Research

Regarding future evaluations of mHealth interventions, more rigorous study designs and strategies are needed to enable us to draw more precise and specific conclusions regarding their effectiveness for diabetes and obesity management. To enhance app design, including user ratings and experiences may be useful in developing evidence-based strategies. The level to which users truly engage with these mHealth apps is not yet clear. Patient-centered self-monitoring with personalized feedback is important in behavioral change and has been shown to improve user engagement and adherence [55]. Designing app functions relevant to the users on the basis of their age and sex, type of diabetes, and geographical location would improve the targeting and effectiveness of mHealth interventions.

To promote an evidence-based approach in mHealth use for diabetes and obesity management, multiple validation tests and, when appropriate, regulations will be needed. Objective and validated measures should be used, in particular, when studying behavioral changes following mHealth interventions. Furthermore, there is a need to identify and focus on high-risk groups (eg, low socioeconomic status populations), as most previous reviews did not include studies conducted in these populations.

In conclusion, findings from the 17 reviews, including 6 meta-analyses published since 2005, suggested promising but limited evidence on the effectiveness of mHealth interventions for diabetes and obesity management. Self-management, monitoring, and use of text messaging and apps are the primary target functions and application types of mHealth investigated in the field. More rigorous study designs should be applied in future studies for assessing the impact of mHealth interventions on diabetes and obesity management. To enhance the effectiveness of mHealth interventions, studies are warranted for the optimal formats and the frequency of contacting patients, using theory-based interventions; for the better tailoring of messages to the specific needs and communication style of recipients; and for enhancing the usability by adapting approaches to recipients with varying degrees of technological and health literacy, thus placing a greater emphasis on maintaining effectiveness over time.

Acknowledgments

This work was supported by a research grant from the US National Institutes of Health, the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U54HD070725), which YW received while he worked at the Johns Hopkins University and established a global center of excellence. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the funder.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological quality of 17 studies based on AMSTAR2 criteria.
[DOCX File , 16 KB - mhealth_v8i4e15400_app1.docx]

Multimedia Appendix 2

Characteristics of 17 review studies on the effectiveness of mobile health interventions for diabetes and obesity management.
[DOCX File , 25 KB - mhealth_v8i4e15400_app2.docx]

Multimedia Appendix 3
Summary of findings from the 17 reviews on the effectiveness of mobile health interventions for diabetes and obesity management.
[DOCX File , 21 KB - mhealth_v8i4e15400_app3.docx]

References

1. Lim H, Nguyen T, Choue R, Wang Y. Sociodemographic disparities in the composition of metabolic syndrome components among adults in South Korea. Diabetes Care 2012 Oct;35(10):2028-2035 [FREE Full text] [doi: 10.2337/dc11-1841] [Medline: 22837361]
2. Wang Y, Wang L, Qu W. New national data show alarming increase in obesity and noncommunicable chronic diseases in China. Eur J Clin Nutr 2017 Jan;71(1):149-150. [doi: 10.1038/ejcn.2016.171] [Medline: 27703162]
3. World Health Organization. 2018 Oct 30. Diabetes URL: http://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2018-07-24]
4. World Health Organization. 2018 Feb 16. Obesity and Overweight URL: http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2018-07-24]
5. World Health Organization. 2018. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016 URL: https://www.who.int/healthinfo/global_burden_disease/estimates/en/ [accessed 2018-07-24]
6. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes Self-management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Clin Diabetes 2016 Apr;34(2):70-80 [FREE Full text] [doi: 10.2337/diaclin.34.2.70] [Medline: 27092016]
7. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016 Apr 9;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
8. McKinsey & Company. 2014 Oct. Overcoming Obesity: An Initial Economic Analysis URL: https://www.mckinsey.com/~/media/McKinsey/Business [accessed 2018-07-24]
9. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol 2011 Nov 8;8(4):228-236. [doi: 10.1038/nrendo.2011.183] [Medline: 22064493]
10. Gortmaker SL, Swinburn BA, Levy D, Carter R, Mabry PL, Finegood DT, et al. Changing the future of obesity: science, policy, and action. Lancet 2011 Aug 27;378(9793):838-847 [FREE Full text] [doi: 10.1016/S0140-6736(11)60815-5] [Medline: 21872752]
11. Jones H, Edwards L, Vallis TM, Ruggiero L, Rossi SR, Rossi JS, Diabetes Stages of Change (DiSC) Study. Changes in diabetes self-care behaviors make a difference in glycemic control: the Diabetes Stages of Change (DiSC) study. Diabetes Care 2003 Mar;26(3):732-737. [doi: 10.2337/diacare.26.3.732] [Medline: 12610030]
12. Collier A, Ghosh S, Hair M, Waugh N. Impact of socioeconomic status and gender on glycaemic control, cardiovascular risk factors and diabetes complications in type 1 and 2 diabetes: a population based analysis from a Scottish region. Diabetes Metab 2015 Apr;41(2):145-151. [doi: 10.1016/j.diabet.2014.09.004] [Medline: 25454092]
13. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
14. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
15. Klonoff DC. The current status of mHealth for diabetes: will it be the next big thing? J Diabetes Sci Technol 2013 May 1;7(3):749-758 [FREE Full text] [doi: 10.1177/193229681300700321] [Medline: 23759409]
16. Kay M, Santos J, Takane M. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011:;64(7):66-;64(7):71.
17. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. Q J Med 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
18. Rho MJ, Kim HS, Chung K, Choi IY. Factors influencing the acceptance of telemedicine for diabetes management. Cluster Comput 2015;18(1):321-331. [doi: 10.1007/s10586-014-0356-1]
19. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
20. Tamony P, Holt R, Barnard K. The role of mobile applications in improving alcohol health literacy in young adults with type 1 diabetes: Help or hindrance? J Diabetes Sci Technol 2015 Aug 6;9(6):1313-1320 [FREE Full text] [doi: 10.1177/1932296815588559] [Medline: 26251369]

21. Wang X, Shu W, Du J, Du M, Wang P, Xue M, et al. Mobile health in the management of type 1 diabetes: a systematic review and meta-analysis. BMC Endocr Disord 2019 Feb 13;19(1):21 [FREE Full text] [doi: 10.1186/s12902-019-0347-6] [Medline: 30760280]
22. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
23. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
24. de Ridder M, Kim J, Jing Y, Khadra M, Nanan R. A systematic review on incentive-driven mobile health technology: as used in diabetes management. J Telemed Telecare 2017 Jan;23(1):26-35. [doi: 10.1177/1357633X15625539] [Medline: 26888421]
25. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
26. Dobson R, Whittaker R, Pfaeffli Dale L, Maddison R. The effectiveness of text message-based self-management interventions for poorly-controlled diabetes: a systematic review. Digit Health 2017;3:2055207617740315 [FREE Full text] [doi: 10.1177/2055207617740315] [Medline: 29942620]
27. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
28. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
29. Mallow JA, Theeke LA, Barnes ER, Whetsel T, Mallow BK. Using mHealth tools to improve rural diabetes care guided by the chronic care model. Online J Rural Nurs Health Care 2014;14(1):43-65 [FREE Full text] [doi: 10.14574/ojrnhc.v14i1.276] [Medline: 26029005]
30. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012 Sep 1;6(5):1185-1196 [FREE Full text] [doi: 10.1177/193229681200600524] [Medline: 23063046]
31. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
32. Park SH, Hwang J, Choi YK. Effect of mobile health on obese adults: a systematic review and meta-analysis. Healthc Inform Res 2019 Jan;25(1):12-26 [FREE Full text] [doi: 10.4258/hir.2019.25.1.12] [Medline: 30788177]
33. Mateo GF, 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]
34. Khokhar B, Jones J, Ronksley PE, Armstrong MJ, Caird J, Rabi D. Effectiveness of mobile electronic devices in weight loss among overweight and obese populations: a systematic review and meta-analysis. BMC Obes 2014;1:22 [FREE Full text] [doi: 10.1186/s40608-014-0022-4] [Medline: 26217509]
35. Bhardwaj NN, Wodajo B, Gochipathala K, Paul DP, Coustasse A. Can mHealth revolutionize the way we manage adult obesity? Perspect Health Inf Manag 2017;14(Spring):1a [FREE Full text] [Medline: 28566984]
36. Darling KE, Sato AF. Systematic review and meta-analysis examining the effectiveness of mobile health technologies in using self-monitoring for pediatric weight management. Child Obes 2017 Oct;13(5):347-355. [doi: 10.1089/chi.2017.0038] [Medline: 28471699]
37. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409 [FREE Full text] [doi: 10.1111/ijpo.12002] [Medline: 25641770]
38. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
39. Paré G, Trudel MC, Jaana M, Kitsiou S. Synthesizing information systems knowledge: a typology of literature reviews. Inf Manag 2015;52(2):183-199. [doi: 10.1016/j.im.2014.08.008]
40. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Br Med J 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
41. Wilhelmsen NC, Eriksson T. Medication adherence interventions and outcomes: an overview of systematic reviews. Eur J Hosp Pharm 2019 Jul;26(4):187-192. [doi: 10.1136/ejhpharm-2018-001725] [Medline: 31338165]

42. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
43. Tronieri J, Wadden T, Diplomate AB, Alamuddin MB. NCBI - NIH. South Dartmouth (MA): MDText.com, Inc; 2000. Endotext: Behavioral Approaches to Obesity Management URL: https://www.ncbi.nlm.nih.gov/books/NBK278952/ [accessed 2019-07-24]
44. International Telecommunication Union (ITU). 2016. The World in 2016: ICT Facts and Figures URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [accessed 2019-07-24]
45. Bailey CJ, Kodack M. Patient adherence to medication requirements for therapy of type 2 diabetes. Int J Clin Pract 2011 Mar;65(3):314-322. [doi: 10.1111/j.1742-1241.2010.02544.x] [Medline: 21314869]
46. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 9;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
47. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
48. Laranjo L, Arguel A, Neves A, Gallagher A, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
49. 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]
50. 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]
51. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
52. Payne H, Moxley V, MacDonald E. Health behavior theory in physical activity game apps: a content analysis. JMIR Serious Games 2015 Jul 13;3(2):e4 [FREE Full text] [doi: 10.2196/games.4187] [Medline: 26168926]
53. Giménez-Pérez G, Recasens A, Simó O, Aguas T, Suárez A, Vila M, et al. Use of communication technologies by people with type 1 diabetes in the social networking era. A chance for improvement. Prim Care Diabetes 2016 Apr;10(2):121-128. [doi: 10.1016/j.pcd.2015.09.002] [Medline: 26428527]
54. Vaala SE, Hood KK, Laffel L, Kumah-Crystal YA, Lybarger CK, Mulvaney SA. Use of commonly available technologies for diabetes information and self-management among adolescents with type 1 diabetes and their parents: a web-based survey study. Interact J Med Res 2015 Dec 29;4(4):e24 [FREE Full text] [doi: 10.2196/ijmr.4504] [Medline: 26715191]
55. Winter SJ, Sheats JL, King AC. The use of behavior change techniques and theory in technologies for cardiovascular disease prevention and treatment in adults: a comprehensive review. Prog Cardiovasc Dis 2016;58(6):605-612 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
mHealth: mobile health
PA: physical activity
PDA: personal digital assistant
RCT: randomized controlled trial
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 08.07.19; peer-reviewed by A Mehdizadeh, R Sun; comments to author 09.09.19; revised version received 04.11.19; accepted 26.01.20; published 28.04.20.

Please cite as:
Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, J Cheskin L
Effectiveness of Mobile Health Interventions on Diabetes and Obesity Treatment and Management: Systematic Review of Systematic Reviews
JMIR Mhealth Uhealth 2020;8(4):e15400
URL: http://mhealth.jmir.org/2020/4/e15400/
doi:10.2196/15400
PMID:32343253

Original Paper

Cyberphysical Human Sexual Behavior Acquisition System (SeBA): Development and Implementation Study in China

Xiaoping Zhou[1], PhD; Jichao Zhao[2], BEng; Xun Liang[3], PhD

[1]Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing University of Civil Engineering and Architecture, Beijing, China
[2]China National Petroleum Corporation Managers Training Institute, Beijing, China
[3]School of Information Science, Qufu Normal University, Qufu, China

Corresponding Author:
Xun Liang, PhD
School of Information Science
Qufu Normal University
#57 Jingxuan West Road
Qufu, 273165
China
Phone: 86 10 68322509
Email: xunliangruc@163.com

Abstract

Background: Sexual health is one of the principal components of human well-being. Traditional methods for observing human sexual behavior typically adopt manual intervention approaches (eg, interviews). However, the data obtained by such traditional approaches suffer from intrinsic bias and limited sample sizes. Sexual behavioral data that are more reflective of the actual situation can be collected by equipping sex toys with sensors.

Objective: To address the limitations of traditional human sexual behavior data observation methods, a novel cyberphysical system is proposed to capture natural human sexual behavior data in China at the nationwide level.

Methods: A cyberphysical human sexual behavior acquisition system (SeBA) was designed and implemented. SeBA jointly utilizes state of the art information and communication technologies such as smart sex toys, smartphones, and mobile social networks. Smart sex toys enable objective collection of data on human sexual behavior, while the mobile social network provides the possibility of partnered sex in a cyberphysical manner. The objectives and function settings are discussed, and the overall framework of the system architecture is presented.

Results: Operation and privacy policies are proposed and the technical solution of SeBA is described. The effectiveness of SeBA was verified based on analysis of users' human sexual behavior data collected from January 2016 to June 2017. A total of 103,424 solo sexual behaviors were recorded involving 13,047 users, and 61,007 partnered sexual behaviors from 7,140 users were observed. The proportions of males and females in the solo and partnered sex groups were fairly consistent with recent statistics on unmarried individuals in China. We also found that only a small portion of individuals provided information on at least one other attribute besides the required input of gender, such as age, height, location, job, sex preferences, purposes, and interests.

Conclusions: To the best of our knowledge, this is the first study to analyze objective human sexual behavior data at the nationwide level. Although the data are restricted to China, this study can provide insight for further research on human sexual behavior based on the huge amount of data available from wireless smart sex toys worldwide. It is anticipated that findings from such objective big data analyses can help deepen our understanding of sexual behavior, as well as improve sexual health and sexual wellness.

(JMIR Mhealth Uhealth 2020;8(4):e12677) doi:10.2196/12677

KEYWORDS

cyberphysical system; sexual behavior; smart sex toys; mobile social network

Introduction

As one of the principal components of human well-being, human sexual health has been studied extensively [1,2]. Human sexual behavior is the manner in which humans experience and express their sexuality, and is a crucial factor to measure and study human sexual health [3,4]. Due to the lack of large-scale objective human sexual behavior data, much remains to be uncovered in the field of human sexual health. Human sexual behavior acquisition is the process of measuring human sexual behavior, including both solo sex and partnered sex. Thus, the development of large-scale objective human sexual behavior acquisition systems can provide new insight into the sexual lives of people. It is anticipated that a human sexual behavior acquisition system capable of capturing real-life sexual data can advance current human sexual research.

Human sexual behavior acquisition is technically challenging. Owing to concerns of sexual privacy, few individuals are willing to be observed during their sexual acts. Therefore, current solutions for capturing sexual behavior data are based on subjective reports and interviews [5-8]. However, individuals are seldom concerned about sexual behavior during an exhilarating session of lovemaking and are therefore likely to be biased (eg, overestimating time). Moreover, the number of subjects included in interviews is generally limited. For clinical purposes, some patients may allow being inspected during their sexual acts [9-11], which could somewhat affect the mood and might not exactly reflect the natural flow of sex. These problems trigger demands for new schemes to objectively measure sexual behavior.

Here, we present a new scheme, termed cyberphysical human **Se**xual **B**ehavior **A**ctivation system (SeBA), which was designed to capture natural sexual behavior. SeBA jointly utilizes state of the art information and communication technologies, including Internet of Things (IoT), big data, smart devices, cloud computing, and mobile social networks. Although the primary objective was to collect data on objective sexual behaviors with smart sex toys at a nationwide level in China, it is also possible to obtain more data from the social perspective (eg, individuals' sexual orientation) [12]. Ultimately, more knowledge on sexual health can be discovered by big data–driven approaches using such large-scale objective human sexual behavior data, thereby contributing to complementing theories of sexuality, improving societal knowledge of the prevalence of sexual behaviors [13], and providing benefits to clinicians working to improve sexual health [14,15].

As a natural human behavior [16], people are engaging in sexual activities at all times in all places worldwide. However, much remains to be uncovered on human sexual behavior due to the lack of objective data. People engage in a variety of sexual acts, ranging from solo sex to partnered sex in varying patterns of frequency for a wide variety of reasons. Currently, no promising solution capable of capturing large-scale objective sexual behavior data is available. Thus, we designed SeBA as a novel objective sexual behavioral data acquisition system.

Sex toys such as dildos and vibrators are primarily used as masturbation tools for a single individual to facilitate their sexual pleasure. By employing sex toys, individuals can reach orgasm naturally without possible psychological effects. In other words, the sexual behavior data from sex toys could be much more objective than those obtained from interviews. Consequently, sexual behavioral data that are more reflective of the actual situation can be collected by equipping sex toys with sensors. This is the primary motivation of the proposed SeBA. A smart device equipped with various sensors is a powerful tool to sense all types of human behaviors, and has been extensively used in a variety of areas [17]. For example, smart sex toys [18] enable the collection of users' sexual behaviors without their awareness, thus providing sexual behavior data in a natural way. However, several challenges remain to obtain sexual data from smart sex toys.

The first and primary problem is the challenge in obtaining human-human sexual behavior data from smart sex toys. Usually, sex toys, including smart sex toys and sex robots [19], are designed for masturbation. Individuals enjoy solo sex in a private space, whereas partnered sex is remarkably different. Therefore, exploring partnered sex can reveal more insight on human sexual behavior. Consequently, breaking through the space limitation of sex toys is a crucial issue.

Second, the recorded sexual behavior data should be sent back to the data center through some form of communication technology. Although sensors are embedded in sex toys to sense accurate human sexual behavior, these data are only stored in the smart sex toys. Undoubtedly, the sex toys, as well as the sexual behavior data, are the properties of users, and third parties cannot collect these data legally. Moreover, the technology of transferring the sexual behavior data from sex toys to the data center also requires the users' permission.

Third, it is difficult to attract nationwide individuals to participate in such studies. One of the limitations of current solutions is the limited number of samples. Therefore, recruiting more individuals can result in more accurate findings on human sexual behavior.

This study proposes a novel framework, SeBA, to address the above issues.

Methods

Function Settings

A smartphone is a handheld personal computer with extensive computing capabilities, including high-speed access to the internet using either wifi or a mobile broadband connection. Moreover, smartphones include support for Bluetooth connectivity. With continuing progress in smartphone and mobile network technologies, the physical distance limitation of communications among individuals can be avoided by providing a cyberspace for connectivity. Currently, mobile social networks are pervasively used in our daily lives. Therefore, we employed a mobile social network as a bridge among smart sex toys through the internet in the development of SeBA.

The mobile social network embedded in SeBA is called SNApp. SNApp can be downloaded online and installed on a

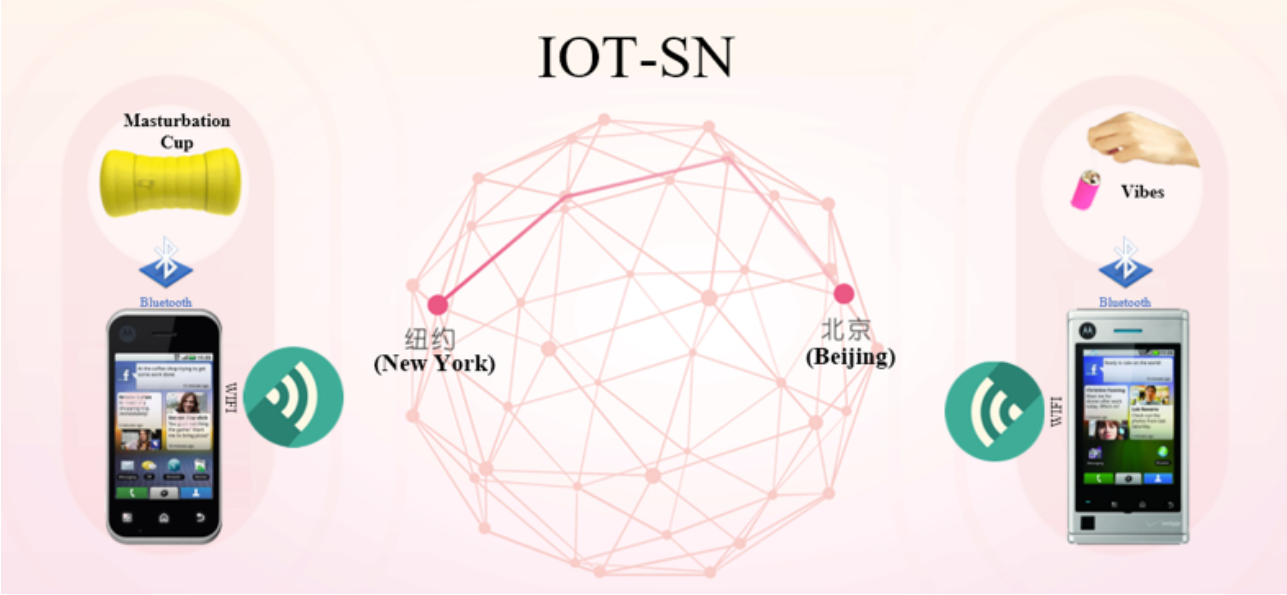
smartphone, allowing smartphones to connect to sex toys through Bluetooth and to the internet through wifi or mobile broadband. Moreover, individuals can connect with others on SNApp. Once a relationship is built in the social network, the linked users can interoperate each others' sex toys, thus enabling partnered sex through SNApp. To some extent, this kind of partnered sex can be treated as typical human-human sex.

With the advantages of SNApp, SeBA solves the three main issues mentioned above. Since the cyberspace in SNApp enables users to contact other users in a private physical space, they can profit from sexual pleasure with assurance of privacy protection. Undoubtedly, SNApp is fundamentally a tool for human-human sex. Moreover, the user behavior data, as well as sexual behavior data, can be transmitted through the mobile network to the data center with the users' permission. Theoretically, individuals worldwide can register and enjoy sexual pleasure in the same platform. From this perspective, SeBA can enable data collection not only at a nationwide but also at a worldwide scale.

SNApp has some distinct features from a traditional mobile social network. Each node in a traditional mobile social network represents a user. However, each node in SNApp connects not only the user but also a smart sex toy. The mobile social network provides the cyberspace, while smart sex toys represent the physical aspect of SeBA. Otherwise, SeBA is a typical cyberphysical system.

Figure 1 presents the typical usage scenario of SeBA. As discussed above, users should have both smart sex toys and SNApp to enable partnered sexual behavior. Smart sex toys are devices that can be controlled via Bluetooth using SNApp. Smart sex toys can be produced in the same forms and functions as traditional sex toys with a key difference being that they include embedded chips enabling smart control from SNApp. Therefore, smart sex toys are based on the IoT concept. SNApp is the core component of SeBA. The functions of SNApp include a user function module (registration, user profile management), user-friendly function module (online user list, friend connection request/response), user communication function module (message communication, partnered sex communication), and sex toy management function module (smart sex toy connection and control). Thus, SNApp is based on a social network, which enables users to cultivate partnered sexual behavior online.

Figure 1. Typical usage scenario of sexual behavior acquisition system (SeBA).



From the perspective of users, SeBA provides a platform or service to facilitate human sexual pleasure. Smart sex toys allow users to enjoy more pleasure from masturbation. Moreover, the mobile social network function enables users to find online partnered sex safely, offering almost the same feeling as penetrative intercourse. SeBA comprehensively utilizes IoT and mobile social network technologies. It is a typical cyberphysical system with the capability to sense natural human sexual behavior.

Data Processing

To analyze the data collected from SeBA, the data were first exported and stored in a MySQL database [20]. The user statistical data were then queried using structured query language. Finally, the selected user statistical data were imported into Microsoft Excel 2016 for visualization.

Results

Operation and Privacy Policies

Since SeBA aims to obtain objective sexual behavioral data, the functions and operation policies focus on satisfying the users' requirements. That is, the users pay more attention to the pleasures from the online partnered sex, with little awareness about the data problem. From this perspective, the products, including both the smart sex toys and SNApp, were designed to achieve a user-friendly experience. Moreover, unlike many other health care studies conducted by governments or hospitals, SeBA operates exclusively on the internet, which enables collecting more data from more areas.

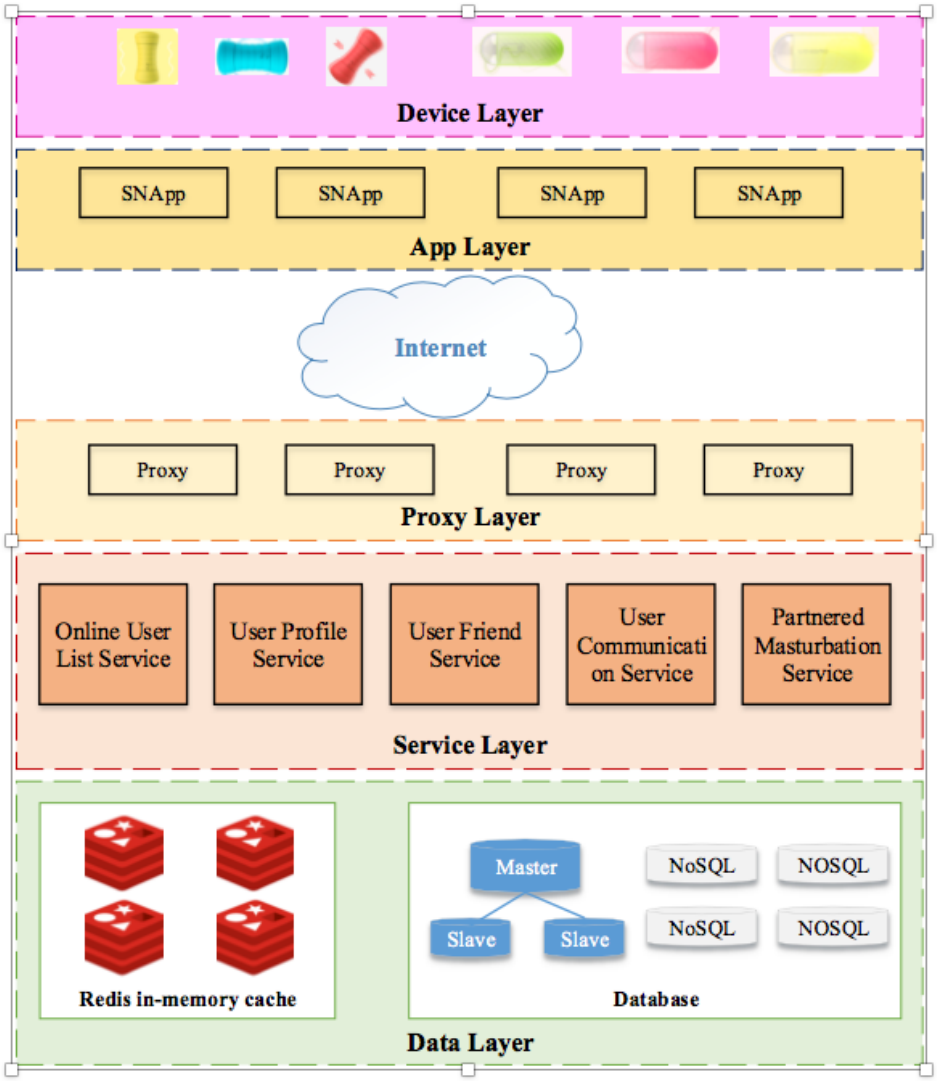
Sex is one of the most important aspects of personal privacy. Even when an individual considers their sexual behavior to be perfectly "normal" and they have nothing to hide, they typically

regard their sex lives as private. Subsequently, all of the data collected are highly private and the details of the data are encrypted to protect the leakage of users' behavioral data.

System Architecture

The system architecture can be roughly divided into two categories according to the users' accessibility: local and the cloud. Figure 2 presents the overall framework of the system architecture.

Figure 2. The overall system architecture of the sexual behavior acquisition (SeBA) system.

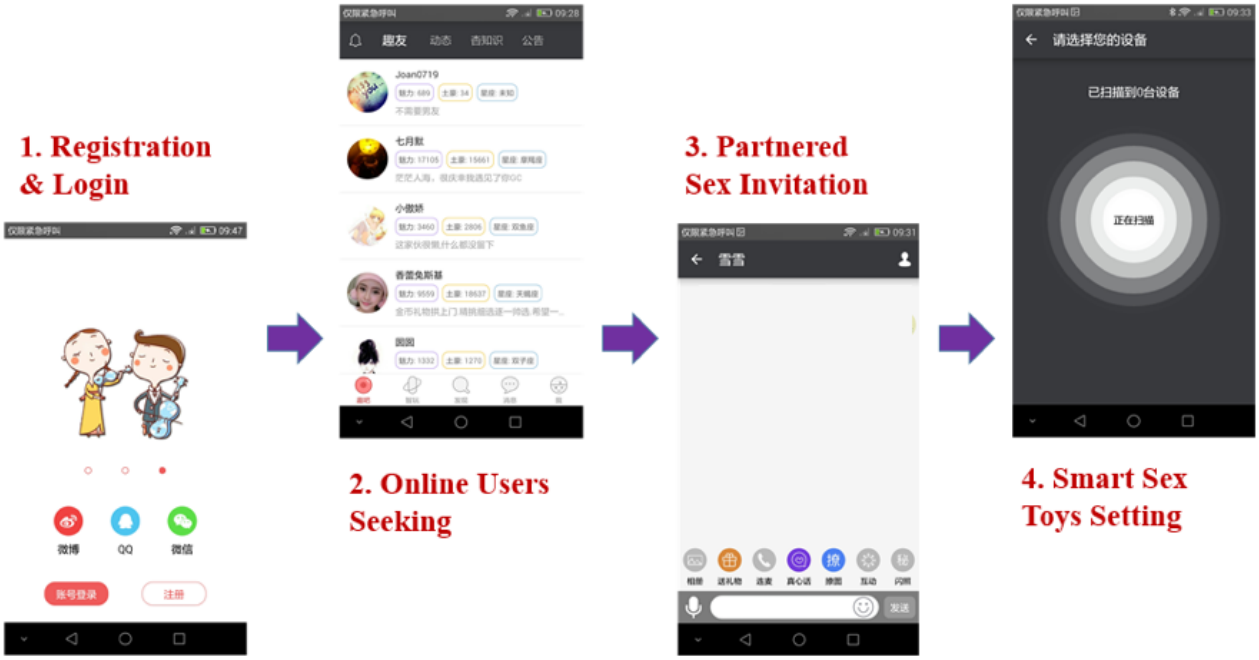


The local part is composed of the hardware and software that users use. It mainly consists of the device layer and the app layer from the technical aspect. The device layer includes all types of smart sex toys, while the app layer is SNApp installed in the users' smartphones. SNApp could be downloaded and installed from the Wolkamo Tech Co Ltd website [21].

The device layer consists of smart sex toys. Two main types of smart sex toys are produced: a smart masturbation cup and a smart vibrator. For the users, the functions of both smart toys are indistinguishable from a traditional masturbation cup and vibrator. Additionally, the smart masturbation cup and smart vibrator can be controlled by SNApp through a Bluetooth connection. Both smart sex toys are produced by Wolkamo Tech Co Ltd (Beijing, China). Since there are no international or national standards for smart sex toy, other brands of smart sex toys are not yet compatible with SeBA.

The SNApp layer is the main user interface. SNApp can be used to control the smart masturbation cup or vibrator, and also to find a sex partner online. Users must register and then sign into SNApp to use its functions. After signing into the app, the user is marked as online and can be found by any other online users on the app. Online users can connect their smart sex toys to SNApp through Bluetooth on the smartphone. After connection, the users can start, control, and stop the sex toys in single sex or partnered sex mode by inviting another online user with a smart sex toy. In partnered sex, the two users can interact with words, pictures, and real-time voice and control their counterparts' smart sex toys. Figure 3 presents the basic steps of SeBA.

Figure 3. Basic usage process of sexual behavior acquisition system (SeBA).



The cloud part is the cloud services provided to support the main functions for users. It consists of three components: proxy layer, service layer, and data layer.

The proxy layer functions in three modes. First, it enables the scale-out load balance. High concurrency is the main feature of online services. The proxy layer dispatches the requests into different services in the service layer to address the high concurrency problem. Second, data compression and data encryption are provided in the proxy layer to avoid data leakage. Third, the proxy layer can act as a firewall for the cloud service, since the outer computers cannot directly connect to the computers serving the data and service layers.

The service layer provides all of the app programming interfaces for SNApp to the data layer with permissions. The main services in the service layer include an online user list service, user profile service, user friend service, user communication service, and partnered sex service. These services correlate to the function settings of SNApp.

The data layer utilizes an in-memory database to cache the real-time retrieval data, and employs a disk-based NoSQL database to store persistent data. Usually, online users are recorded in the in-memory database, since these data are frequently used to ease the users in seeking online sex partners. Currently, all other data are stored in the disk-based database.

Effectiveness

The system was primary launched online on January 1, 2016. Up to June 8, 2017, 103,424 single sexual behaviors were recorded, involving 13,047 users, including 77.85% owning sex toys. Table 1 provides the general statistics of the users, demonstrating a greater proportion of males than females, which were completely consistent with the proportions of sex toy-owning users. With respect to partnered sex, 61,007 sexual behaviors were collected, with a slightly higher proportion of male users. We verified that these percentages were fairly consistent with recent statistics on unmarried individuals in China from the National Bureau of Statistics of China, which reported that among the 182,568 unmarried people more than 15 years of age (sampling rate 0.837‰), 107,984 (59.15%) were males and 74,584 (40.85%) were females [22].

Table 1. Summary of user statistics of the dataset.

User type	Total users (N)	Males (n, %)	Females (n, %)
Registered	210,104	176,904 (84.20%)	33,200 (15.80%)
Sex Toy Owners	16,760	11,037 (65.85%)	5723 (34.15%)
Solo Sex	13,047	8559 (65.60%)	4488 (34.40%)
Partnered Sex	7140	3973 (55.64%)	3167 (44.36%)

Gender is a mandatory requirement when registering an account on SeBA. SeBA explicitly reminds the user on the user registration page that their chosen gender cannot be revised after selected. It was assumed that most users will choose to be honest about their gender to find appropriate sex partners according to their preferences on SeBA.

The other attributes, including age, height, location, job, sex preferences, purposes, and interests, can be entered and revised

at any time in the app setting page. Interestingly, only a small portion of individuals included details on any of these other attributes besides the required gender field. Table 2 summarizes the numbers of sex toy–owning users according to the attributes provided.

Table 2. Sex toy–owning users according to attribute information provided.

Attribute provided	Total sex toy users (n, % of total)	Males (n, % of total sex toy users)	Females (n, % of total sex toy users)
Height	508 (3.03%)	392 (77.2%)	116 (22.8%)
Age	485 (2.89%)	376 (77.5%)	109 (22.5%)
Location	427 (2.55%)	344 (80.6%)	83 (19.4%)
Job	261 (1.56%)	204 (78.2%)	57 (21.8%)

With respect to age, most SeBA users were found to be young, with a maximum age of 30 years among the 485 users who supplied an age. Figure 4 presents the distribution of the age of these 485 users, showing a predominance of users between 22 and 28 years old. The average heights of the male and female users were 176.6 cm and 163.46 cm, respectively. Figure 5 displays the distribution of the height of the 508 users that provided this information. The height of most male users in SeBA ranged between 170 cm and 182 cm, whereas most female users reported a height between 160 cm and 170 cm. This observation is consistent with actual statistics in China [23]. Among the 427 users who entered a location, more than 20 provinces or cities in China included at least 5 individuals registered in SeBA. From this perspective, we consider that SeBA has the ability to sensor the sexual behavior data at a nationwide level. Table 3 lists the top 15 provinces or cities with the most users.

Figure 4. Distribution of user age in sexual behavior acquisition system (SeBA).

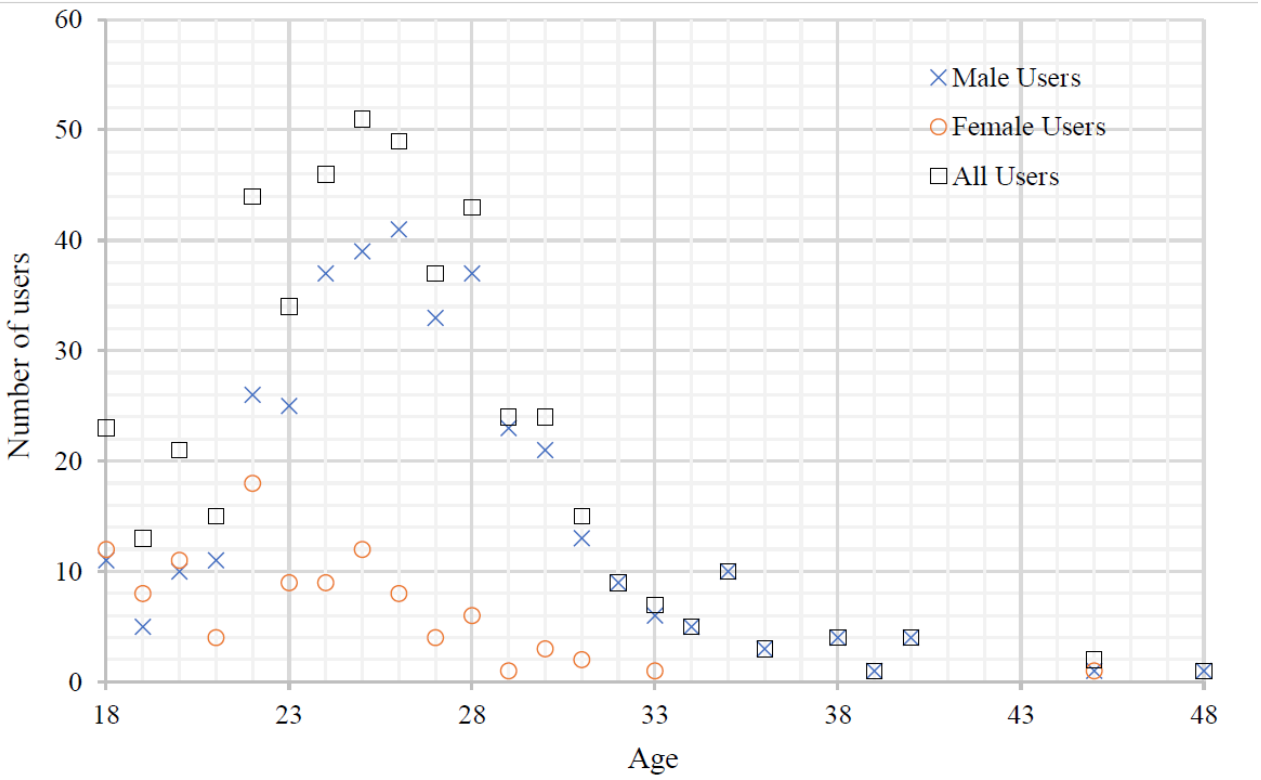


Figure 5. Distribution of height among sexual behavior acquisition system (SeBA).

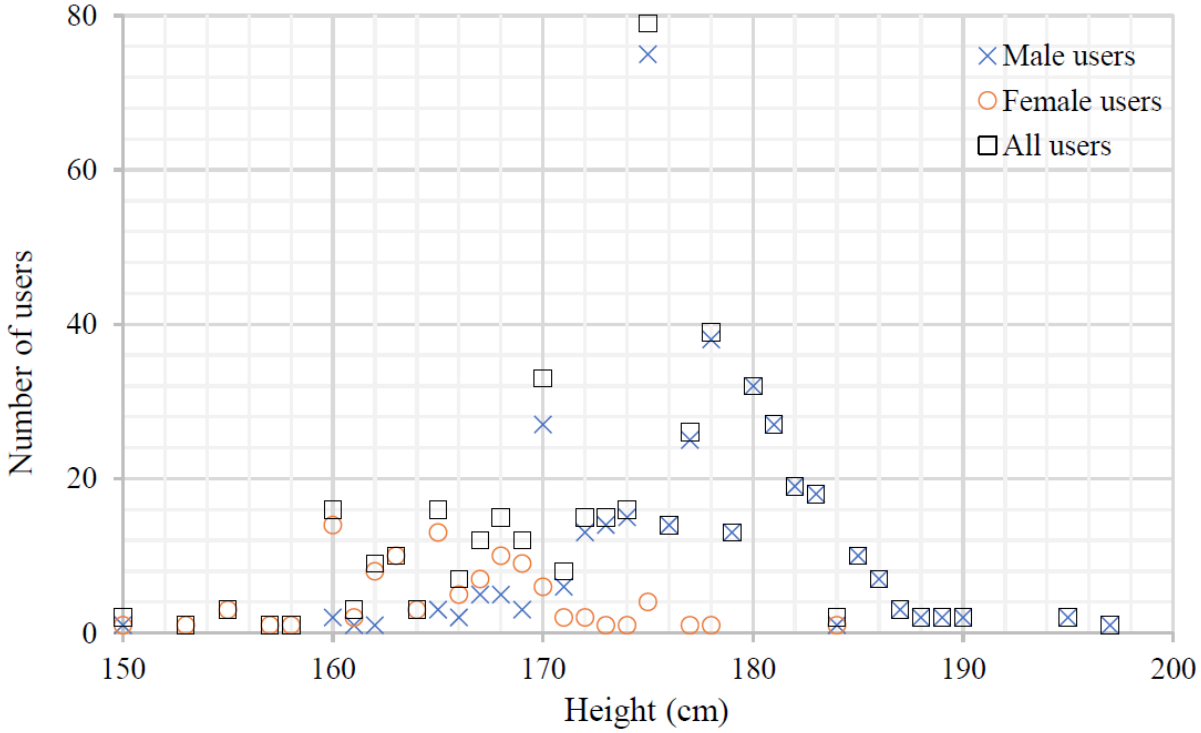


Table 3. Top-ranked locations of the 427 users that provided this attribute information.

Location	Male users (n)	Female users (n)	All users (n)
Beijing	56	17	73
Guangdong	49	10	59
Shanghai	25	3	28
Jiangsu	20	2	22
Zhejiang	14	3	17
Fujian	12	4	16
Hong Kong	9	5	14
Liaoning	13	1	14
Sichuan	11	1	12
Hubei	7	3	10
Hebei	9	1	10
Shanxi	9	1	10
Tianjin	9	0	9
Shaanxi	9	0	9
Macau	7	2	9

Of note, the large majority of users that provided additional attribute information were male (Table 2). Only a few females were willing to fill in additional information on their SeBA profiles. From this perspective, sexual desire plays a crucial role in online partnered sex, and most online sex behaviors are independent of age and location, but rather depend on gender selection. This finding partially confirms that users on SeBA are sexual satisfaction–oriented [24].

Discussion

Principal Findings

A novel cyberphysical human sexual behavior acquisition system (SeBA) was designed and implemented in China at a nationwide level. SeBA jointly utilizes state of the art information and communication technologies such as smart sex toys, smartphones, and a mobile social network. Operation and privacy policies and technical solutions of SeBA were presented,

followed by preliminary data processing of collected data. The effectiveness of SeBA, in terms of the ability to collect nationwide data on human sexual behavior, was verified by more than 1 year of user data. A total of 103,424 solo sexual behaviors were recorded involving 13,047 users, and 61,007 partnered sexual behaviors from 7140 users were observed. The proportions of males and females engaging in solo and partnered sex with the app were fairly consistent with recent statistics on unmarried individuals in China. We also found that only a small portion of individuals filled in the fields of one or more other attributes besides gender, including age, height, location, job, sex preferences, purposes, and interests.

Limitations

This is a pioneering study on collecting large-scale human sexual behavior data jointly using smart sex toys and mobile social networks. Admittedly, there are some limitations in the instrument used to collect sexual behavior. First, SeBA is currently only compatible with the smart sex toys produced by Wolkamo Tech Co Ltd due to the lack of international or national standards for smart sex toys. Second, the dataset collected may be limited to individuals who use smart sex toys and agree to share their toys on a social network. It is notable that most of the sex toys available are not yet smart. As more concerns on smart sex toys emerge, an increasing amount of sex toys are expected to be upgraded to a smart version, and thus smart sex toys from more brands will be compatible with SeBA. Third, the dataset may have failed to capture the intensities and placements of the smart sex toys in both solo and partnered sex users. Additionally, the sex that occurs via smart sex toys may not be the same as real human-human sex. With further development of smart sensors, more behavioral data can certainly be collected, resulting in more results in this area. Although the data were restricted to China, this study can provide insight for general sexual studies using a huge amount of data obtained from wireless smart sex toys worldwide.

Conclusions

This study presents a new paradigm for human sexual behavior studies by demonstrating that smart sex toys and mobile social networks can be used to obtain objective human sexual behavior data. It is anticipated that the findings from analyzing these more objective data can help deepen our understanding of human sexual behavior, as well as help to improve sexual health and sexual wellness [25-28]. This study may contribute to demystifying the ultimate secrets of human sexuality in the era of wireless IoT and big data [29].

Acknowledgments

This work was supported by the National Social Science Foundation of China (18ZDA309), National Natural Science Foundation of China (71531012, 71601013, 71271211), Natural Science Foundation of Beijing (4202017, 4172032), Youth Talent Support Program of Beijing Municipal Education Commission (CIT&TCD201904050), and Youth Talent Project of Beijing University of Civil Engineering & Architecture. The authors would like to thank Wolkamo Tech Co, Ltd for technical support and Baotong Mi for the statistical analysis of SeBA-derived data.

Conflicts of Interest

None declared.

References

1. Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, et al. Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). Lancet 2013 Nov;382(9907):1781-1794. [doi: 10.1016/s0140-6736(13)62035-8]
2. McClintock MK, Dale W, Laumann EO, Waite L. Empirical redefinition of comprehensive health and well-being in the older adults of the United States. Proc Natl Acad Sci U S A 2016 May 31;113(22):E3071-E3080 [FREE Full text] [doi: 10.1073/pnas.1514968113] [Medline: 27185911]
3. Oram S, Stöckl H, Busza J, Howard LM, Zimmerman C. Prevalence and risk of violence and the physical, mental, and sexual health problems associated with human trafficking: systematic review. PLoS Med 2012 May 29;9(5):e1001224 [FREE Full text] [doi: 10.1371/journal.pmed.1001224] [Medline: 22666182]
4. Vanden Broucke H, Everaert K, Peersman W, Claes H, Vanderschueren D, Van Kampen M. Ejaculation latency times and their relationship to penile sensitivity in men with normal sexual function. J Urol 2007 Jan;177(1):237-240. [doi: 10.1016/j.juro.2006.08.071] [Medline: 17162053]
5. Miller SA, Byers ES. Actual and desired duration of foreplay and intercourse: discordance and misperceptions within heterosexual couples. J Sex Res 2004 Aug;41(3):301-309. [doi: 10.1080/00224490409552237] [Medline: 15497058]
6. Corty EW, Guardiani JM. Canadian and American sex therapists' perceptions of normal and abnormal ejaculatory latencies: how long should intercourse last? J Sex Med 2008 May;5(5):1251-1256. [doi: 10.1111/j.1743-6109.2008.00797.x] [Medline: 18331255]
7. Smith AMA, Patrick K, Heywood W, Pitts MK, Richters J, Shelley JM, et al. Sexual practices and the duration of last heterosexual encounter: findings from the Australian longitudinal study of health and relationships. J Sex Res 2012;49(5):487-494. [doi: 10.1080/00224499.2011.598247] [Medline: 21797717]

8. Amidu N, Owiredu WKBA, Dapare PPM, Antuamwine BB. Perceptions of normal and abnormal ejaculatory latency times: an observational study in Ghanaian males and females. Eur J Med Res 2015 Sep 04;20:75 [FREE Full text] [doi: 10.1186/s40001-015-0169-6] [Medline: 26340933]
9. Waldinger MD, Quinn P, Dilleen M, Mundayat R, Schweitzer DH, Boolell M. A multinational population survey of intravaginal ejaculation latency time. J Sex Med 2005 Jul;2(4):492-497. [doi: 10.1111/j.1743-6109.2005.00070.x] [Medline: 16422843]
10. Shabsigh R, Patrick DL, Rowland DL, Bull SA, Tesfaye F, Rothman M. Perceived control over ejaculation is central to treatment benefit in men with premature ejaculation: results from phase III trials with dapoxetine. BJU Int 2008 Sep;102(7):824-828. [doi: 10.1111/j.1464-410X.2008.07845.x] [Medline: 18647300]
11. Guo L, Liu Y, Wang X, Yuan M, Yu Y, Zhang X, et al. Significance of penile hypersensitivity in premature ejaculation. Sci Rep 2017 Sep 05;7(1):10441. [doi: 10.1038/s41598-017-09155-8] [Medline: 28874780]
12. Garofalo R, Wolf RC, Kessel S, Palfrey J, DuRant RH. The association between health risk behaviors and sexual orientation among a school-based sample of adolescents. Pediatrics 1998 May 01;101(5):895-902. [doi: 10.1542/peds.101.5.895] [Medline: 9565422]
13. Laumann EO, Paik A, Rosen RC. Sexual dysfunction in the United States: prevalence and predictors. JAMA 1999 Feb 10;281(6):537-544. [doi: 10.1001/jama.281.6.537] [Medline: 10022110]
14. Kleinplatz PJ. History of the Treatment of Female Sexual Dysfunction(s). Annu Rev Clin Psychol 2018 May 07;14(1):29-54. [doi: 10.1146/annurev-clinpsy-050817-084802] [Medline: 29356577]
15. Hatzimouratidis K, Amar E, Eardley I, Giuliano F, Hatzichristou D, Montorsi F, European Association of Urology. Guidelines on male sexual dysfunction: erectile dysfunction and premature ejaculation. Eur Urol 2010 May;57(5):804-814. [doi: 10.1016/j.eururo.2010.02.020] [Medline: 20189712]
16. Georgiadis JR, Kringelbach ML, Pfaus JG. Sex for fun: a synthesis of human and animal neurobiology. Nat Rev Urol 2012 Sep 28;9(9):486-498. [doi: 10.1038/nrurol.2012.151] [Medline: 22926422]
17. Atzori L, Iera A, Morabito G. From "smart objects" to "social objects": The next evolutionary step of the internet of things. IEEE Commun Mag 2014 Jan;52(1):97-105. [doi: 10.1109/mcom.2014.6710070]
18. Bardzell J, Bardzell S. Pleasure is your birthright: digitally enabled designer sex toys as a case of third-wave HCI. : Association for Computing Machinery; 2011 May Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 2011; Vancouver, BC, Canada p. 257-266. [doi: 10.1145/1978942.1978979]
19. Richardson K. Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines. IEEE Technol Soc Mag 2016 Jun;35(2):46-53. [doi: 10.1109/mts.2016.2554421]
20. MySQL. URL: https://www.mysql.com/ [accessed 2020-02-25]
21. Wolkamo Co Ltd. URL: http://www.wolkamo.com/ [accessed 2020-02-25]
22. National Bureau of Statistics of China. China Statistical Yearbook (2016). URL: http://www.stats.gov.cn/tjsj/ndsj/2017/indexch.htm [accessed 2018-02-21]
23. NCD Risk Factor Collaboration (NCD-RisC). A century of trends in adult human height. Elife 2016 Jul 26;5:e13410. [doi: 10.7554/eLife.13410] [Medline: 27458798]
24. Robbins CL, Schick V, Reece M, Herbenick D, Sanders SA, Dodge B, et al. Prevalence, frequency, and associations of masturbation with partnered sexual behaviors among US adolescents. Arch Pediatr Adolesc Med 2011 Dec 05;165(12):1087-1093. [doi: 10.1001/archpediatrics.2011.142] [Medline: 21810625]
25. Chivers ML, Seto MC, Blanchard R. Gender and sexual orientation differences in sexual response to sexual activities versus gender of actors in sexual films. J Pers Soc Psychol 2007 Dec;93(6):1108-1121. [doi: 10.1037/0022-3514.93.6.1108] [Medline: 18072857]
26. Puts DA, Welling LL, Burriss RP, Dawood K. Men's masculinity and attractiveness predict their female partners' reported orgasm frequency and timing. Evol Hum Behav 2012 Jan;33(1):1-9. [doi: 10.1016/j.evolhumbehav.2011.03.003]
27. Armstrong EA, England P, Fogarty ACK. Accounting for Women's Orgasm and Sexual Enjoyment in College Hookups and Relationships. Am Sociol Rev 2012 May 07;77(3):435-462. [doi: 10.1177/0003122412445802]
28. Giami A, Laumann EO, Gagnon JH, Michael RT, Michaels S, Michael RT, et al. The Social Organization of Sexuality. Sexual Practices in the United States. Population (French Edition) 1997 Nov;52(6):1548. [doi: 10.2307/1534637]
29. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manag 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]

Abbreviations

IoT: Internet of Things
SeBA: cyberphysical human sexual behavior acquisition system

Edited by G Eysenbach; submitted 02.11.18; peer-reviewed by T Clavier, T Abdulai, D Callander; comments to author 01.10.19; revised version received 26.11.19; accepted 28.01.20; published 03.04.20.

Please cite as:
Zhou X, Zhao J, Liang X
Cyberphysical Human Sexual Behavior Acquisition System (SeBA): Development and Implementation Study in China
JMIR Mhealth Uhealth 2020;8(4):e12677
URL: https://mhealth.jmir.org/2020/4/e12677
doi:10.2196/12677
PMID:32271153

Tutorial

Developing Mental or Behavioral Health Mobile Apps for Pilot Studies by Leveraging Survey Platforms: A Do-it-Yourself Process

Philip I Chow[1], PhD

Center for Behavioral Health and Technology, Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Philip I Chow, PhD
Center for Behavioral Health and Technology
Department of Psychiatry and Neurobehavioral Sciences
University of Virginia
560 Ray C. Hunt Dr.
Charlottesville, VA, 22908
United States
Phone: 1 434 924 5401
Email: philip.i.chow@gmail.com

Abstract

Background: Behavioral health researchers are increasingly recognizing the potential of mobile phone apps to deliver empirically supported treatments. However, current options for developing apps typically require large amounts of expertise or money.

Objective: This paper aims to describe a pragmatic do-it-yourself approach for researchers to create and pilot an Android mobile phone app using existing survey software (eg, Qualtrics survey platform).

Methods: This study was conducted at an academic research center in the United States focused on developing and evaluating behavioral health technologies. The process outlined in this paper was derived and condensed from the steps to building an existing app intervention, iCanThrive, which was developed to enhance mental well-being in women cancer survivors.

Results: This paper describes an inexpensive, practical process that uses a widely available survey software, such as Qualtrics, to create and pilot a mobile phone intervention that is presented to participants as a Web viewer app that is downloaded from the Google Play store. Health researchers who are interested in using this process to pilot apps are encouraged to inquire about the survey platforms available to them, the level of security those survey platforms provide, and the regulatory guidelines set forth by their institution.

Conclusions: As app interventions continue to gain interest among researchers and consumers alike, it is important to find new ways to efficiently develop and pilot app interventions before committing a large amount of resources. Mobile phone app interventions are an important component to discovering new ways to reach and support individuals with behavioral or mental health disorders.

(JMIR Mhealth Uhealth 2020;8(4):e15561) doi:10.2196/15561

KEYWORDS

app; mental health; mHealth

Introduction

Background

Considering the worldwide burden of mental and behavioral health disorders, researchers are recognizing the importance of exploring alternative treatment delivery options, given the following drawbacks of in-person treatment: financial cost [1], time investment [2], social stigma [3], and a shortage of trained health care providers [4-6]. Increasingly, health researchers are discovering the importance of using mobile phone apps to fill a critical health care gap [7-9]. In contrast to in-person health delivery, in many circumstances, app interventions can increase accessibility to users [10,11], are more affordable, and can offer a more efficient use of time. As literature on the efficacy of behavioral and mental health apps continues to flourish [10,12,13], developing and piloting apps remains a significant challenge for many health researchers who lack any requisite experience in native app coding and app development. Having a method to quickly develop, iterate, and pilot app interventions

can lead to a more agile mobile health (mHealth) research lifecycle [14].

Options for App Development

Currently, 2 options generally available to health researchers who wish to develop and test their own app interventions are to (1) write their own software code or (2) have someone else (usually a software engineer) generate the software code. As many health researchers lack expertise in software coding, an attractive option is to procure the services of someone who does. Some academic settings make it possible to leverage the programming services of students (eg, engineering students), given that the number of individuals with programming experience has increased in recent years. Despite these potential options, many health researchers still opt to pay private app developers that primarily focus on building native apps that run on either the Android or iPhone operating systems. Thus, many health researchers who are skilled in intervention development and validation, and are interested in piloting their own intervention apps, never get the opportunity to test their ideas because of financial constraints. This limits the field of app development to researchers with funding, private for-profit corporations, and software engineers. Therefore, it is not surprising that despite a proliferation of health apps available to the public, only a small fraction have been empirically tested and validated [13,15].

In addition to cost, paying a third-party developer to build a native app has additional drawbacks. First, native apps must be adapted to the context of each operating system [16]. Despite some considerable strengths of native apps, such as the ability to leverage mobile phone sensors for passive data collection, they can be vulnerable to crashing because of software updates, and some of them may be incompatible with certain types of mobile phones [17]. Native software code that is developed to run on Android mobile phones cannot be reused for developing an iPhone app, and vice versa [17]. If the unique strengths of a native app are not required for a pilot research trial, researchers may want to cheaply develop and pilot an app intervention without having to commit a large amount of resources. Paying someone else to develop an app can also lead to delayed progress caused by (1) breakdowns in communication between researchers and app developers and (2) scheduling issues on behalf of developers (eg, a developer may decide to deprioritize an app if another more lucrative contract is competing for their time and attention). Additional funds are often necessary for app maintenance, data storage, and making changes to the app once the trial has begun. Thus, a significant amount of money and time can be wasted if a pilot study evaluating an app yields null findings.

A Do-It-Yourself Approach to App Development

This paper is intended for mental and behavioral health researchers who possess content knowledge in a scientific domain (eg, depression, anxiety, and nutrition) but lack critical funding and skills in app development and coding to pilot their own app ideas. Researchers who have the ability to pay an app developer are strongly encouraged to consider that option. The purpose of this paper was to present an alternative, do-it-yourself process for developing and piloting an app by leveraging a widely available survey software, typically referred to as a Web-based app or a hybrid app [16]. In this type of app, the intervention content is built in, and hosted by, the survey platform. All data are also stored by the survey platform and are accessible only to the researcher. Whenever a user launches the app from their mobile phone, the intervention is displayed on their screen through their Web browser. The apps that can be constructed using this process are generally limited to those that administer intervention content through text, interactive exercises, audio files, and video. These apps are not able to passively collect fine-grained data through mobile phone sensors or execute complex sensor-driven interventions (eg, a just-in-time adaptive intervention [18]). Researchers who are interested in having these functionalities in their intervention should consider a native mobile phone app. Although there are several industry platforms for prototyping apps (eg, InVision and Proto.io), many of these are tools for designing an app rather than building an app for data collection in a pilot study. One key advantage of using a secure Web platform such as Qualtrics for highly sensitive data (Qualtrics, Provo, Utah) and Research Electronic Data Capture (REDCap; Vanderbilt University, Nashville, Tennesse) is that, in addition to providing researchers with full control over the intervention development process, these platforms are often compliant with local laws and regulations (eg, Health Insurance Portability and Accountability Act [HIPAA]) for storing protected health information data.

The app intervention discussed in this paper, iCanThrive, is meant to serve as an example of the app building steps provided. It was built using the Qualtrics survey platform for highly sensitive data portal and teaches evidence-based skills grounded in cognitive therapy, acceptance-based therapies, and positive psychology. The app contains brief, intuitive exercises that focus on identifying and challenging distorted thoughts, problem solving, reducing worry, reducing stress through breathing and mindfulness exercises, fostering gratitude, promoting values, savoring positive experiences, and increasing emotional awareness. It was designed to be deployed on users' mobile phones when and where it is needed. The app is available for download on the Google Play store for US Android users. Google Play is the official app store for the Android operating system and is available in over 145 countries. In a recently completed pilot, the app was downloaded and used by 23 cancer survivors. The participants did not report any issues in downloading or using the app. A paper describing the acceptability and preliminary efficacy of iCanThrive is currently under revision [19].

Methods

Five phases were identified to developing a Web-based app (Table 1), which is browser based and hosted in Qualtrics. The phases outlined below are not exhaustive, and researchers are encouraged to consult with the appropriate individuals at their institution (eg, data security experts and institutional review board members) before determining how best (and whether) to undertake this work at their institution. The general phases of the app development process are presented first, followed by a description of how these phases were reflected in the development of the iCanThrive app.

Table 1. General phases of app development using Web-based survey software.

Phase	Description
Establish a theoretical framework	Using existing models of electronic/mobile health, mental health, and cancer to develop a theory-informed intervention.
Wireframe and develop content	Begin to develop content and create a visual representation of how content will be delivered through the app.
Build the intervention in a survey platform	Insert content in an existing survey development platform (Qualtrics), using existing tools to create an interactive and app-like user experience. Test frequently.
Post app on the Google Play store	Using the survey Web link, post the app on the Google Play store for download.
Data quality and output	Before beginning a pilot study with actual participants, ensure that the data are being collected and relevant information (eg, log-ins and session durations) is being collected.

Results

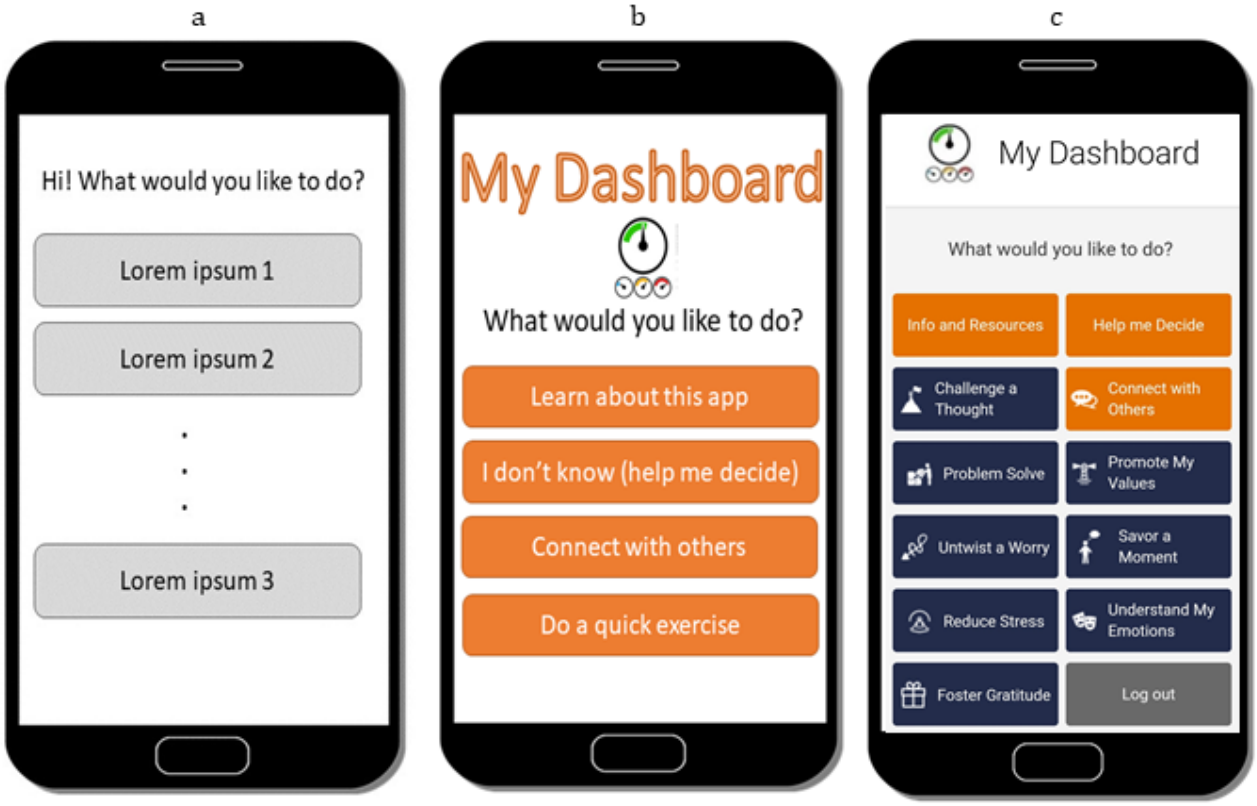
Phase 1: Establish a Theoretical Framework

The first step is to ground the intervention in a theoretical framework based on models of behavioral change [20-22]. It is recommended that researchers incorporate existing electronic health/mHealth intervention models into their conceptualizations. Existing models of internet interventions [23] and mobile interventions, such as the behavioral intervention technologies model [24], can be useful in developing the theoretical framework for app-based interventions. For example, researchers who are designing an intervention that is informed by the theory of planned behavior [20] may seek to modify people's beliefs about the degree to which they have control over some aspect of their behavior (eg, exercising and smoking). Combined with an mHealth intervention model [24] that promotes the use of brief and targeted app sessions, they may decide to provide brief educational content that can be digested by users in short periods or add interactive components (eg, videos) to enhance user experience.

Phase 2: Wireframe and Content Development

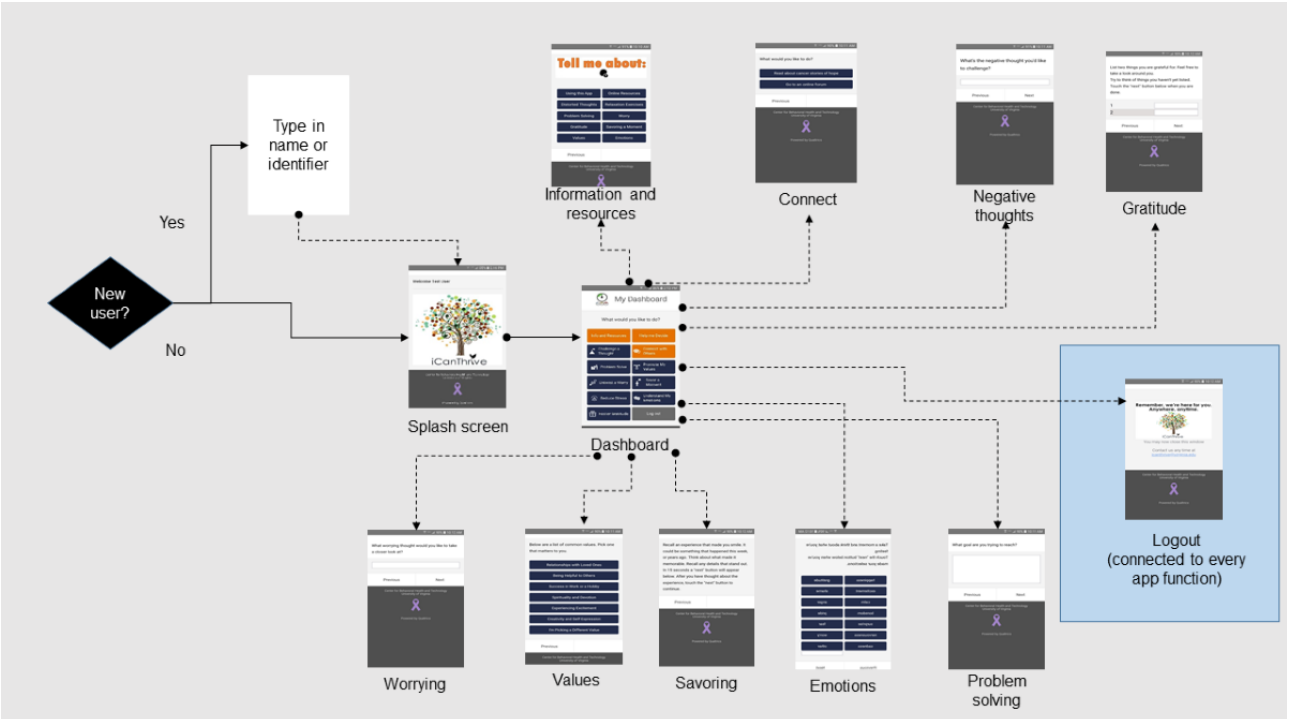
A wireframe is a visual guide of the skeletal framework of the app. As highlighted by others [25], the purpose of creating a wireframe is to determine how best to deliver and meter content across different app pages to achieve the desired effect. In this phase, researchers develop content and determine how best to present the content in a way that is engaging, intuitive, and achieves the purpose of the intervention [23,24]. Researchers can create a wireframe from widely available software, such as PowerPoint (Microsoft Office, Redmond, Washington) and InVision (InVisionApp Inc, New York, New York), or they may choose to draw their wireframe using a paper-and-pencil method. There is also a range of affordable and user-friendly wireframing tools that researchers can search for on the Web, such as Adobe Xd (Adobe, San Jose, California) and Lucidchart (Lucid Software Inc, South Jordan, Utah). Although researchers should expect to iterate during the app construction phase, the wireframe provides an essential blueprint from which to start. Researchers may wish to support users with an overview of the app, how to use it, and expectations users should have of the app. For example, by providing a help option from a main page, researchers can ensure that users can easily access instructions by promoting an intuitive user interface [26]. Another way of promoting user interface is by embedding email links within the app itself to support users and to be responsive to their questions and concerns. Examples of a wireframe process can be seen in Figure 1. The sketching and outlining of app pages allow for quick iterations of existing prototypes until a desired form is achieved.

Figure 1. Examples of 2 previous wireframes for the iCanThrive main dashboard page using PowerPoint. From left to right: (a) early iteration sketch, (b) a more visually appealing iteration, and (c) actual screenshot of the iCanThrive dashboard page.



In addition to prototyping the content and layout of app pages, researchers are encouraged to track how different pages are connected within the app (Figure 2). This enables a high-level perspective of how the app functions and the paths a user can adopt as they navigate the app.

Figure 2. Example of a schematic that tracks how a user can navigate the app by showing how the front app pages and functions are connected.

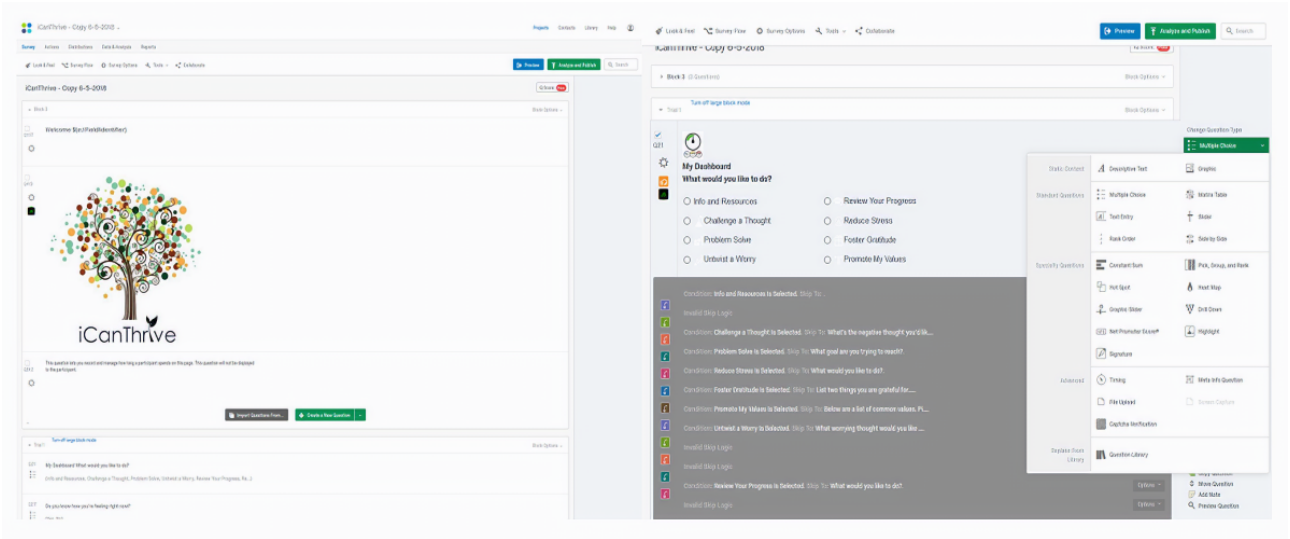


Phase 3: Build the Intervention in a Web Survey Platform

The next phase involves actually building the intervention in a secure Web survey platform. The iCanThrive intervention was built in the Qualtrics survey platform designed to collect and store highly sensitive data. Universities often have subscriptions to survey platforms such as Qualtrics and REDCap, which have similar functions and are typically free for their faculty and students to use for research purposes. Researchers should inquire about the options at their institution and become familiar with using the survey platform available to them. If the intention is to collect sensitive data, researchers should use a survey platform that is secure and compliant with security, privacy, and ethical standards (eg, HIPAA).

Similar to other survey platforms, Qualtrics has preset options for how to display content through different question types. As it is designed to create surveys rather than interventions, Qualtrics naturally labels a new item as a *question*, which is then grouped by *blocks* (Figure 3). Thus, each app page in an intervention (eg, iCanThrive) is actually a different Qualtrics *question* embedded in a single block (Figure 3). An intervention is essentially an elaborate survey with embedded skip and display logic and other features available in the survey platform. Every time a user launches the app, they are completing a Qualtrics survey, which is then stored by the survey platform. Researchers are encouraged to explore the different preset options for creating and displaying content (see Figure 3 for a list of available options in Qualtrics). For example, the *Multiple-Choice* option is useful for creating a checklist of options to which users can respond (Figure 3), whereas the *Graphic* option is convenient for inserting educational figures or logos. This option can be used to create a splash screen, which contributes to an authentic app feel that a user sees when they launch the app from their phone. Question types should be chosen strategically to fulfill the purpose of each app page. For example, the *Text Entry* option is useful for allowing users to freely respond or generate their own examples using their phone's keyboard. A text entry field may also be used to have a user type in a unique subject identifier every time they launch the app, which enables researchers to track the number of app launches for each user. Tracking usage in this way can also allow researchers to monitor app launches across different devices (eg, if a user gets a new phone). After choosing a preset option and inserting content, the *Preview Question* option is helpful for evaluating the page to achieve the desired format.

Figure 3. Screenshots of the iCanThrive splash screen and dashboard pages.

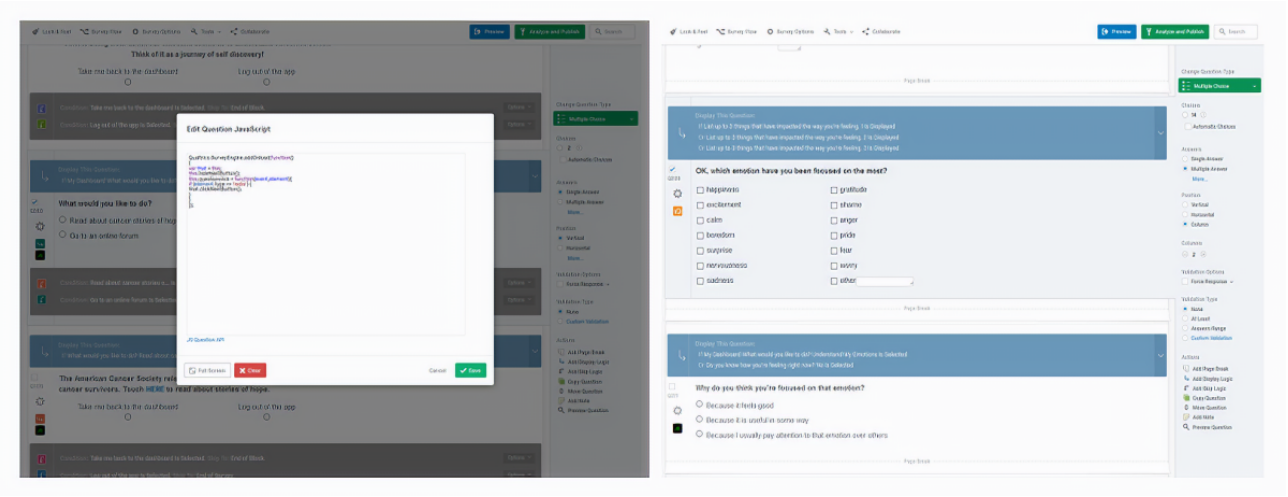


In addition to creating content, Qualtrics has useful tools for connecting different pages within the intervention. For example, after creating a new page, researchers are able to link that page to other existing pages through the *AddSkip Logic* option. This allows researchers to tailor intervention content based on a user's responses. For example, researchers can directly link an app landing page to other pages within the app. In addition, researchers are able to create customized navigation routes through the *Add Display Logic* option found in each Qualtrics question. Researchers can use this function to present users with different pages based on their responses, such as providing additional educational content if a user indicates that they are struggling with some aspect of the intervention. Once again, researchers are encouraged to explore different options to create the type of navigation they desire within the intervention.

Qualtrics has additional features that enable researchers to enhance a user's experience through programmable code (eg, JavaScript). For example, in the preset version of Qualtrics, after indicating a response, individuals are asked to select a *next* button to proceed to the next page. Researchers who have some rudimentary background in JavaScript can directly insert their own code in a Qualtrics question that will enable users to bypass this step, allowing them to automatically jump to the appropriate page when they make a selection to a multiple-choice question (see Figure 4 for the JavaScript code that was used). It should be noted that the use of programmable code is optional in the app process outlined in this paper. Health researchers who do not have the time to acquire any coding skills are encouraged to bypass the use of code. Qualtrics also enables researchers to change the overall presentation of the intervention through the *Look & Feel* tab, where researchers can enhance user experience by inserting customized style sheets (eg, to change the colors

of response buttons) or use existing drop-down options such as selecting different color schemes. It is important to note that researchers who lack the requisite experience to make such changes (eg, creating customized style sheets) yet are interested in exploring these options, may wish to pay a Web developer to enact the desired changes. Researchers are strongly encouraged to test their app frequently before conducting a trial with actual cancer survivors. As the functions of the app are executed within the robust Qualtrics survey platform, health researchers may wish to evaluate the app based on aesthetic qualities (eg, color scheme, font sizes, and graphics), intuitiveness of using and navigating the app (eg, sophistication of language in the app and use of too many required text fields that may impede progress), and whether the app pages are connected in the intended ways (eg, skip logic redirects the user to the appropriate page in the app). Researchers should consider what resources are available to them in evaluating the app. For example, students can be asked to evaluate whether the app pages are linked in the appropriate ways. End users (eg, cancer survivors) may be asked to trial the app to evaluate aesthetic qualities and the language used in the app pages. Though it is outside the scope of this paper, there is abundant literature on involving cancer populations in the development and testing of interventions [27-31].

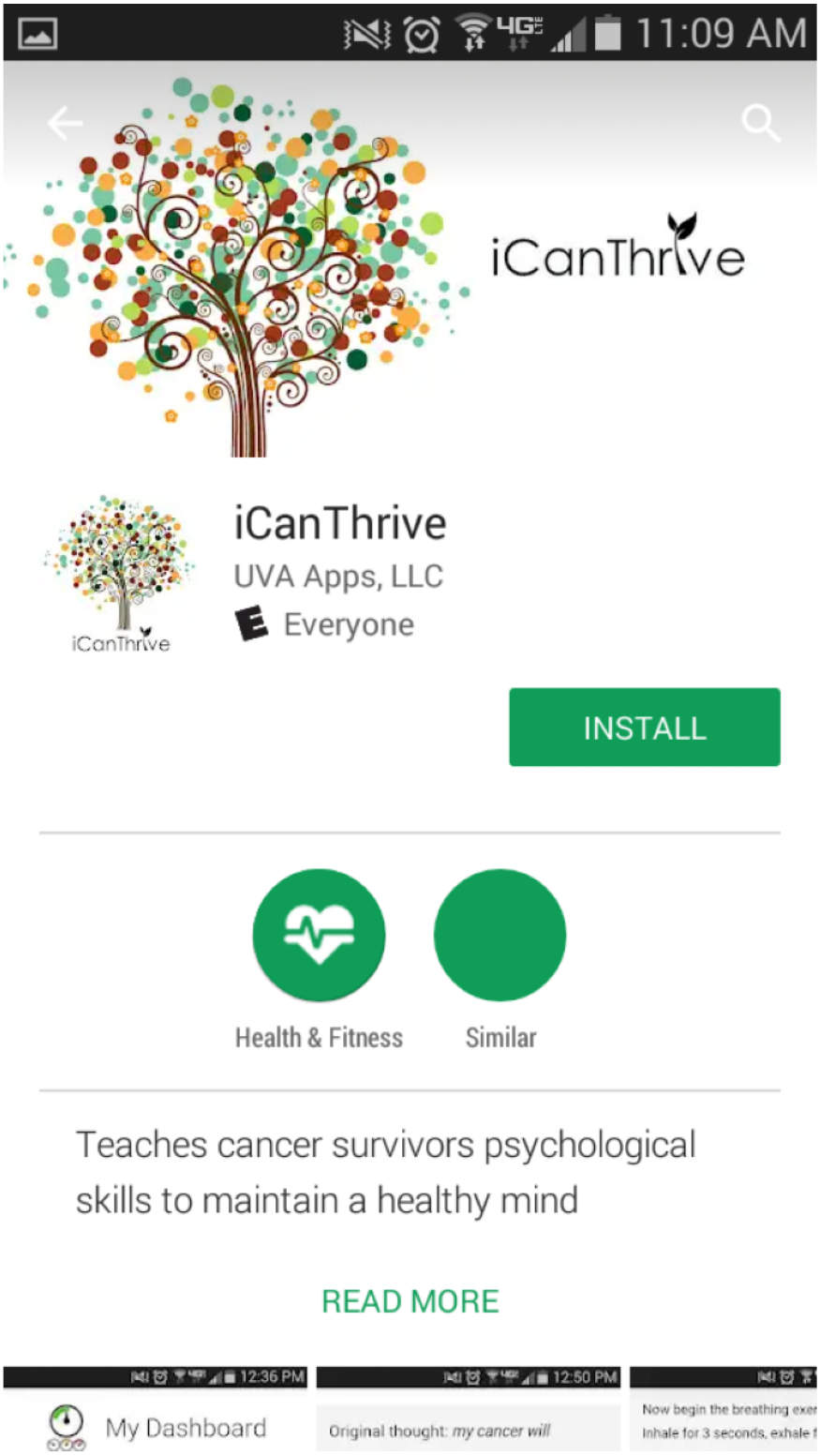
Figure 4. Screenshots of the custom JavaScript and using the multiple-choice question type to create intervention content in the iCanThrive app.



Phase 4: Post the App on the Google Play Store

Qualtrics provides a unique URL for each survey that is created in its platform. The URL is the unique Web address that is linked to that survey. The URL of the survey that contains the intervention content can be used to post a Web viewer app to the Google Play store for user download (Figure 5). As the process of posting an app on the Google Play store is beyond the scope of this paper and is the same for all types of Android apps, researchers are encouraged to learn about this process through Google's support pages or other reputable sources. To post an app requires a one-time registration free (US $25 as of 2018), and Google Play allows for the release of apps to select groups of users as alpha or beta tests.

Figure 5. Screenshot of the iCanThrive app on the Google Play store.



Phase 5: Data Quality and Output

Similar to other survey platforms, Qualtrics can export data in a convenient spreadsheet on demand (Figure 5). It should be noted that researchers should be aware that there are 2 sources within Qualtrics from which to download data: (1) recorded responses, which is composed of sessions in which a user completes a session by *submitting* their responses, and (2) responses in progress, which is composed of sessions in which a user begins a session but does not submit their response or prematurely closes their Web browser (these types of data should be downloaded by the researcher every month of the trial, as responses in progress are not saved in Qualtrics in perpetuity). As seen in Figure 6, data are presented in a wide table format, whereby columns represent different variables found within the intervention and each row represents a separate log-in from a user. Entries are presented in chronological order, although researchers may choose to represent the data in whatever way

suits them best. Importantly, information such as app launches can be easily calculated from the first few columns of the data spreadsheet, as for every log-in attempt, Qualtrics records the date, time, and duration in seconds. It should be noted that because of its design as a survey platform, the order of columns in the data spreadsheet corresponds to how the question items are ordered in the actual survey. However, researchers who wish to examine how users responded to specific intervention content (within each app launch) and navigated through the app may still obtain this information by sorting through the columns. For example, column *N* in Figure 6 contains what module users selected when they first encountered the landing screen in iCanThrive, and by scanning the rows of the data file, it is possible to understand how users navigated through the app at each log-in.

Figure 6. Screenshot of a hypothetical data spreadsheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	StartDate	EndDate	Status	Progress	Duration (	Finished	RecordedD	Distributio	UserLangu	Q642_Firs	Q642_Las	Q642_Pag	Q642_Clic	Q21	Q217	Q207	Q207_12_	Q208_1
2	Start Date	End Date	Response	Progress	Duration (	Finished	Recorded	Distributio	User Langı	Timing - Fi	Timing - La	Timing - Pa	Timing - Cl	My	Do you kn	Take a	Take a	List up to 3
3	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId	{"ImportId
4	########	########	Survey Pre	100	32	TRUE	########	preview	EN					Info and Resources				
5	########	########	Survey Pre	100	5	TRUE	########	preview	EN					Info and Resources				
6	########	########	Survey Pre	100	21	TRUE	########	preview	EN					Info and Resources				
7	########	########	Survey Pre	100	15	TRUE	########	preview	EN					Info and Resources				
8	########	########	Survey Pre	100	13	TRUE	########	preview	EN					Info and Resources				
9	########	########	Survey Pre	100	14	TRUE	########	preview	EN					Challenge a Thought				
10	########	########	Survey Pre	100	9	TRUE	########	preview	EN					Challenge a Thought				
11	########	########	Survey Pre	100	30	TRUE	########	preview	EN					Problem Solve				
12	########	########	Survey Pre	100	10	TRUE	########	preview	EN					Challenge a Thought				
13	########	########	Survey Pre	100	11	TRUE	########	preview	EN					Info and Resources				
14	########	########	Survey Pre	100	6	TRUE	########	preview	EN					Foster Gratitude				
15	########	########	Survey Pre	100	49	TRUE	########	preview	EN					Problem Solve				
16	########	########	Survey Pre	100	53	TRUE	########	preview	EN					Info and Resources				
17	########	########	Survey Pre	100	19	TRUE	########	preview	EN					Log out				
18	########	########	IP Address	100	138	TRUE	########	anonymou	EN					Info and Resources				
19	########	########	IP Address	100	300	TRUE	########	anonymou	EN					Info and Resources				
20	########	########	Survey Pre	100	11	TRUE	########	preview	EN					Challenge a Thought				
21	########	########	Survey Pre	100	23	TRUE	########	preview	EN					Promote My Values				
22	########	########	Survey Pre	100	26	TRUE	########	preview	EN					Info and Resources				
23	########	########	Survey Pre	100	156	TRUE	########	preview	EN					Log out				
24	########	########	IP Address	100	67	TRUE	########	anonymou	EN					Problem Solve				
25	########	########	IP Address	100	65	TRUE	########	anonymou	EN					Foster Gratitude				
26	########	########	IP Address	100	57589	TRUE	########	anonymou	EN					Info and Resources				
27	########	########	IP Address	100	412	TRUE	########	anonymou	EN					Info and Resources				
28	########	########	IP Address	100	50	TRUE	########	anonymou	EN					Promote My Values				
29	########	########	IP Address	100	302	TRUE	########	anonymou	EN					Challenge a Thought				

Example: iCanThrive

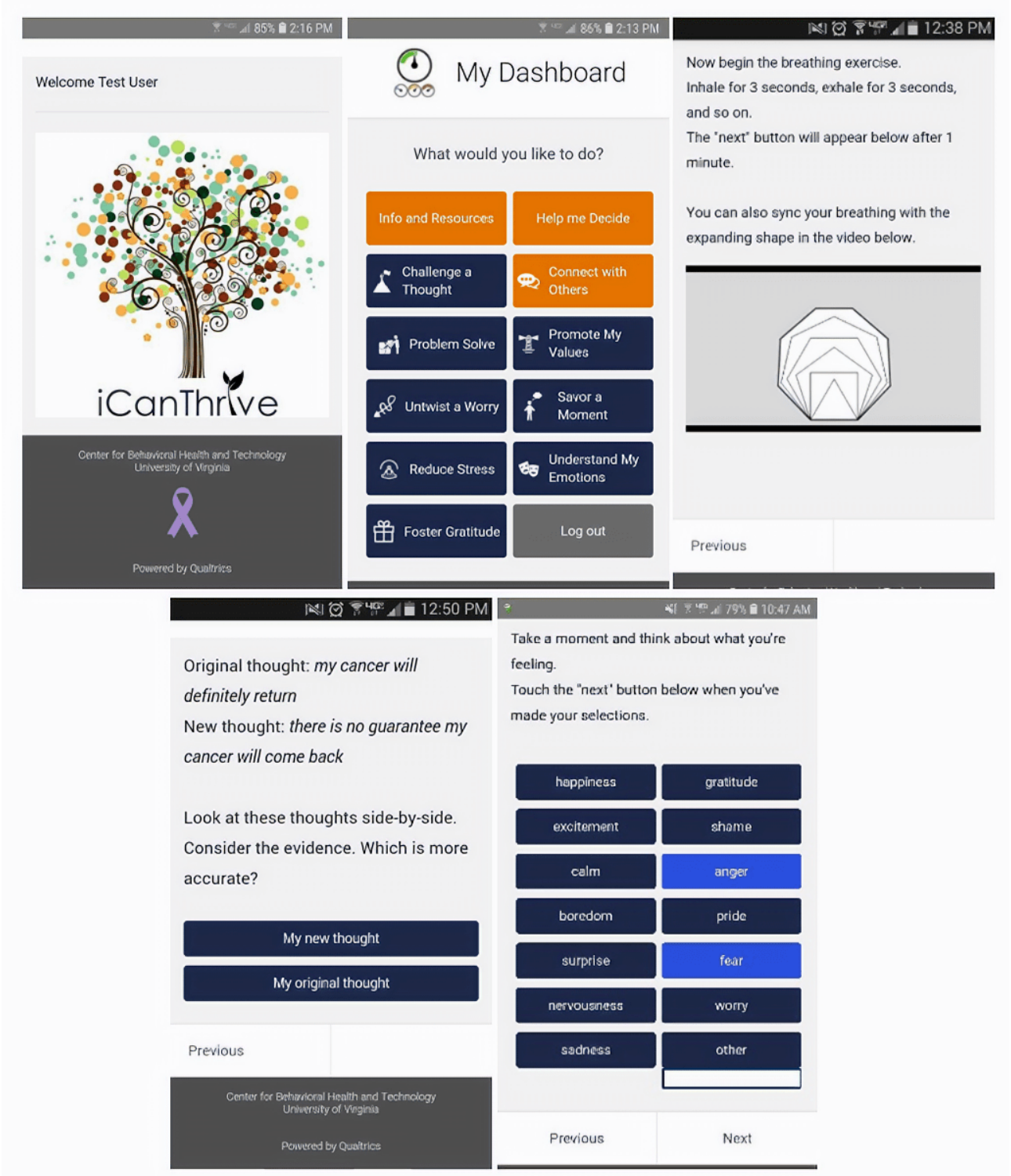
Phase 1: Establish a Theoretical Framework

The iCanThrive app was intended for users to launch where or when they desired and follows a cognitive behavioral therapy (CBT) framework, whereby patients acquire skills by practicing them in daily life. This approach was informed by findings that suggest that people use apps in short and frequent bursts [32] and was believed to be particularly relevant to cancer survivors given the considerable amount of time of cancer treatment [2] and its impact on leisure time [33]. The model of internet interventions [34] informed the use of video content and exercises to promote mental health. Thus, iCanThrive was designed to increase mental well-being by delivering brief interactive exercises that cancer survivors could perform in daily life, when and where they were needed.

Phase 2: Wireframe and Content Development

The iCanThrive app was developed with a detailed wireframe that informed the development and placement of each app page (Figures 1 and 2). Upon launching the iCanThrive app, users are presented with a brief splash screen that contains the app logo (Figure 7, left), before automatically directing them to the *My Dashboard* screen (Figure 7, second from left). From this screen, users can navigate freely by selecting the appropriate option on their phone screen. They can access any of the 8 intervention modules, learn more about the psychological constructs the app targets, or be guided to 1 of the 8 intervention modules based on a series of questions that assess the user's current state.

Figure 7. Screenshots of iCanThrive app pages.



The *My Dashboard* screen was a critical component of the skeletal framework of the app, a central point from which users can easily navigate. For example, if a user picks the *Challenge a Thought* module from the *My Dashboard* screen, they are guided through an interactive exercise that leads them through the steps of recording a distorted thought, classifying the thought, and ultimately generating a new alternative thought. At the end of the exercise, users are given an option to log out of the app (which submits their responses) or navigate back to the *My Dashboard* screen. Each app page in the *Challenge a Thought* module was articulated in the iCanThrive wireframe before the app was constructed and follows evidence-based approaches in CBT. An additional consideration that informed the design of the app was wanting users to complete exercises in as little time as possible. Users can complete most exercises in iCanThrive in less than 2 min. The exercises require few instructions to complete, and any instructions are clearly presented on the screen.

If a user chooses the *Info and Resources* option from the *My Dashboard* page, they are brought to a page that contains an

overview of the app, how to use it, and expectations that users should have of the app. Users may also directly contact the iCanThrive research staff through embedded email links within the app itself.

Phase 3: Build the Intervention in a Web Survey Platform

Each app page in iCanThrive is actually a different Qualtrics *question* embedded in a single block (Figure 4). This option was chosen to create the iCanThrive splash screen, which users see every time they launch the app. The *My Dashboard* landing page in iCanThrive is linked to many other pages within the app. Users are automatically sent to the first page of the module they select (eg, *Challenge a Thought* and *Understand My Emotions*) from the *My Dashboard* page. The app also contains customized navigation routes. For example, in the *Reduce Stress* module in iCanThrive, after completing a breathing exercise, users are asked if they are more or less relaxed than they were before the exercise. If a user indicates they are more relaxed, they are sent to a page that congratulates them for their success. If a user indicates they are less relaxed than before, they are sent to a page that normalizes their experience and encourages them to try again. The use of an app to support the delivery of content and simple exercises has been shown to be effective in reducing mental health issues [35-37].

All components of the app, iterative testing, and app modifications were performed by the lead author (who is not a software developer), and therefore, there were no costs in the development of iCanThrive other than investigator time. Testing of an app page was performed after each modification to ensure there were no errors. Every page in iCanThrive went through 2 to 15 iterations, depending on the complexity of the content and interactive content. After all app pages were tested and assembled, a Web developer was [37] paid less than US $400 to make cosmetic improvements to the iCanThrive program. Researchers may wish to apply a more user-centered design to developing their own app-based interventions [38].

Phase 4: Post App on the Google Play Store

The iCanThrive app was posted on the Google Play store in May 2018. To further customize the app to the individual user, a local Web developer enabled a few lines of code that prompts users to write their name upon launching the app for the first time. Subsequent app launches will display the user's name at each app launch (Figure 7), which will also be listed in the final data spreadsheet.

Phase 5: Data Output

As seen in Figure 6, iCanThrive data are presented in a wide table format. As each row represents an instance in which the program was initiated, a health researcher can track the number of app launches.

Discussion

Principal Findings

This paper is intended to provide mental health researchers with few resources with a do-it-yourself process for developing and piloting their own Web-based app. The app development process outlined in this paper is intended to assist health researchers with limited resources to *pilot* an app to establish a proof of concept. This method may not be suitable, nor is recommended, for health researchers who (1) have funds to pay an app developer or (2) are conducting a large-scale trial that requires a more polished app than can be afforded using the proposed process. Readers who are interested in other mHealth platforms that can enable the creation of a mobile intervention should visit the Web pages for LifeGuide (University of Southampton, Southampton, UK) and Chorus (University of California, Los Angeles, California).

The steps presented are not exhaustive. Additional considerations should be stressed when applying the process presented in this paper. If researchers intend to collect personal health information from their participants, they must ensure that the survey platform they use is compatible for collecting and storing such information. Researchers are strongly encouraged to assess options at their own institutions and to use secure survey platforms (eg, REDCap) that are able to collect and store sensitive data. Both within the app itself and when posting their app on the Google Play store, researchers should include privacy policy and terms of use sections that explicitly state the conditions upon which users download and use the app, the expectations users should have when using the app, and a disclaimer regarding that the user acknowledges the conditions presented when using the app. This information increases transparency to users and offers protection for researchers. Finally, depending on the specific population and their needs, researchers may wish to adopt a co-design approach to designing and constructing their app interventions [39]. Specifically, to learn about the needs, attitudes, and preferences of their end users, health researchers may wish to conduct mixed method studies to understand how to best tailor an app intervention [40] to inform phase 1 (establishing a theoretical framework) or phase 2 (wireframe and content development) of the app building process [27-31]. Information regarding ways to better tailor the app to a specific population should account for the limitations of the survey platform being leveraged (eg, inability to facilitate live video conferencing).

The worldwide proliferation of personal-use mobile phones has provided mental and behavioral health researchers with an opportunity to deliver app-based interventions to those in need. Mobile phone apps are affordable, highly accessible, and capable of delivering empirically informed health interventions. For example, studies find that digital interventions that mirror the content of in-person therapy can perform just as well in reducing mood symptoms [41,42]. Mobile phone apps can be a tremendous resource for those in underserved areas, as the majority of mental and behavioral health specialists practice in densely populated regions, leading to a disparity in access to care for many ethnic minority individuals and those that reside in rural areas [43-45]. Although this paper used iCanThrive, a mental well-being intervention, to illustrate the process of building an app, the process outlined in this paper can in theory be used to create other types of health interventions. For example, health researchers who wish to build an app to reduce smoking may wish to use an app to deliver educational content on the harms of smoking. Researchers interested in diet and

exercise may specify portions of an app to collect information on meals and activity levels.

Limitations

The method described in this paper may not be applicable to developing app-based interventions for iOS (ie, Apple) devices. The iOS app store has strict guidelines for publishing apps and tend to only publish native apps. This process does not allow for fine-grained passive data collection that leverages mobile phone sensors. Finally, the app process described in this paper is not relevant for health researchers who are interested in developing an app that enables live video conferencing.

Conclusions

Enabling researchers to cheaply and rapidly develop and test mobile interventions, through processes such as the one outlined in this paper, can ultimately lead to increased access to evidence-based behavioral and mental health interventions.

Acknowledgments

This research was aided by grant number 157675 from the Virginia Kincaid Cancer Research Fund and grant number IRG 81-001-26 from the American Cancer Society. The author wishes to thank Gabe Heath and Brendan Murphy for their consultation and expertise.

Conflicts of Interest

None declared.

References

1. Chi M. The hidden cost of cancer: helping clients cope with financial toxicity. Clin Soc Work J 2017;47(3):249-257. [doi: 10.1007/s10615-017-0640-7]
2. Yabroff K, Davis W, Lamont E, Fahey A, Topor M, Brown M, et al. Patient time costs associated with cancer care. J Natl Cancer Inst 2007 Jan 3;99(1):14-23. [doi: 10.1093/jnci/djk001] [Medline: 17202109]
3. Holland JC, Kelly BJ, Weinberger MI. Why psychosocial care is difficult to integrate into routine cancer care: stigma is the elephant in the room. J Natl Compr Canc Netw 2010 Apr;8(4):362-366. [doi: 10.6004/jnccn.2010.0028] [Medline: 20410331]
4. Davis AS, McIntosh DE, Phelps L, Kehle TJ. Addressing the shortage of school psychologists: a summative overview. Psychol Schs 2004;41(4):489-495. [doi: 10.1002/pits.10192]
5. Shafran R, Clark D, Fairburn C, Arntz A, Barlow D, Ehlers A, et al. Mind the gap: Improving the dissemination of CBT. Behav Res Ther 2009 Nov;47(11):902-909. [doi: 10.1016/j.brat.2009.07.003] [Medline: 19664756]
6. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
7. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
8. Muñoz RF, Chavira DA, Himle JA, Koerner K, Muroff J, Reynolds J, et al. Digital apothecaries: a vision for making health care interventions accessible worldwide. Mhealth 2018;4:18 [FREE Full text] [doi: 10.21037/mhealth.2018.05.04] [Medline: 30050914]
9. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
10. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
11. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010 Dec 17;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]
12. Anthes E. Pocket psychiatry: mobile mental-health apps have exploded onto the market, but few have been thoroughly tested. Nature 2016;532:20-23 [FREE Full text]
13. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
14. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018;1:46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
15. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, 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]

16. Joorabchi ME, Mesbah A, Kruchten P. Real Challenges in Mobile App Development. In: 2013 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement. 2013 Presented at: ESEM'13; October 10-11, 2013; Baltimore, MD, USA. [doi: 10.1109/ESEM.2013.9]
17. Xanthopoulos S, Xinogalos S. A Comparative Analysis of Cross-Platform Development Approaches for Mobile Applications. In: Proceedings of the 6th Balkan Conference in Informatics. 2013 Presented at: BCI'13; September 19 - 21, 2013; Thessaloniki, Greece p. 213-220. [doi: 10.1145/2490257.2490292]
18. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
19. Chow P, Drago F, Kennedy E, Cohn W. A novel mobile phone app intervention with phone coaching to reduce symptoms of depression in survivors of women's cancer: pre-post pilot study. JMIR Cancer 2020 Feb 6;6(1):e15750 [FREE Full text] [doi: 10.2196/15750] [Medline: 32027314]
20. Ajzen I. From intentions to actions: A theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control. Berlin, Heidelberg: Springer; 1985:11-39.
21. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
22. Strecher V, Rosenstock I. The health belief model. In: Glanz K, Lewis FM, Rimer BK, editors. Health Behavior and Health Education: Theory, Research, and Practice, second edition. San Francisco: Jossey-Bass Publishers; 1997:41-59.
23. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
24. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
25. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
26. Wasserman AI. Software Engineering Issues for Mobile Application Development. In: Proceedings of the FSE/SDP workshop on Future of software engineering research. 2010 Presented at: FoSER'10; November 7 - 8, 2010; New Mexico, Santa Fe, USA p. 397-400. [doi: 10.1145/1882362.1882443]
27. Northouse L, Schafenacker A, Barr KL, Katapodi M, Yoon H, Brittain K, et al. A tailored web-based psychoeducational intervention for cancer patients and their family caregivers. Cancer Nurs 2014;37(5):321-330 [FREE Full text] [doi: 10.1097/NCC.0000000000000159] [Medline: 24945270]
28. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
29. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
30. Jensen JD, King AJ, Carcioppolo N, Davis L. Why are tailored messages more effective? A multiple mediation analysis of a breast cancer screening intervention. J Commun 2012 Oct;62(5):851-868 [FREE Full text] [doi: 10.1111/j.1460-2466.2012.01668.x] [Medline: 26405350]
31. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
32. Oulasvirta A, Tamminen S, Roto V, Kuorelahti J. Interaction in 4-second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2005 Presented at: CHI'05; April 2 - 7, 2005; Portland, OR, USA p. 919-928. [doi: 10.1145/1054972.1055101]
33. Taylor K, Currow D. A prospective study of patient identified unmet activity of daily living needs among cancer patients at a comprehensive cancer care centre. Aust Occ Ther J 2003 Jun;50(2):79-85. [doi: 10.1046/j.1440-1630.2003.00327.x]
34. Ritterband L, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
35. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017 Jan 5;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
36. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, et al. Comparison of the effects of coaching and receipt of app recommendations on depression, anxiety, and engagement in the IntelliCare platform: Factorial randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e13609 [FREE Full text] [doi: 10.2196/13609] [Medline: 31464192]

37. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv 2016 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
38. Kangas E, Kinnunen T. Applying user-centered design to mobile application development. Commun ACM 2005;48(7):55-59. [doi: 10.1145/1070838.1070866]
39. Verbiest M, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: A case study in New Zealand. Transl Behav Med 2019 Jul 16;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]
40. Bradbury K, Watts S, Arden-Close E, Yardley L, Lewith G. Developing digital interventions: a methodological guide. Evid Based Complement Alternat Med 2014;2014:561320 [FREE Full text] [doi: 10.1155/2014/561320] [Medline: 24648848]
41. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
42. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
43. Brown A, Rice SM, Rickwood DJ, Parker AG. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia Pac Psychiatry 2016 Mar;8(1):3-22. [doi: 10.1111/appy.12199] [Medline: 26238088]
44. Creedon TB, Cook BL. Access to mental health care increased but not for substance use, while disparities remain. Health Aff (Millwood) 2016 Jun 1;35(6):1017-1021. [doi: 10.1377/hlthaff.2016.0098] [Medline: 27269017]
45. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Public Health 2015 Jun;129(6):611-620. [doi: 10.1016/j.puhe.2015.04.001] [Medline: 26025176]

Abbreviations

CBT: cognitive behavioral therapy
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 19.07.19; peer-reviewed by S Schueller, C Escobar-Viera, L Appel; comments to author 14.10.19; revised version received 01.11.19; accepted 24.01.20; published 20.04.20.

Please cite as:
Chow PI
Developing Mental or Behavioral Health Mobile Apps for Pilot Studies by Leveraging Survey Platforms: A Do-it-Yourself Process
JMIR Mhealth Uhealth 2020;8(4):e15561
URL: https://mhealth.jmir.org/2020/4/e15561
doi:10.2196/15561
PMID:32310143

Original Paper

App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study

Ronny Bergquist[1], MSc; Beatrix Vereijken[1], PhD; Sabato Mellone[2], PhD; Mattia Corzani[2], MSc; Jorunn L Helbostad[1], PhD; Kristin Taraldsen[1], PhD

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Bologna, Italy

Corresponding Author:
Ronny Bergquist, MSc
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Edvard Griegs gate 8
Trondheim, 7030
Norway
Phone: 47 47634462
Email: ronny.bergquist@ntnu.no

Abstract

Background: Objective measures of physical function in older adults are widely used to predict health outcomes such as disability, institutionalization, and mortality. App-based clinical tests allow users to assess their own physical function and have objective tracking of changes over time by use of their smartphones. Such tests can potentially guide interventions remotely and provide more detailed prognostic information about the participant's physical performance for the users, therapists, and other health care personnel. We developed 3 smartphone apps with instrumented versions of the Timed Up and Go (Self-TUG), tandem stance (Self-Tandem), and Five Times Sit-to-Stand (Self-STS) tests.

Objective: This study aimed to test the usability of 3 smartphone app–based self-tests of physical function using an iterative design.

Methods: The apps were tested in 3 iterations: the first (n=189) and second (n=134) in a lab setting and the third (n=20) in a separate home-based study. Participants were healthy adults between 60 and 80 years of age. Assessors observed while participants self-administered the tests without any guidance. Errors were recorded, and usability problems were defined. Problems were addressed in each subsequent iteration. Perceived usability in the home-based setting was assessed by use of the System Usability Scale, the User Experience Questionnaire, and semi-structured interviews.

Results: In the first iteration, 7 usability problems were identified; 42 (42/189, 22.0%) and 127 (127/189, 67.2%) participants were able to correctly perform the Self-TUG and Self-Tandem, respectively. In the second iteration, errors caused by the problems identified in the first iteration were drastically reduced, and 108 (108/134, 83.1%) and 106 (106/134, 79.1%) of the participants correctly performed the Self-TUG and Self-Tandem, respectively. The first version of the Self-STS was also tested in this iteration, and 40 (40/134, 30.1%) of the participants performed it correctly. For the third usability test, the 7 usability problems initially identified were further improved. Testing the apps in a home setting gave rise to some new usability problems, and for Self-TUG and Self-STS, the rates of correctly performed trials were slightly reduced from the second version, while for Self-Tandem, the rate increased. The mean System Usability Scale score was 77.63 points (SD 16.1 points), and 80-95% of the participants reported the highest or second highest positive rating on all items in the User Experience Questionnaire.

Conclusions: The study results suggest that the apps have the potential to be used to self-test physical function in seniors in a nonsupervised home-based setting. The participants reported a high degree of ease of use. Evaluating the usability in a home setting allowed us to identify new usability problems that could affect the validity of the tests. These usability problems are not easily found in the lab setting, indicating that, if possible, app usability should be evaluated in both settings. Before being made available to end users, the apps require further improvements and validation.

(JMIR Mhealth Uhealth 2020;8(4):e16507) doi:10.2196/16507

KEYWORDS

physical function; mHealth app; usability; older people; seniors

Introduction

At the time of retirement, at the age of 60-70 years, many people experience a significant decline in physical activity levels [1], and balance, gait, and mobility typically start to decline at a higher rate than before [2,3]. Thus, detection of changes in physical function at an early stage could be crucial to improve or prevent future declines in physical function and to sustain physical function over time. Objective assessment of physical function in health care settings is resource-demanding and therefore limited to people with a pressing need to have their function assessed, such as individuals who have experienced falls or who have been diagnosed with a condition known to affect physical functioning. Because functional decline typically occurs slowly, it might not pose an issue for the individual until their ability to perform activities of daily life is affected. Consequently, it might not be obvious why younger or well-functioning seniors should have their physical function assessed until it has come to this stage.

Innovations in mobile health (mHealth) technology have paved the way for new possibilities in assessing physical function. Most smartphones are equipped with sensors such as accelerometers, gyroscopes, and magnetometers and have high computational power; therefore, smartphones can be considered an inertial measurement unit enabling an objective and reliable assessment of physical function [4]. Considering that seniors are the fastest growing group of smartphone users [5] and that, in 2017, 42% of adults aged 65 or older in the United States owned a smartphone [6], there is great potential for using smartphones as a tool for self-assessing physical function [7]. Furthermore, well-designed and evidence-based apps represent new opportunities in preventive strategies for the senior population as a valuable tool in helping to make changes in their lives that can prevent functional decline.

Three such smartphone apps for self-assessment of physical function were developed as part of the PreventIT (early risk detection and prevention in aging people by self-administered ICT-supported assessment and a behavioral change intervention, delivered by use of smartphones and smartwatches) project. PreventIT was a European Union Horizon 2020 Personalising Health and Care project. The apps allow users to self-administer instrumented versions of the Timed Up and Go (Self-TUG), tandem stance (Self-Tandem), and Five Times Sit-to-Stand (Self-STS) tests in order to measure mobility and dynamic balance, static balance, and leg strength, respectively.

When developing an mHealth app for self-assessment of physical function, the usability of the app must be carefully considered, as it has been shown to be a fundamental determinant for technology adoption among older adults [8]. Usability is defined in the official International Organization for Standardization (ISO) guidelines as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [9]. Furthermore, when measuring aspects of one’s health, the accuracy of the results relies on correct administration of the test. Thus, any usability problem associated with using an app-based test should be identified and addressed before it is made available to end users. This is usually done through several iterations of testing with target user groups, ideally until no major usability problems exist with regards to using the apps and administering the test. Usability studies are most often carried out in a lab setting, which is convenient and offers a high degree of control, as opposed to field-based usability testing. However, field-based testing, which, in this context, would be a home setting, provides insight into how the system is used under more realistic situations. Depending on the system being tested and the phase of development, usability should ideally be tested in both lab and home settings.

The overall aim of this study was to evaluate whether people in our target group of seniors between the ages of 60 years and 80 years were able to reliably self-administer the tests on their own using the smartphone, apps, and instructions we provided without any interaction with the assessors. In this paper, we describe the 3 iterations of usability testing with target user groups that were needed to identify all major usability problems. Each iteration consisted of a development phase and subsequent testing phase. In the first 2 iterations, we performed the usability tests in a controlled lab setting, where the assessors had prepared the test setup and necessary materials for the participants beforehand. For the third testing phase, participants were in their own homes, where they needed to prepare the test setup themselves by following instructions presented within the apps. This study does not address the topic of algorithms for signals and data processing nor how to present specific information and feedback to the users based on the test results.

Methods

Design Overview

We developed 3 app-based self-tests of physical function within the European Union Horizon 2020 project PreventIT [10]. Technology development in PreventIT followed the ISO standard 9241-210 [9] on user-centered development of products, and an iterative design approach was used to develop and test the usability of the apps. Because our target group is community dwellers and not clinical patients, we did not follow the ISO norm for medical devices. The target group of the apps was community-dwelling people aged 60 years and older, able to walk independently, and without any cognitive, functional hearing, or visual impairments. The overall aim of the mobile-based, self-administrable functional tests is early identification of risk for age-related functional decline by extracting relevant digital biomarkers from the smartphone-embedded inertial sensors. The intended context of use for the apps is to guide preventive intervention strategies for the general population.

An initial version (version 1) of the Self-TUG and Self-Tandem apps was included for the first iteration. The apps were upgraded based on the results of this testing, and the Self-STS was added

as a third self-test. All 3 apps were tested under similar conditions during the second iteration (apps version 2). After further upgrades, version 3 of the apps was tested in a summative usability evaluation with a new group of volunteers in a home setting.

Participants

We included participants from two studies. First, we included participants from a multicenter, 3-armed, feasibility randomized controlled trial conducted within the PreventIT project. For the first and second iterations, we included participants from the main study if they had performed the self-administration of the apps during baseline (iteration 1) and follow-up (iteration 2). The inclusion criteria for the participants are described in detail in the protocol paper for the PreventIT trial [10]. In short, for iterations 1 and 2, we included 189 and 134 community-dwelling adults, respectively, aged between 61 and 70 years from Trondheim, Norway; Stuttgart, Germany; and Amsterdam, the Netherlands.

For iteration 3, we included 20 community-dwelling adults ranging in age from 60 years to 80 years (mean 68.7 years, SD 5.2 years) in Trondheim, Norway. Inclusion criteria were community-dwelling status, age between 60 years and 80 years, ability to walk 500 meters independently, Norwegian-speaking, ability to hear sound from a smartphone, and current user of a smartphone. Participants were excluded if they reported any severe cardiovascular, pulmonary, neurological, or mental diseases.

Description and Development of the Apps – From Version 1 to Version 3

We developed the apps using Android Studio 3.1.2 (Google, Mountain View, CA). Versions 1 and 2 of the apps were installed on a Samsung Galaxy S3 (Samsung, Seoul, Korea), while version 3 was installed on a Samsung Galaxy S8 (Samsung, Seoul, Korea).

Self-Timed Up and Go, Self-Sit to Stand, and Self-Tandem Apps

We created separate apps for each of the clinical tests (TUG, Five Times STS, and tandem stance). The apps were developed to be used as standalone tests, so one or more tests could be skipped if participants felt unsafe or did not want to perform a test. The TUG is a measure of mobility, in which the participant is timed while rising up from a chair, walking 3 meters, turning around, walking back, and sitting down again. In the Five Times STS, the participant is supposed to stand up from a chair and sit down again repeatedly 5 times as fast as possible, while being timed. In the tandem stance, the participant is supposed to place one foot in front of the other, heel-to-toe, in a straight line for 15 seconds, if possible. The Self-TUG uses an algorithm to detect the different phases of the TUG and the transitions between them (ie, sit-to-stand, walking, turning, turn-to-sit) from the sensor signals. Further, it calculates features from these phases, such as duration, velocity, jerkiness, and signal range, as well as gait features including number of steps, step duration, and gait speed. For the Self-STS, the algorithms analyze the sensor signals and calculate several features from the whole task, transitions, and separate sit-to-stand and stand-to-sit phases of each repetition. Finally, for Self-Tandem, the algorithms analyze the sensor signals and calculate features such as signal frequency, ellipse area, velocity, sway path, jerkiness, signal range, and spectral entropy.

Version 1

A multidisciplinary team designed the apps with emphasis on ease of use for the target group, corresponding to the term "perceived satisfaction" in the ISO terminology [9]. This included displaying buttons and icons in relatively large sizes and using contrasting colors on a white background. In addition, to ease the demands on working memory, the app screens were designed with as few elements and text as possible.

All apps are based on the same structure (Figure 1). For example, when opening the self-TUG app, a green "play" button appears. Pressing the button prompts a dialog box with a 5-second countdown and a red stop button. The countdown gives the user time to attach the smartphone to the lower back by means of a waist belt case (see Procedures). After the countdown and as soon as no movement is detected by the inertial sensors, an audio signal initiates the start of the test. At the end of the test, when the user is again sitting still, an audio signal indicates that the test is completed. The Self-STS has the same structure (ie, audio signals for both the start and end of the test when the participant is sitting still). One important difference for the Self-Tandem is that the start and end signals are activated by time and not by movement. Thus, the audio start signal is initiated immediately after a 5-second countdown, followed by the end signal after 15 seconds.

Version 2

In version 2 of the Self-TUG and Self-STS, the smartphone was worn in the front trouser pocket instead of the waist belt case. We also integrated instruction videos into the apps. By pressing "play," a dialog box appears with a question asking whether the participants want to see the instruction video (with a yes/no choice; Figure 2). Pressing "yes" starts the instruction video for how to perform the self-test. Pressing "no" results in the question "Do you want to start the test?" with the options "Yes" or "No, play the instruction video." A reminder of what to do (insert phone in pocket for Self-TUG and Self-STS, hold against chest for Self-Tandem) was added to the countdown dialog box. The apps were otherwise similar in structure as in version 1.

Version 3

The upgrades made for the third version of the apps included new instructional videos, with updated voiceover and footage, and new graphical elements in the video to emphasize important details of how to perform the tests (Multimedia Appendix 1) as well as a new menu structure where the user could choose to view instructions or start the test. Instructions consisted of a submenu with two options: watch the instructions for how to prepare the test setup or how to perform the test.

In addition, new features (Figure 3) included a warning message that popped up if a user tried to perform a test without having watched both instructions; voiceover that instructed the user on what to do once the test sequence had been initiated (ie, "Put the phone gently in your right pocket. Sit down and wait for my

instructions"); and real-time verbal feedback based on the inertial signals from the smartphone (eg, "sit down," "get up from the chair," "proceed with the test," and count of repetitions for the Self-STS). The instruction videos were made for Norwegian study participants; thus, voiceover and text elements were in Norwegian. The text on the menu and dialog box was automatically adapted to the system language of the phone.

Figure 1. Screenshots of the first version of the Self-Timed Up and Go test.

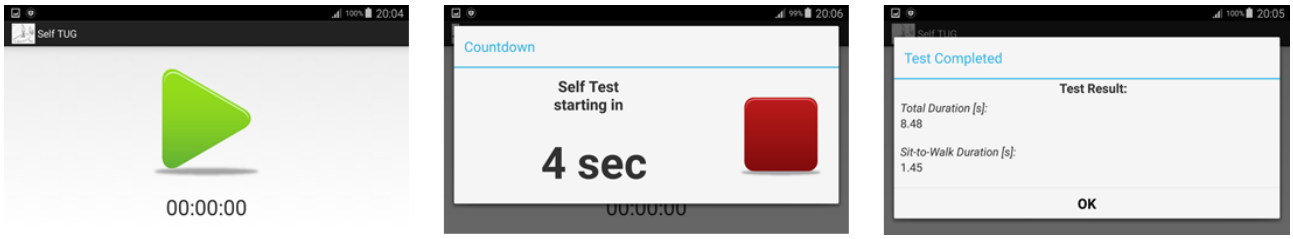


Figure 2. Screenshots of the second version of the Self-Timed Up and Go test.

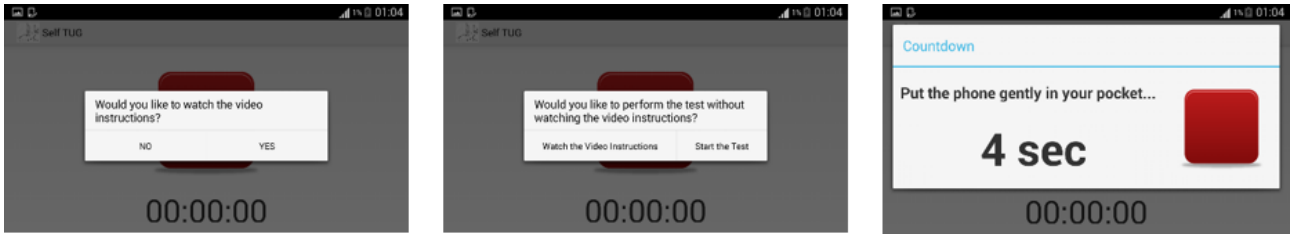
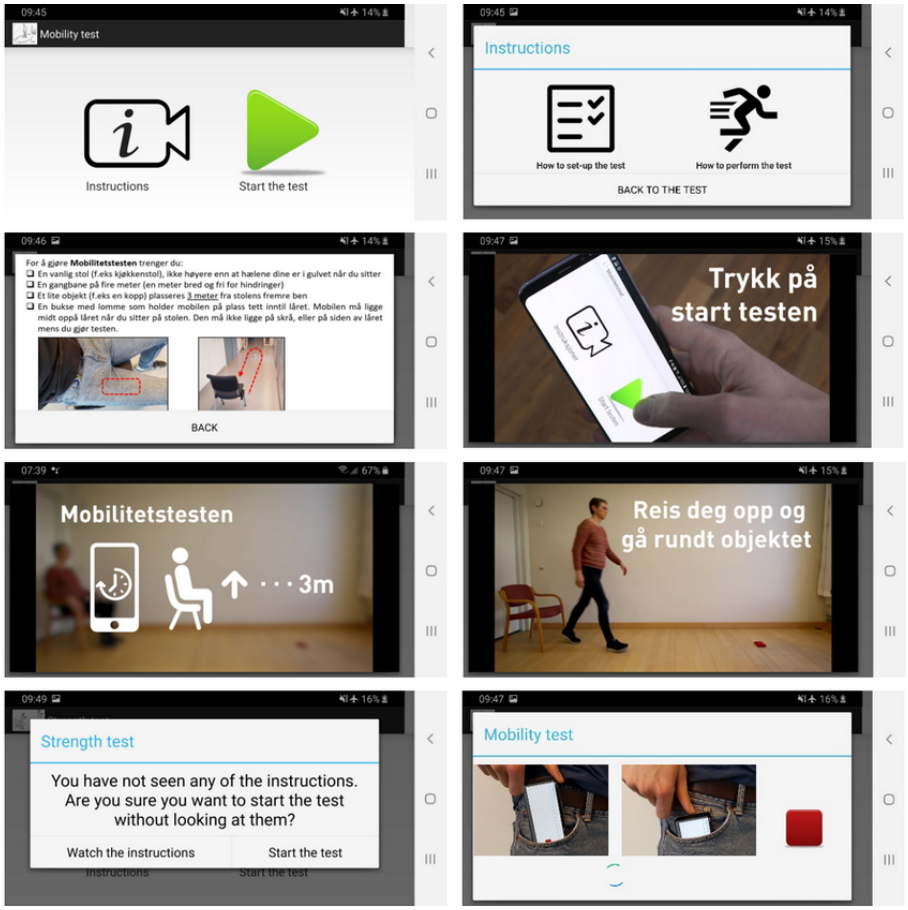


Figure 3. Screenshots of the third version of the Self-Timed Up and Go, including the start menu, instructions menu, test setup, instruction video screenshots, warning prompt when trying to start the test without having opened the instructions first, and instructions after starting the test.



Procedures

Testing of Version 1

The testing of version 1 was carried out in a lab setting by trained assessors. Before testing, the assessors prepared the setup, which included a chair placed against the wall with a 3-meter walkway in front of it and a beanbag at the 3-meter mark. Following a standardized introduction of the general purpose and procedures of a usability test, the participants were asked to self-administer the tests using the app on the

smartphone. The participants did not receive any guidance from the assessors during the tests, and the materials they needed to perform the test were placed on a chair in front of them. This included the smartphone, written instructions, and a belt case for wearing the phone during the Self-TUG.

The assessors observed while the participants attempted to self-administer the tests, recording issues and errors on a sheet with predefined errors and issues that we expected in addition to a free-text box to record other errors and issues.

Testing of Version 2

For the second step, the Self-STS was added to the self-test battery. Testing happened under the same conditions as during testing of the first version, with one exception for the Self-TUG and Self-STS, namely that the smartphone placement was changed from the belt case to the front trousers pocket.

Testing of Version 3

An instructor visited the participants in their homes to get a realistic impression of how the system would be used in a real-world home setting. After a standardized introduction, the participants were asked to prepare and self-administer each of the 3 self-tests 3 times, without guidance from the instructor. One GoPro camera (GoPro, San Mateo, CA) was attached with a harness to the participant's chest and worn during the test sequence to record the participant's interaction with the smartphone. A second GoPro camera was placed in the room in a position where all movements could be recorded. The participants were encouraged to think aloud when using the system. After performing the self-tests, we asked participants to complete two questionnaires: User Experience Questionnaire (UEQ) and Systems Usability Scale (SUS) [11]. This was followed by an audiotaped semistructured interview that was developed specifically for this study, where we aimed to collect end users' views on topics relevant to the apps, such as user experience, feedback/results, suggestions for improvements, and general usefulness of the apps.

Data Processing and Analyses

We defined errors as deviations from the test instructions and counted the number of errors from the clinical record forms in the first and second iterations and from video recordings in the third iteration (Multimedia Appendix 2). SUS scores, based on 5-point Likert-scale items, were averaged for each participant and converted into a usability score with a range from 0 to 100 (with a higher score indicating better usability). UEQ Likert-scale items were scored from 0 (highly agree) to 4 (highly disagree), and frequencies of responses within each category across all items were calculated.

The interview transcripts were analyzed using thematic analysis [12] to identify relevant themes. Quotes were extracted for each subtheme and translated from Norwegian to English for analysis. The questions presented to the participants were "What did you think about using these apps to test your physical function?" and "Do you have any ideas for how the apps can be improved?"

Ethics

The data collection was performed in accordance with the Helsinki ethical guidelines. The first and second usability testing phases were approved by the ethical committees in Norway (REK midt, 2016/1891), Stuttgart (registration number 770/2016BO1), and Amsterdam (METc VUmc registration number 2016.539 [NL59977.029.16]). The Norwegian Center for Research Data approved that the data protection for the third usability testing was in accordance with current regulations (ref. no. 391684). All participants included in this study gave their informed consent.

Results

Participants' characteristics are presented in Table 1.

Table 1. Participant characteristics.

Cohort	App version	n	Age (years), mean (SD)	Male gender, n (%)	Has smartphone experience, n (%)	Years of education, mean (SD)
PreventIT study	1	189	66.3 (2.4)	90 (47.4)	157 (83.1)	15.6 (4.6)
PreventIT study	2	134	66.3 (2.5)	64 (47.8)	108 (80.0)	15.9 (4.8)
Summative usability evaluation	3	20	68.7 (5.2)	11 (55.0)	20 (100.0)	__[a]

[a]Data were not collected.

The usability problems identified, numbers of participants who experienced these problems, and what was done to eliminate or reduce these problems are presented in Tables 2-4.

Table 2. Usability problems in version 1 of the Self-Timed Up and Go and Self-Tandem apps, rate of errors, and solutions (n=189).

Problem ID	Usability problem	Rate of errors/trials	Improvements made
1	Incorrect performance	120/378 (32%)	Added instruction video
2	Performed test without starting app	22/378 (6%)	Implemented instruction video that clearly demonstrates that the play button needs to be pressed before performing the test
3	Did not sit still and wait for start signal after test was started	23/189 (12%)	Added instruction video (demonstrating sitting still and waiting for the start signal before starting the test) and shortened the delay in the algorithm to limit any confusion
4	Incorrect placement of phone	32/378 (8%)	Changed placement to front pocket for Self-TUG[a] and Self-STS[b] and added a reminder in the countdown screen on what to do first (eg, “put the phone gently in your pocket”)
5	Did not hear/perceive instructions	18/378 (5%)	Changed placement to front pocket for Self-TUG and Self-STS
6	Accidentally cancelled the test	15/378 (4%)	Not possible to override the home button function in the android OS, change of placement the preferred solution to reduce this problem

[a]TUG: Timed Up and Go.

[b]STS: Sit to Stand.

Table 3. Usability problems in version 2 of all 3 self-tests, rate of errors, and solutions (n=134).

Problem ID	Usability problem	Rate of errors/trials	Improvements made
1	Incorrect performance	66/402 (16%)	Added new, improved instructions to the videos (new voiceover and added graphical elements to draw attention to the details of the test procedures); added a warning message that appears if trying to start the test without watching instructions; and added real-time TTS[a] voice feedback on the number of repetitions in the Self-STS[b]
2	Started performing test (during instruction video) without starting the test in the app	28/402 (7%)	Changed structure of the app: main window now has two separate buttons, one for “start test” and one for “instructions”
3	Did not sit still and wait for start signal after test was started in the app	39/268 (15%)	Added real-time verbal step-by-step instructions that are initiated after the test is started in the app
4	Incorrect placement of phone	11/402 (3%)	Added real-time verbal instruction explaining where to place the phone and when to do this
5	Did not hear/perceive instructions	4/402 (1%)	Changed settings in the app so that the volume is always on maximum levels during testing, to prevent participants from accidentally pressing the “volume down” button
6	Accidentally cancelled the test	8/402 (2%)	Reduced the size of the “stop” button

[a]TTS: text-to-speech.

[b]STS: Sit to Stand.

Table 4. Usability problems identified in version 3 of all 3 self-tests and the rate of errors (n=20).

Problem ID	Usability problem	Rate of errors/trials
1	Incorrect performance	19/60 (32%)
2	Performed test (during instruction video) without starting the test in the app	5/60 (8%)
3	Did not sit still and wait for start signal after test was started in the app	0/40 (0%)
4	Incorrect placement of phone	0/60 (0%)
5	Did not hear/perceive instructions	2/60 (3%)
6	Accidentally cancelled the test	1/60 (2%)

Iteration 1

In total, at least 1 error was made in 120 of 378 (32%) trials during the first usability testing with the Self-TUG and Self-Tandem. Forgetting or misunderstanding the written instructions were the leading causes of errors. In order to reduce the errors caused by this usability problem, we created video instructions to replace the written instructions.

Iteration 2

In the second usability test, errors due to usability problem 1 (incorrect performance of test) were made in 66 of 402 trials (16%). Percentage of errors due to usability problems 2 (performs test without starting app) and 3 (did not sit still and wait for start signal after test was started) increased from the first usability test, while the frequency of problems 4-6 (incorrect placement of phone, did not hear/perceive instructions, accidentally cancelled the test, respectively) decreased.

Iteration 3

In the third summative usability evaluation, errors due to usability problem 1 (incorrect performance of test) were made in 19 of 60 (32%) trials. Usability problems 3 (did not sit still and wait for start signal after test was started) and 4 (incorrect placement of phone) were eliminated, while the frequencies of usability problems 2 (performs test without starting app), 5 (did not hear/perceive instructions), and 6 (accidentally cancels the test) remained similar.

Table 5 presents an overview of the proportions of correctly performed (first) trials of self-tests for all tests in all iterations.

Table 5. Number of correctly performed self-tests (first trial) with versions 1, 2, and 3.

	Self-TUG[a], n (%)	Self-STS[b], n (%)	Self-Tandem, n (%)
Testing of version 1	42 (22.0)	N/A[c]	127 (67.2)
Testing of version 2	108 (83.1)	40 (30.1)	106 (79.1)
Testing of version 3	14 (70.0)	5 (25.0)	18 (90.0)

[a]TUG: Timed Up and Go.

[b]STS: Sit to Stand.

[c]Not yet developed.

Perceived Ease of Use

UEQ scores for iteration 3 are presented in Table 6, indicating a positive or very positive user experience on all 6 items. Seven sub-themes of perceived ease of use were identified in the analysis of interview transcripts and are presented in Multimedia Appendix 3 with accompanying sample quotes, mapped to proposed solutions.

Table 6. Frequency of scores across the 6 items in the User Experience Questionnaire administered in the summative user evaluation (n=19).

Likert scale items	Strongly agree	Agree	Neither agree/disagree	Disagree	Strongly disagree
The set-up instructions were clear and easy to follow	14	4	1	0	0
The text in the app was easy to read	15	4	0	0	0
The buttons in the app were easy to discern from other elements	12	5	0	2	0
The signals were easy to hear	15	3	0	1	0
It was easy to navigate around in the apps	10	6	0	1	1
The instruction videos were clear and easy to follow	14	5	0	0	0

The mean score on the SUS for version 3 was 77.63 points (SD 16.1 points, range 42.5-97.4 points). Of the 20 participants, 14 participants scored the SUS above 66.5 points, which is the average SUS score for cell phones [13].

Discussion

Principal Findings

This paper describes the development and usability testing of the Self-TUG, Self-STS, and Self-Tandem through 2 iterations in the lab and 1 in a home setting. Our aim was to develop app-based, self-administrable tests of physical function that participants could use with a high degree of effectiveness and perceived ease of use. The first phase of testing revealed usability problems that affected the validity of the test results, illustrating a clear need for improvements. We addressed all usability problems by making changes to the app design, test algorithms, and test setup, which led to a large decrease in the number of trial errors in the second usability testing. The work on the third version of the apps then started, which included updating and adding new instructions for a version fully adapted for use in a home-based setting.

The results from the SUS, UEQ, and thematic analysis from the usability testing in the home setting indicated that the participants experienced high levels of perceived ease of use when using the apps. Still, errors were made that may affect the validity of the tests, most of which were caused by misunderstanding the instructions. As an example, the most common error for Self-STS was not performing it as fast as

possible, which was the main reason why only 25.0% (5/20) performed it correctly on their first attempt. This is lower than in the second version tested in the lab (40/134, 30.1%). This misunderstanding was caused by a delay in the real-time counting of repetitions that was implemented in the third version of Self-STS. The verbal announcement of repetitions, which is also done by the assessor when the original Five Times STS is administered in the clinic, was implemented to motivate the participant to perform it faster and as a way for the participant to keep track. However, as can be seen in the sample quotes from the participant interviews (Multimedia Appendix 3), there was a delay in the real-time feedback, making many of the participants stop and wait in a standing position for the TTS to announce the repetition before sitting down. This slowed down the performance and thus impaired instead of improving the validity of the test.

During the Self-TUG, a common error was to measure an incorrect distance for the walkway during the set-up. Although the instructions state that the walkway should be 3 meters, the participant responses indicate that they did not consider it crucial to measure exactly 3 meters. However, it has to be exactly 3 meters if the total test duration, walking duration, or gait speed is to be used as an outcome measure, as these features rely on a standardized distance walked. A clarification in the instruction, where it is specified that the walkway needs to follow a straight line of exactly 3 meters, could be one way to increase the reliability of the test. However, as the distance walked by the participant cannot be accurately measured by the app, another approach could be to only exploit the distance-independent signal features, such as those calculated from the sit-to-stand, turning, and turn-to-sit phases. This will improve the system reliability in assessing motor performance, but it will not ensure full compatibility with the standard clinical measure of the total test duration.

Another common error with the Self-TUG was to press “Start test” without watching the instructions first. Although we had implemented a pop-up warning if this happened, a bug prevented this from happening in 5 of the 7 times this occurred. For the 2 participants who received the pop-up warning, 1 ended up watching the instruction video and performing the test correctly, while the other ignored the warning and proceeded to perform the test without watching the instruction video, thus performing it incorrectly. Because of the bug, we do not have sufficient information to make a safe claim regarding the effectiveness of the warning message. However, we assume that this problem will be resolved by fixing the bug and specifying in the warning message that a correct trial depends on having watched the instruction video first.

A common usability problem observed with the Self-Tandem, and also mentioned by many of the participants in the interviews, was the discrepancy between the instructions and actual duration required to stand in the tandem position. The instructions state that the participant is supposed to place the feet in tandem, hold the phone against the chest, and, after hearing the start signal, keep as still as possible for 15 seconds. What often happened in the current version, however, was that the app tried to detect and verify the position of the smartphone after the participant had been instructed to place the feet in a tandem position. The TTS then instructed the participant to hold the tandem position and keep as still as possible for 15 seconds, until the end signal. Depending on how fast they placed the phone against the chest, the participant could thus stand up to 25 seconds in total. However, if the instructions had said that the participant should assume the tandem position after hearing the start signal, different people would likely need a different amount of time to get into the correct position, thereby risking that we would get less than 15 seconds of actual tandem balancing. The outcome measure in Self-Tandem is mediolateral sway, which was found to be a strong predictor of age-related decline in a study in which an eyes-open condition was used [14]. We therefore designed the test in a way that would ensure, or at least increase the chance, that we would have at least 15 seconds of the participant standing in tandem.

A limitation with the Self-Tandem test is that we cannot infer whether the participant was keeping the correct tandem position for the entire 15 seconds from the inertial sensor signals. This is not true for the Self-STS and Self-TUG tests, where the correct performance of all phases of the test can be identified reliably from the signal. A potential solution could be to implement a pop-up question where the user self-reports whether they actually held the position for the entire duration. Such a solution has been implemented in the mHealth app “Steady” [14]. Steady is a falls risk app that consists of a health history questionnaire, 4 balance tasks (eyes open, eyes closed, tandem, and single leg), and a 30-second sit-to-stand test. The binary outcome measure of whether a user is able to complete a static steady-state balance task in various conditions and durations, such as those used in Steady, has been used extensively in studies assessing healthy young seniors [15].Therefore, adding this feature could potentially increase the Self-Tandem’s diagnostic/prognostic abilities.

Implications and Future Work

The 3 iterations of usability testing described in this paper were sufficient to identify all major usability problems with the self-tests. The only problem remaining after the third cycle is the real-time counting in the Self-STS, described earlier in the discussion, which can easily be fixed.

We have demonstrated what challenges can be expected when developing app-based tests of physical function for seniors and how solutions to specific usability problems identified in one iteration affected the same problems in the next iterations (Tables 2-5). In addition, we described the perspectives of the seniors regarding their experience of using the apps to self-administer the tests (Multimedia Appendix 3). Another interesting insight is how going from the lab to a home-setting influenced the type of usability problems observed, in particular those related to the test setup. In the lab setting, the setup was prepared beforehand, whereas in the home setting, participants needed to follow the instructions in the app describing how to measure the walkway in the TUG, secure the chair for Self-STS, and perform the Self-Tandem in a spot with a secure object within hands reach, without any guidance from the assessor.

The next step in the developmental process of the apps is to implement the solutions proposed in Multimedia Appendix 3 to address the remaining usability problems and conduct new

usability tests to ensure that the apps are ready to be used by the target group to self-administer the tests safely and correctly. Furthermore, the algorithms used for signal processing in Self-TUG and Self-STS need to be validated with the changed placement of the smartphone from the lower back (version 1) to the thigh (versions 2 and 3). Although we experienced some issues with this new placement in the usability testing (eg, some trousers were too loose), with the smartphone tilting down on either side of the thigh and making the trial invalid, we nevertheless believe that this solution offers the best trade-off between motion detection ability on one side and ease of use on the other side.

Our app-based tests of physical function could be applicable in many contexts, and different contexts may require different test outcomes. In the current version of the apps, the results presented to the user after performing the tests are total durations for the Self-TUG and Self-STS and sway path distance in the Self-Tandem. As discussed, however, these might not be feasible to exploit from an unsupervised test, where correct test set-up cannot be verified. The data processed by the inertial sensors within the smartphone provide additional features, and we aim to review existing literature to identify which of the signal features from instrumented versions of the TUG, Five Times STS, and standing tandem are the most predictive of functional decline in seniors. Given that these features can be reliably measured with the smartphone worn in the trouser pocket, they will be exploited as outcomes presented to the app users.

Although many tests of physical function have been instrumented by the use of smartphones, the authors are only aware of one other app that is developed for unsupervised self-assessment, the Steady app [14]. What separates the Steady app from ours is the type of tests implemented in the app. In addition to static balance and repeated sit-to-stands, we integrated the instrumented TUG. Furthermore, we performed usability assessments of the app in the participants' own homes, in contrast to Steady, where an unoccupied apartment was used for all non-lab test sessions in order to mimic a home environment.

Limitations

In our first 2 usability tests, the apps were tested by 189 and 134 participants, respectively. Although this was very useful for identifying what did and did not work well, we might have achieved similar results with fewer participants. Earlier studies have suggested that as few as 12 test users can be sufficient to detect the majority of usability problems [16]. Thus, with shorter and faster test cycles, the apps could potentially have been at a more mature stage today.

The participants in the summative usability evaluation differed to those from the PreventIT study in terms of age and smartphone experience. This makes it more difficult to say something about the impact that each app improvement had, as opposed to testing all app versions in 3 different, but homogenous, cohorts. However, we see it as an advantage that the apps are also tested in a slightly older cohort, as these participants can help us identify problems that could be more relevant to how they experience the usability of the apps, as compared to seniors that are younger or more experienced with technology in their daily life. Furthermore, the self-tests have not been validated in persons with tremor or pathologies; thus, the results do not necessarily generalize to these populations.

ISO's definition of usability comprises 3 main aspects: effectiveness, which is the accuracy and completeness with which users achieve certain goals; efficiency, which is the relationship between the accuracy and completeness with which users achieve certain goals and the resources expended in achieving them; and satisfaction, which is the user's comfort with and positive attitudes towards the use of the system [17]. Efficiency was not measured in our usability studies. It is often measured as task completion time or learning time, but in the context of testing physical function, where the time spent on completing a task also depends on the person's physical abilities, we did not consider task completion time to be an appropriate outcome measure of usability, but rather of functional level of the participant.

Although we have assessed the usability of these apps and identified solutions to the remaining usability problems, the validity of the outcome measures from the tests also needs to be further investigated before being made available to end users. Another point worth mentioning is that the correct use of the apps and, accordingly, valid test results could be ensured by giving the end users a one-time demonstration of how to use apps and perform the tests correctly. This could be given in a home visit or in an appointment at the lab or clinic, depending on the context of use.

Conclusion

The study results suggest that the apps have the potential to be offered as a solution for self-testing of physical function in a nonsupervised, home-based setting. Participants found the apps easy to use. The summative user evaluation in a home setting revealed important usability problems that were not identified in the lab, highlighting the importance of utilizing both test settings when assessing app usability. The current version of the apps has some remaining usability problems that can affect the test results, indicating that the apps need to be further improved and then validated before being made available to end users.

Acknowledgments

The authors wish to thank Andreas Hansen Schille and Jonas Langset Hustad for their contribution to the editing of the instruction videos and Gemma Goodall for proofreading the manuscript. We thank all participants for being part of the study and providing us with valuable input.

This study was supported by PreventIT, which received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 689238. The content is solely the responsibility of the authors and does not necessarily represent the official views of the European Commission. The European Commission had no role in the design, execution, interpretation, or writing of the study.

Conflicts of Interest

RB, BV, MC, JLH, and KT declare no conflict of interest. SM holds a share of the mHealth Technologies srl company, which owns the rights to some of the movement analysis algorithms.

Multimedia Appendix 1
Video instruction available within version 3 of the Self-TUG app, demonstrating how to perform the Self-TUG, with voiceover in Norwegian.
[MP4 File (MP4 Video), 21012 KB - mhealth_v8i4e16507_app1.mp4]

Multimedia Appendix 2
Detailed description of usability problems.
[DOCX File , 16 KB - mhealth_v8i4e16507_app2.docx]

Multimedia Appendix 3
Sub-themes within perceived ease of use and sample quotes from participant interviews following the third iteration of usability testing and proposed solutions.
[DOCX File , 15 KB - mhealth_v8i4e16507_app3.docx]

References

1. Olesen K, Rod NH, Madsen IEH, Bonde JP, Rugulies R. Does retirement reduce the risk of mental disorders? A national registry-linkage study of treatment for mental disorders before and after retirement of 245,082 Danish residents. Occup Environ Med 2015 May 26;72(5):366-372 [FREE Full text] [doi: 10.1136/oemed-2014-102228] [Medline: 25814269]
2. Choy NL, Brauer S, Nitz J. Changes in postural stability in women aged 20 to 80 years. J Gerontol A Biol Sci Med Sci 2003 Jun 01;58(6):525-530. [doi: 10.1093/gerona/58.6.m525] [Medline: 12807923]
3. Isles R, Choy N, Steer M, Nitz J. Normal values of balance tests in women aged 20-80. J Am Geriatr Soc 2004 Aug;52(8):1367-1372. [doi: 10.1111/j.1532-5415.2004.52370.x] [Medline: 15271128]
4. Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. Gait Posture 2012 May;36(1):163-165. [doi: 10.1016/j.gaitpost.2012.02.006] [Medline: 22421189]
5. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
6. Anderson MP, A. Pew Research Center. 2017. Technology use among seniors URL: https://www.silvergroup.asia/wp-content/uploads/2017/07/Technology-use-among-seniors-_-Pew-Research-Center.pdf [accessed 2019-08-01]
7. Weiss A, Herman T, Plotnik M, Brozgol M, Giladi N, Hausdorff J. An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers. Physiol Meas 2011 Dec;32(12):2003-2018. [doi: 10.1088/0967-3334/32/12/009] [Medline: 22094550]
8. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
9. ISO. Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability. Geneva: International Organization of Standardization; 1998. Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability Genewa, Switzerland 1998; URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en [accessed 2019-08-01]
10. Taraldsen K, Mikolaizak AS, Maier AB, Boulton E, Aminian K, van Ancum J, et al. Protocol for the PreventIT feasibility randomised controlled trial of a lifestyle-integrated exercise intervention in young older adults. BMJ Open 2019 Mar 20;9(3):e023526 [FREE Full text] [doi: 10.1136/bmjopen-2018-023526] [Medline: 30898801]
11. Brooke J. SUS - A quick and dirty usability scale. In: Jordan P, Thomas B, Weerdmeester B, McClelland I, editors. Usability Evaluation In Industry. London: Taylor and Francis; 1996:189-194.
12. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
13. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
14. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]

15. Bergquist R, Weber M, Schwenk M, Ulseth S, Helbostad J, Vereijken B, et al. Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. BMC Geriatr 2019 Jan 09;19(1):9 [FREE Full text] [doi: 10.1186/s12877-018-1011-0] [Medline: 30626340]
16. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT. 169166: ACM; 1993 Presented at: Conference on Human Factors in Computing Systems; 1993; Amsterdam, The Netherlands p. A-13. [doi: 10.1145/169059.169166]
17. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? 332455: ACM; 2000 Presented at: The SIGCHI conference on Human Factors in Computing Systems; 2000; The Hague, The Netherlands p. a-52. [doi: 10.1145/332040.332455]

Abbreviations

ISO: International Organization for Standardization.
mHealth: mobile health.
STS: Sit-to-Stand.
SUS: Systems Usability Scale.
TUG: Timed Up and Go.
UEQ: User Experience Questionnaire.

Edited by C Dias; submitted 09.10.19; peer-reviewed by F Lamers, K Hsieh, E Sadeghi-Demneh, JP Leuteritz, J Knitza; comments to author 19.11.19; revised version received 07.01.20; accepted 21.02.20; published 27.04.20.

Please cite as:
Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K
App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study
JMIR Mhealth Uhealth 2020;8(4):e16507
URL: http://mhealth.jmir.org/2020/4/e16507/
doi:10.2196/16507
PMID:32338616

Original Paper

Usability of Food Size Aids in Mobile Dietary Reporting Apps for Young Adults: Randomized Controlled Trial

Ying-Chieh Liu[1,2], PhD; Sheng-Tang Wu[3], MD; Shan-Ju Lin[1], MSc; Chien-Hung Chen[4], MSc; Yu-Sheng Lin[2], MD; Hsin-Yun Chen[5], MSc

[1]Department of Industrial Design, College of Management, Chang Gung University, Taoyuan, Taiwan

[2]Health Promotion Center, Department of Internal Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]Division of Urology, Department of Surgery, Tri-Service General Hospital and National Defense Medical Center, Taipei, Taiwan

[4]Cyber Security Technology Institute, Institute for Information Industry, Taipei, Taiwan

[5]Department of Nutrition Therapy, Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Ying-Chieh Liu, PhD
Department of Industrial Design
College of Management
Chang Gung University
No 259, Wen-Hwa 1st Road
Kwei-Shan
Taoyuan, 333
Taiwan
Phone: 886 3 2118800 ext 5250
Email: ycl30@mail.cgu.edu.tw

Abstract

Background: Young adults are more likely to use self-managed dietary reporting apps. However, there is scant research examining the user experience of different measurement approaches for mobile dietary reporting apps when dealing with a wide variety of food shapes and container sizes.

Objective: Field user experience testing was conducted under actual meal conditions to assess the accuracy, efficiency, and subjective reaction of three food portion measurement methods embedded in a developed mobile app. Key-in–based aid (KBA), commonly used in many current apps, relies on the user's ability to key in volumes or weights. Photo-based aid (PBA) extends traditional assessment methods, allowing users to scroll, observe, and select a reduced-size image from a set of options. Gesture-based aid (GBA) is a new experimental approach in which the user makes finger movements on the screen to roughly describe food portion boundaries accompanied by a background reference.

Methods: A group of 124 young adults aged 19 to 26 years was recruited for a head-to-head randomized comparison and divided into 3 groups: a KBA (n=42) control group and PBA (n=41) and GBA (n=41) experimental groups. In total, 3 meals (ie, breakfast, lunch, and dinner) were served in a university cafeteria. Participants were provided with 25 dishes and beverages for selection, with a variety of food shapes and containers that reflect everyday life conditions. The accuracy of and time spent on realistic interaction during food portion estimation and the subjective reaction of each aid were recorded and analyzed.

Results: Participants in the KBA group provided the highest accuracy in terms of hash brown weight (P=.004) and outperformed PBA or GBA for many soft drinks in cups. PBA had the best results for a cylindrical hot dog (P<.001), irregularly shaped pork chop (P<.001), and green tea beverage (660 mL; P<.001). GBA outperformed PBA for most drinks, and GBA outperformed KBA for some vegetables. The GBA group spent significantly more time assessing food items than the KBA and PBA groups. For each aid, the overall subjective reaction based on the score of the System Usability Scale was not significantly different.

Conclusions: Experimental results show that each aid had some distinguishing advantages. In terms of user acceptance, participants considered all 3 aids to be usable. Furthermore, users' subjective opinions regarding measurement accuracy contradicted the empirical findings. Future work will consider the use of each aid based on food or container shape and integrate the various advantages of the 3 different aids for better results. Our findings on the use of portion size aids are based on realistic and diverse food items, providing a useful reference for future app improvement of an effective, evidence-based, and acceptable feature.

Trial Registration: International Standard Randomized Controlled Trial Registry ISRCTN36710750; http://www.controlled-trials.com/ISRCTN36710750.

(JMIR Mhealth Uhealth 2020;8(4):e14543) doi:10.2196/14543

KEYWORDS

portion size measurement; prototype; user-centered design; dietary reporting; mobile health; randomized controlled trial

Introduction

Background

Young adults are increasingly concerned with healthy eating habits [1-3]. Younger people are more likely to use mobile health (mHealth) apps [4,5], and a broad range of mHealth solutions is available to facilitate self-management of diet, fitness, and weight control [6-9]. Self-managed dietary intake apps are among the most popular apps in the health and fitness category [10]. One major area of mHealth apps has focused on the tracking of daily food intake, nutrient information, and calorie counting based on users' assessment of their dietary intake [11,12]. However, little research has examined the effectiveness and user acceptance of such apps [13]. A better understanding of how users actually use mHealth apps is required [9].

Assessment accuracy of dietary intake has a significant impact on energy and nutrient intake calculations [14], and many studies have stressed the need to reduce dietary intake measurement errors [15,16]. One of the major challenges inherent in such reporting is the high degree of variation of food shapes, types, and containers in real-world settings [17]. Some technological approaches have sought to facilitate food size measurement, including the use of augmented reality [18,19]; digital photographs [20-22]; comparative food size measurements [23,24]; wearable cameras [25]; and combinations of text, images, and voice recordings in digital devices [26]. However, these studies were mostly based on highly contrived experimental conditions with little or no follow-up in real-world dining contexts. Few studies have compared mobile dietary reporting apps to evaluate the effectiveness of various methods in measuring relative portion size.

Objectives

This study investigates the effectiveness and limitations of three food size measurement methods in a dietary intake reporting app. Each method has a specific user interaction design and requires different cognitive abilities for task completion. To investigate the practical usage experience, this study seeks to assess the relative effectiveness and acceptability of these different features in tests of young adults reporting their intakes in realistic contexts.

Methods

General Overview of the App and its Reporting Procedure

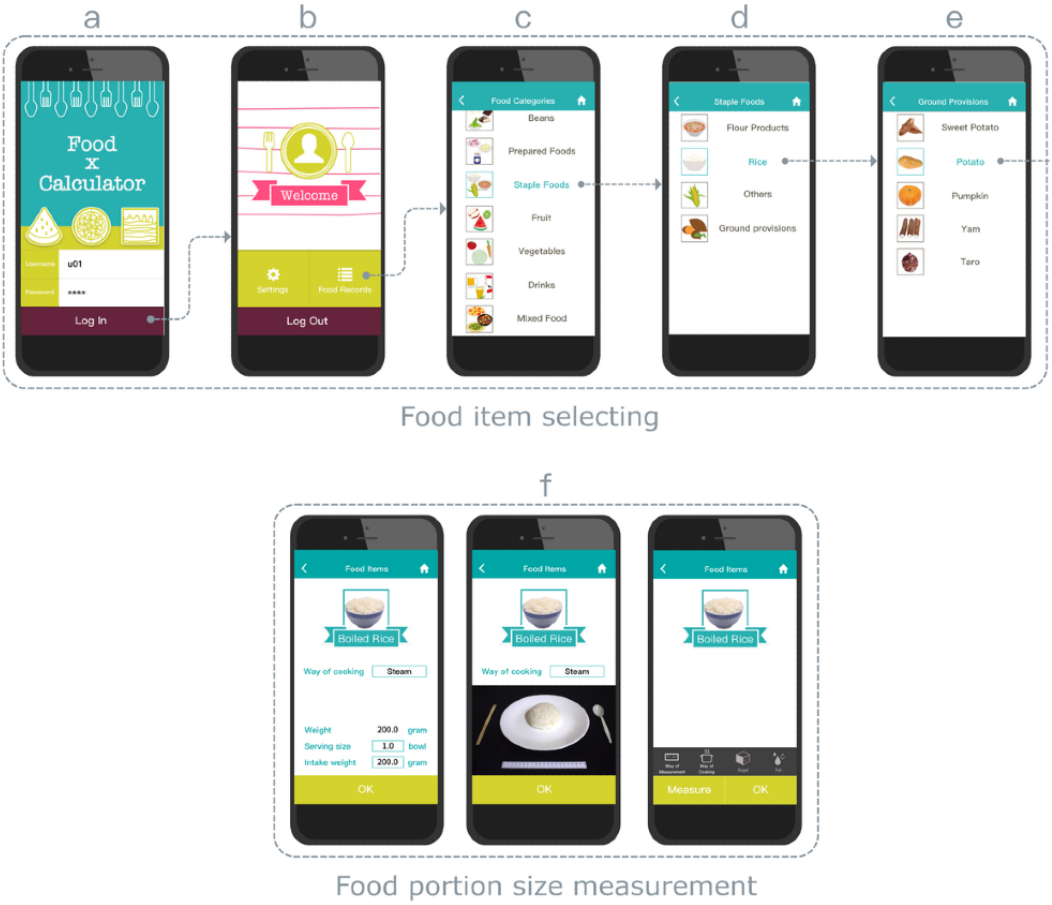
On the basis of user-centered design approaches [27], a new mobile app named DigiDiet (Digital Diet) was developed; it was originally designed for use by patients with metabolic syndrome to improve eating habits. However, DigiDiet functions can also be applied to the dietary tracking and weight management needs of patients with other chronic diseases.

The app's reporting procedures primarily consist of 2 parts. The first part identifies the food or beverage item to be consumed, using a set of tree-like scrollable lists (steps c and d in Figure 1). The first screen provides a scrollable list of food groups, showing 7 groups at a time. Each field (step c in Figure 1) presents a list headed by a colorful image and text description of the group. The second and third screens (steps d and e in Figure 1), respectively, show subgroups and the main ingredients of the selected item. This part allows users to synthesize a dish using 2 or 3 food items, allowing them to account for a broad range of dishes and food items that may not exist in a predefined tree structure food database (see the study by Liu et al [28] for further details).

The second part (see step f in Figure 1) identifies the portion amount (in terms of portion size or weight) of the item to be consumed. This stage includes 2 distinct functions. First, having selected the main food ingredient, food attributes are displayed in individual tabs. Second, the user taps a particular food attribute (eg, *sugar*, *fat*, and *preparation*) and selects a descriptive value (eg, *half sugar*, *low-fat milk*, and *stir fry*).

Before this research, we had developed prototypes for dietary reporting [28,29]. The major enhancement presented in this study is the inclusion of various portion size aids to help users describe the amount of food consumed. Various users described portion sizes in different ways [11], using weight or volume references, household measures (eg, cup, plate, or bowl), and references against commonly used objects (eg, a credit card) or against hand measures (eg, fist or open palm).

Figure 1. High-level description of dietary intake reporting procedure.



Three Measurement Aids

Three different measurement designs were developed to support portion measurement. Key-in–based aid (KBA), which is commonly used in many current apps, relies on the user's ability to key in units of measurement for weight, volume, or household utensils. Photo-based aid (PBA) extends traditional assessment methods [30]. Instead of using life-size paper-based photographs, the aid allows users to scroll, observe, and select a reduced-size image from a set of options. Gesture-based aid (GBA) is a new experimental approach developed based on user finger movements on the screen to roughly describe food volume [31]. Boundaries of the food cross-section and height are prescribed by the user accompanied by a background reference.

In the KBA, the user determines the quantity of the food serving (eg, *1* piece, *0.5* cup, or *1.5* slices), and the KBA then outputs a portion weight in grams (see Multimedia Appendix 1). This approach also allows users to enter specific food weights (in grams) to address the converted amount in proportion to the size of the food serving. As shown in step a in Figure 2, the user taps the serving size field to bring up a 12-button field to log the quantity of food size in terms of a specific unit (eg, piece, bowl, or slice). On the basis of the food item and serving size or food weight, the system then outputs the total calories (step d in Figure 2).

The PBA is an analogical conversion method in which users are provided with life-size food item images on paper [17]. For the purposes of this study, these images are reduced in size and presented on the smartphone screen (see Multimedia Appendix 2). In PBA, users use a one-to-multiple photo selection design to visualize different food quantities. PBA is based on our previous results [29] investigating the potential of applying such measurement methods on mobile devices. We found that this mobile adaptation of PBA can be used to effectively support portion assessment in laboratory testing. This research extends this testing to real-world conditions, including a high degree of variation in food items and container shapes. The pictures of the food items were photographed by the research assistant in accordance with the procedures addressed in the study by Liu et al [29], and these images were integrated into the PBA. Each picture provides a size comparative reference for users. As seen in step c of Figure 3, the user selects one of the several representative photos indicating various proportions. Each photograph includes several objects, including a plate, a pair of chopsticks, a 20-cm ruler, and a spoon (step b in Figure 3). The user then swipes left or right to select the appropriate food portion that best represents the object (steps a-c in Figure 3). Once the user finds an appropriate size relationship, they confirm the selection (step c in Figure 3) and complete the action (step d in Figure 3).

GBA is a new approach that uses on-screen movements to describe portion volume in 3 dimensions (see Multimedia Appendix 3). GBA is an improvement of our previous research that only described 2-dimensional (2D) food shapes [29]. Portion size is described using a top view and a side view, with the user deforming a basic 2D shape (eg, circle, ellipse, rectangle, polygon, or bowl) to fit the outline from each direction. An

irregular 2D shape could be described by repetitive scribbling or dotting to form the desired shape. As shown in step b of Figure 4, the user observes the actual food item and selects an appropriate base shape to initiate measurement, using a credit card as a basis for size comparison. The user then drags and resizes the food item's top view image to an appropriate ratio of a credit card (steps c and d in Figure 4). Next, the user describes and shades the appropriate area of the side view to address the actual food height, as shown in steps e and f of Figure 4. The estimated volume of the actual food is multiplied by the area and height. The user then submits the query for calculation and storage, as shown in steps g and h of Figure 4.

Figure 2. Key-in–based aid interface design and operation.

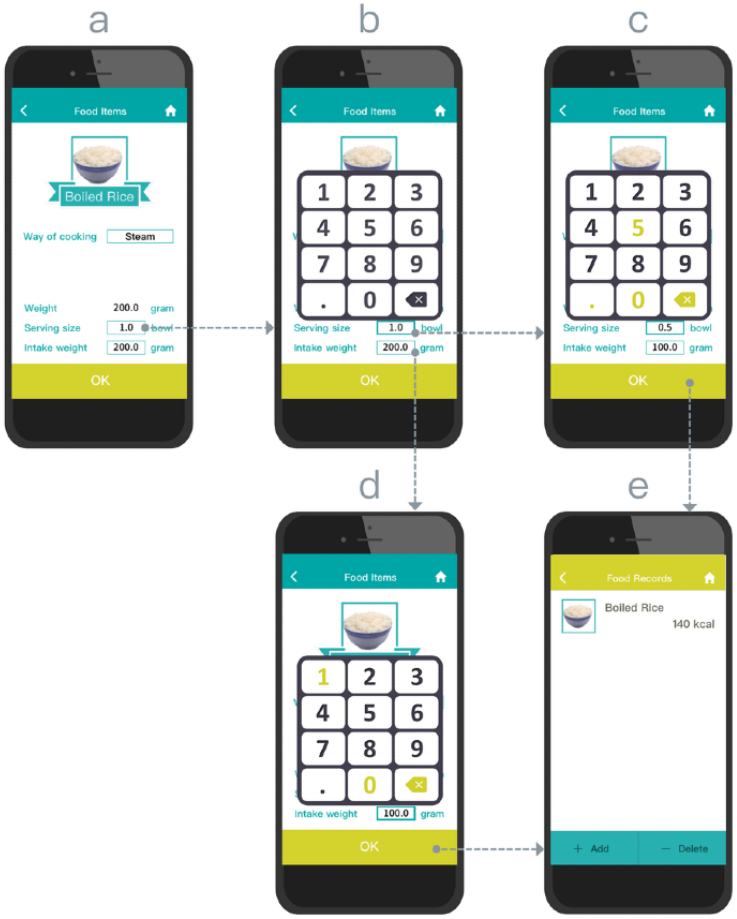


Figure 3. Photo-based aid interface design and operation.

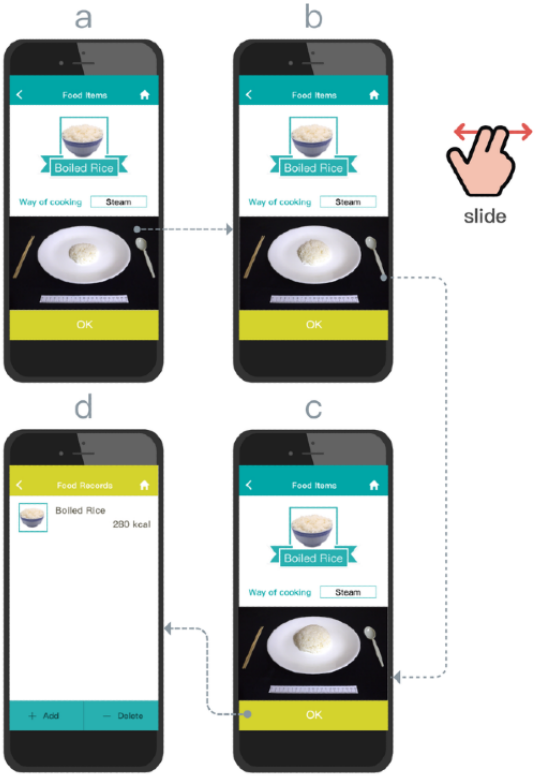
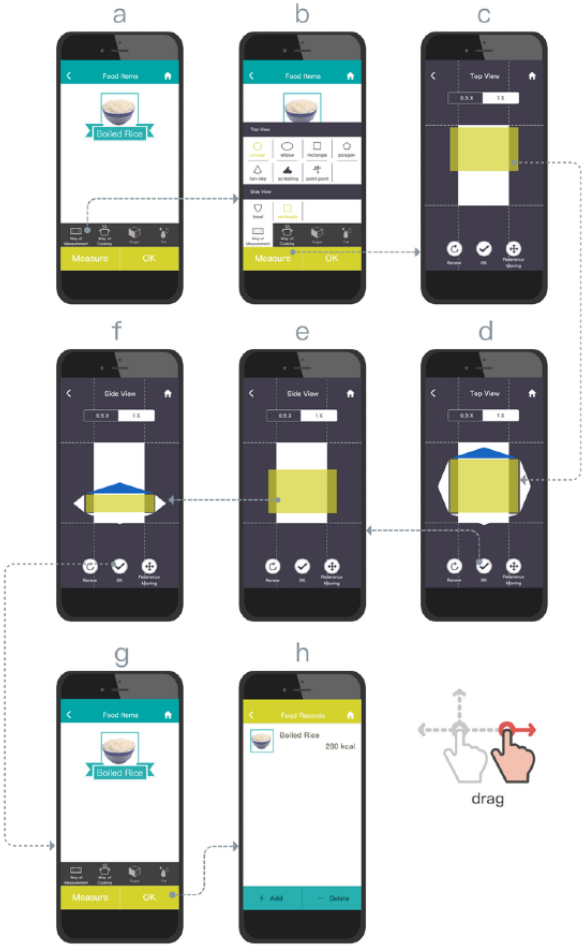


Figure 4. Gesture-based aid interface design and operation.



Study Design for Evaluation

The study was a single-site, assessor-masked, three-armed, parallel-group randomized controlled trial. Each group used a different measurement function to evaluate and compare their respective performance and user perception. An institutional review board (201601817B0) and the ethics committee at the Chang Gung Memorial Hospital approved the design of this study. The study was implemented in the student cafeteria of the Chang Gung University, Taoyuan, Taiwan, between April 2017 and November 2017.

Figure 5 summarizes the experimental process. Respondents were first recruited from among the student population of the Chang Gung University and the Chang Gung University of Science and Technology, both in Taoyuan, Taiwan. Each participant was required to use the developed app for dietary reporting of breakfast, lunch, and dinner in a single day. Eligible participants were university students aged between 19 and 25 years; capable of reading and operating the app on their smartphone; and without diabetes, high cholesterol, or high blood pressure. The exclusion criteria included participants who were currently under any form of dietary control, currently engaged in deliberate weight loss, on medication, or pregnant.

Each participant completed a questionnaire to collect background and baseline data, including age, gender, BMI, academic department, experience with nutrition education, smartphone usage, usage of related apps, and cooking experience (Table 1). Each participant was asked to select intended meal items from a list of 25 food dishes available for consumption in the university cafeteria. Each dish was assessed by a senior dietitian in terms of availability, commonality, preference, and diversity of culinary styles. The food portion, drink containers, and dishes selected were consistent with those commonly seen in the participants' campus cafeteria (see Multimedia Appendix 4). Considering participants' eating habits, we deliberately included both Asian and Western-style dishes, specifically Western food for breakfast and dinner and Asian food for lunch. To further reduce the difficulty of using the app, we prepared specific food items that were considered relatively conducive to food size measurement and recording for the first meal (eg, breakfast). Breakfast included a hash brown, ham, a hot dog, and a beverage. For lunch and dinner, the list allowed participants to select a staple food, the main course, 2 types of vegetables, 2 dishes with mixed food ingredients, and 1 beverage.

Figure 5. Participant flowchart. SUS: System Usability Scale.

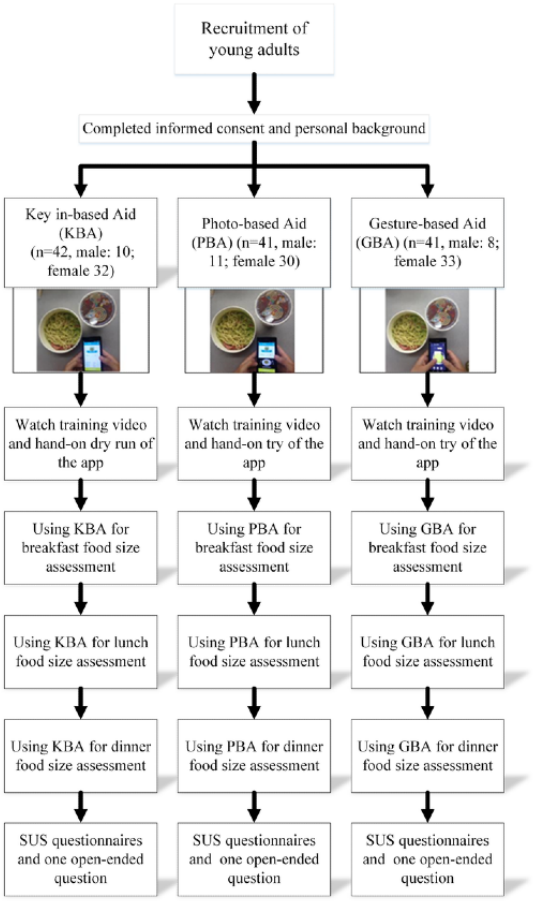


Table 1. Participant characteristics among the key-in–based aid, photo-based aid, and gesture-based aid groups.

Variables	Key-in–based aid (n=42)	Photo-based aid (n=41)	Gesture-based aid (n=41)	Total (N=124)	*P* value
Gender, n (%)					.74
Male	10 (24)	11 (27)	8 (20)	29 (23.4)	
Female	32 (76)	30 (73)	33 (81)	95 (76.6)	
Age (years)[a]					.19
19-20, n (%)	16 (38)	21 (51)	13 (32)	50 (40.3)	
21-26, n (%)	26 (62)	20 (49)	28 (68)	74 (59.7)	
Mean (SD)	21.02 (1.54)	20.83 (1.50)	21.34 (1.74)	100	.34
BMI[a] (kg/m^2), mean (SD)	21.86 (3.47)	21.86 (2.66)	21.32 (2.90)	100	.65
Academic department, n (%)					.54
Health care management	19 (45)	21 (51)	21 (51)	61 (49.2)	
Industrial design	6 (14)	4 (10)	6 (15)	16 (12.9)	
Nursing	7 (17)	3 (7)	5 (12)	15 (12.1)	
Industrial business management	3 (7)	1 (3)	4 (10)	8 (6.5)	
Other	7 (17)	12 (29)	5 (12)	24 (19.4)	
Experience of using diet and nutrition apps , n (%)					.74
Yes	10 (24)	8 (20)	7 (17)	25 (20.2)	
No	32 (76)	33 (80)	34 (83)	99 (79.8)	
Experience of using Android phones, n (%)					.91
Yes	38 (90)	38 (93)	39 (95)	115 (92.7)	
No	4 (10)	3 (7)	2 (5)	9 (7.3)	
Experience of nutrition-related courses , n (%)					.74
Yes	20 (48)	21 (51)	23 (56)	64 (51.6)	
No	22 (52)	20 (49)	18 (44)	60 (48.4)	
Experience of general health education, n (%)					.83
Yes	26 (62)	28 (68)	27 (66)	81 (65.3)	
No	16 (38.1)	13 (32)	14 (34)	43 (34.7)	
Experience in cooking, n (%)					.87
Yes	39 (93)	40 (98)	39 (95)	118 (95.2)	
No	3 (7)	1 (2)	2 (5)	6 (4.8)	

[a]Age and BMI data were analyzed with analysis of variance.

Randomization

A 1:1:1 computer randomization was used to equally assign subjects into 1 control group using KBA and 2 experimental groups, respectively, using PBA and GBA to record 3 meals. To ensure that group assignment was random, we used an SAS (SAS Institute Inc) randomization procedure to generate a randomized scheduling program [32]. Sample size determination performance was assessed in terms of accuracy and task completion time (in seconds). The required sample size of each group was determined based on previous similar studies [7]. A minimal sample size of 40 participants was determined for each group.

Evaluation Outcomes

In total, 3 outcome types were assessed to evaluate the effectiveness of using mobile apps for dietary measurements, including accuracy in terms of the absolute difference between the actual food item weight and system-assessed weight, the respondent's task duration, and the participant's perception of efficacy.

Accuracy

The first outcome (absolute weight difference) was expressed as a percentage of the difference divided by the actual weight of the food item. The various food items (eg, hash brown, ham, or hot dog for breakfast and chicken leg or pork chop for lunch) were measured before serving. For soft drinks, each cup was filled to a specified level before serving, and the actual weight was determined by the cup size. Ingredients for spaghetti were preweighed by the chef before cooking. Staple foods, vegetables, and dishes with mixed food ingredients were served at a predetermined weight.

Task Duration

The assessment duration for each participant was automatically collected through the mobile app. For KBA, the operating duration covers the time from when the participant first begins to input a standard food serving quantity or food weights until the participant taps the *complete* button on the screen. For PBA, the duration covers the time from when the participant tapped the feature button until the participant tapped the *complete* button. For GBA, the duration covers the time from when the participant tapped the measurement button until the participant tapped the *complete* button.

Perception

Participants' perceptions of the utility of each app were measured using the System Usability Scale (SUS) [33], with 10 items scored using a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Following the study by Sauro and Lewis [33], a mean SUS score above 64.7 was considered above average. One open-ended question was included to collect each user's usage experience and suggestions for design improvements.

Assessment Procedures

The experiment was performed by 2 research assistants. The assessment was scheduled by appointment and implemented on an individual basis. Informed consent was explained to and obtained from each participant. All participants used a 4.7-inch Android smartphone for the test, and all participant trials were conducted on a single day. Each participant first watched an instructional video explaining the operation of the mobile app and the measurement method each participant was assigned to use. Following the written and video instructions, the researchers spent several minutes teaching each participant how to navigate to ensure familiarity with app operation and features and conducted a *dry run*, which involved assessing portion sizes of 4 real food items (bacon, black tea, sweet pepper pork strip, and sausage and spaghetti with cream sauce) and 1 food item (tofu with carrots), which was presented in terms of text and portion description. Participants were encouraged to use the system to assess these items until they felt comfortable with the app operations. For the trial, participants were informed that their time to completion was also a performance consideration. Respondents were asked to record their food items before eating. All participants were asked to record each meal (breakfast, lunch, and dinner) at the time it was served, regardless of whether or how much they actually consumed. Examples of the 3 meals are shown in Multimedia Appendix 4.

Analysis

We conducted a chi-square test for participant gender and background characteristics and applied analysis of variance for the participants' age and BMI. Using the difference in weight (measured in grams) and response duration (in seconds) as continuous variables, we applied the Kruskal-Wallis test to the 3 groups and the Mann-Whitney *U* test for multiple two-group comparisons based on the intention-to-treat principle. We also conducted a chi-square test to assess the SUS questionnaire to compare the 3 groups and an independent *t* test for the multiple two-group comparisons. All statistical tests were two tailed, and a *P* value below .05 indicated statistical significance. All statistical analyses were performed using SAS, version 9.1.3 (SAS Institute).

Written responses to the open-ended question were analyzed by 2 research assistants to check the meanings of each participant's response from different perspectives. Discrepancies between the 2 reviewers were discussed, and a consensus was reached under the project leader's supervision. Items related to usability and design improvement were highlighted and grouped into specific classes. Individual responses in each category were extracted and counted to obtain a cumulative number for each item.

Results

Participant Characteristics

A total of 158 subjects were registered, 135 subjects were scheduled to participate in the experiment, and eventually, 124 subjects completed the study (Figure 3); 11 subjects either failed to record all meals or voluntarily withdrew before completion. The PBA and GBA groups each had 41 respondents, and the KBA group had 42 respondents (see Table 1). Of the respondents who completed the test, 23.4% (29/124) were male; 40.3% (50/124) were aged 19 to 20 years, whereas the remainder were aged 21 to 26 years. The mean BMI of all participants was 21.68 kg/m^2, with an SD of 3.02. Nearly half (61/124, 49.2%) of the subjects were health care management students, followed by students from miscellaneous departments (24/124, 19.4%), industrial design (16/124, 12.9%), nursing (15/124, 12.1%), and industrial business management (8/124, 6.5%). In terms of previous relevant experience, 95.2% (118/124) of respondents reported having cooking experience, followed by 92.7% (115/124) of respondents who reported using Android operating system phones, 65.3% (81/124) of respondents who reported having general health education, 51.6% (64/124) of respondents who reported taking nutrition-related courses, and 20.2% (25/124) of respondents who reported using diet and nutrition apps. The baseline information distributions did not reveal significant differences among the 3 groups, thereby confirming randomized allocation.

Weight Comparison Errors

Table 2 summarizes the weight estimation error for all food items among the 3 measurement methods. The results were described in the order of breakfast, lunch, dinner, and beverages, as shown in Table 2.

Table 2. Weight comparison error among the key-in based aid, photo-based aid, and gesture-based aid (absolute value).

Meal course and food ingredient	Estimating error in weight (%), mean (SD)			Overall	Key-in vs photo	Photo vs gesture	Key-in vs gesture
	Key-in (n=42)	Photo (n=41)	Gesture (n=41)	*P* value	*P* value	*P* value	*P* value
Breakfast (plate)							
Hash browns	18.9 (22.9)	24.6 (17.5)	39.4 (37.1)	.004	.03	.13	.002
Ham	25.2 (12.9)	33.2 (14.6)	106.6 (315.2)	.05	.001	.84	.09
Hot dog	53.4 (22.8)	22.9 (8.7)	47.9 (67.9)	<.001	<.001	.008	.003
Lunch (bento cuboid)							
Staple foods							
Rice	39.7 (31.9)	28.0 (14.3)	48.3 (46.6)	.32	.57	.12	.33
Chow mein	11.9 (19.6)	33.3 (44.2)	67.3 (91.4)	.07	.40	.16	.03
Main courses							
Chicken leg	37.5 (14.4)	32.9 (12.8)	80.3 (83.9)	.29	.12	.25	.77
Pork chop	37.4 (9.3)	24.2 (13.8)	78.0 (45.9)	.001	.08	.001	.06
Vegetables							
Cabbage	79.4 (74.4)	49.5 (53.3)	43.5 (36.8)	.09	.09	.63	.05
White shoots	77.2 (69.4)	28.0 (6.6)	35.0 (21.4)	.03	.009	.59	.06
Loofah	49.6 (31.9)	29.2 (20.1)	51.4 (51.0)	.10	.02	.21	.64
Green beans	55.3 (88.8)	22.9 (8.2)	32.4 (25.6)	.84	.84	.79	.59
Green pepper	51.8 (50.9)	16.7 (16.0)	44.1 (30.2)	.12	.17	.04	.77
Dish with 2 ingredients							
Green pepper	238.5 (237.6)	38.5 (18.5)	80.6 (82.5)	.01	.002	.62	.05
Shredded pork	412.8 (523.3)	7.7 (27.7)	107.0 (124.0)	<.001	<.001	<.001	.26
Dish with 2 ingredients							
Tomato	152.5 (293.2)	68.5 (48.7)	76.7 (59.5)	.72	.56	.51	.55
Scrambled eggs	129.4 (105.8)	45.2 (31.7)	90.2 (68.8)	.008	.02	.001	.45
Dish with 3 ingredients							
Cabbage	106.6 (95.4)	25.9 (19.8)	36.8 (28.5)	<.001	<.001	.13	.008
Bacon	128.6 (218.0)	3.7 (19.2)	155.7 (159.8)	<.001	<.001	<.001	.07
Black fungus	332.8 (397.5)	22.2 (42.4)	78.4 (82.7)	<.001	<.001	<.001	.10
Dish with 3 ingredients							
Fried bean curd	41.9 (40.8)	64.3 (63.3)	210.5 (214.3)	.009	.96	.02	.002
Green pepper	58.3 (63.4)	26.2 (29.8)	55.0 (62.7)	.18	.13	.11	.64
Shredded carrot	99.2 (156.3)	31.0 (8.9)	57.3 (30.6)	.03	.13	.003	.60
Dinner (cylindrical bowl)							
Dish with 2 ingredients							
Bacon	60.6 (19.5)	76.5 (8.5)	75.7 (78.0)	.003	<.001	.08	.65
Spaghetti	41.6 (31.1)	29.6 (25.3)	39.6 (37.7)	.74	.52	.69	.55
Dish with 2 ingredients							
German sausage	36.1 (31.0)	37.8 (25.9)	192.6 (141.6)	<.001	.65	<.001	<.001
Spaghetti	59.2 (68.4)	28.5 (25.3)	53.2 (44.8)	.21	.24	.06	.96
Beverages[a] (conical cup)							

Meal course and food ingredient	Estimating error in weight (%), mean (SD)			Overall	Key-in vs photo	Photo vs gesture	Key-in vs gesture
	Key-in (n=42)	Photo (n=41)	Gesture (n=41)	*P* value	*P* value	*P* value	*P* value
Orange juice (390 mL)	22.3 (9.7)	38.1 (22.6)	20.3 (10.5)	.04	.06	.03	.32
Black tea (390 mL)	14.1 (9.6)	51.3 (15.7)	19.4 (10.7)	.001	.003	.002	.29
Soy milk (390 mL)	31.5 (21.0)	48.7 (15.1)	24.6 (17.9)	<.001	.006	<.001	.07
Orange juice (490 mL)	9.7 (13.2)	21.0 (14.1)	23.2 (9.1)	<.001	<.001	.19	<.001
Black tea (490 mL)	10.0 (13.8)	19.3 (10.2)	26.2 (14.2)	.002	.02	.04	.002
Green tea (500 mL)	19.6 (18.9)	16.0 (9.1)	22.4 (11.7)	.25	.99	.08	.32
Green tea (660 mL)	7.5 (8.1)	8.9 (21.2)	20.4 (12.2)	<.001	.02	<.001	.009

[a]Beverages include all the beverages for breakfast, lunch, and dinner.

Breakfast

Looking at the overall *P* value, all the 3 food items for the breakfast courses were found to incur a significant difference in terms of weight estimation errors. PBA users performed best in the hot dog group (*P*<.001), whereas KBA outperformed PBA or GBA in measuring the hash brown and ham portion sizes.

Lunch

All 19 food ingredients present in the food items chosen by the participants were analyzed. As for the staple foods, no significant differences were found. The pork chop in the lunch main course was significantly different (*P*<.001), with PBA outperforming GBA. In the vegetable course, only 1 of the 5 ingredients (ie, white shoots) showed a significant difference (*P*=.03), with PBA outperforming KBA. The other 4 vegetables (ie, cabbage, loofah, green beans, and green pepper) showed no significant differences.

Two lunch items included 2 ingredients each, and analysis of these 4 ingredients found significant differences for 3 (ie, green pepper, shredded pork, and scrambled eggs). Furthermore, another 2 lunch items included 3 ingredients each, and analysis of these 6 ingredients found significant differences for 5 ingredients: cabbage, bacon, black fungus, fried bean curd, and shredded carrot. Of these 5 ingredients, PBA outperformed KBA for cabbage, bacon, and black fungus. PBA also outperformed GBA for all ingredients except cabbage.

Dinner

For the dinner food items, each of the 2 spaghetti dishes had 2 ingredients. In total, 4 ingredients were described. The measurement errors for bacon and German sausage showed significant differences (*P*<.01). For bacon, both KBA and GBA outperformed PBA, whereas KBA and PBA outperformed GBA in measuring the German sausage in the two-group comparison.

Beverages

Six beverages showed significant differences (*P*<.05), with the exception of the 500-mL beverage (ie, green tea). KBA significantly outperformed the other 2 aids for the two 490-mL beverages (ie, orange juice and black tea). PBA provided the best results for the 660-mL beverages. For the three 390-mL beverages, GBA significantly outperformed PBA, whereas KBA and GBA produced similar results.

Task Duration

Table 3 summarizes the response time required to estimate the food portion sizes for all food items.

Table 3. Portion assessment duration (absolute value).

Meal course and food ingredient	Assessment duration (seconds), mean (SD)			Overall	Key-in vs photo	Photo vs gesture	Key-in vs gesture
	Key-in (n=42)	Photo (n=41)	Gesture (n=41)	*P* value	*P* value	*P* value	*P* value
Breakfast							
Hash browns	8.8 (5.8)	8.4 (2.6)	90.5 (39.6)	<.001	.35	<.001	<.001
Ham	10.9 (7.5)	11.8 (7.5)	83.6 (40.1)	<.001	.33	<.001	<.001
Hot dog	15.4 (9.7)	10.3 (4.0)	86.4 (41.2)	<.001	.001	<.001	<.001
Lunch							
Staple food							
Rice	9.8 (7.6)	12.1 (6.1)	78.6 (33.3)	<.001	.04	<.001	<.001
Chow mein	7.2 (3.1)	8.0 (5.8)	86.0 (28.5)	.005	.95	.003	.008
Main course							
Chicken leg	9.6 (6.4)	9.5 (4.8)	95.5 (34.5)	<.001	.48	<.001	<.001
Porkchop	9.0 (7.7)	15.3 (4.7)	97.2 (39.1)	<.001	.02	<.001	<.001
Vegetables							
Cabbage	12.6 (8.1)	11.2 (6.7)	71.9 (29.2)	<.001	.67	<.001	<.001
White shoots	14.3 (10.0)	8.7 (4.7)	72.6 (32.6)	<.001	.04	<.001	<.001
Loofah	17.4 (18.1)	9.6 (8.2)	70.1 (33.4)	<.001	.008	<.001	<.001
Green beans	9.5 (4.6)	9.9 (7.0)	53.2 (16.0)	<.001	.88	.002	<.001
Green pepper	11.1 (2.9)	11.1 (2.1)	59.7 (30.5)	<.001	>.99	.002	.002
Dishes with 2 ingredients							
Green pepper+shredded pork	15.8 (8.8)	12.5 (6.3)	70.6 (21.5)	<.001	.31	<.001	<.001
Tomato+scrambled eggs	22.4 (19.0)	15.0 (7.6)	71.3 (29.6)	<.001	.14	<.001	<.001
Dishes with 3 ingredients							
Cabbage+bacon+black fungus	23.8 (20.1)	15.3 (7.3)	91.8 (40.3)	<.001	.18	<.001	<.001
Fried bean curd+green pepper+shredded carrot	33.2 (13.7)	22.6 (12.2)	84.3 (36.4)	<.001	.03	<.001	<.001
Dinner							
Bacon spaghetti	34.1 (15.9)	25.3 (27.1)	97.1 (47.5)	<.001	.002	<.001	<.001
German sausage spaghetti	25.5 (11.5)	24.4 (26.7)	91.9 (53.0)	<.001	.09	<.001	<.001
Beverages							
Orange juice (390 mL)	13.0 (9.6)	15.1 (7.1)	93.6 (29.1)	<.001	.13	<.001	<.001
Black tea (390 mL)	13.8 (5.9)	15.5 (5.4)	96.2 (29.1)	<.001	.65	<.001	.001
Soy milk (390 mL)	13.9 (9.2)	12.0 (5.6)	94.2 (46.4)	<.001	.99	<.001	<.001
Orange juice (490 mL)	11.5 (7.6)	10.3 (3.3)	78.3 (29.4)	<.001	.83	<.001	<.001
Black tea (490 mL)	9.3 (5.1)	12.5 (5.6)	84.2 (35.6)	<.001	.05	<.001	<.001
Green tea (500 mL)	13.4 (5.2)	12.6 (4.2)	90.2 (42.0)	<.001	.57	<.001	<.001
Green tea (660 mL)	13.8 (8.6)	10.6 (3.5)	93.4 (38.7)	<.001	.28	<.001	<.001

Breakfast

All the food dishes (3 for breakfast, 4 for lunch, and 2 for dinner) showed significant differences. In the two-group comparison, KBA and PBA took significantly less time than GBA for all the ingredients, that is, GBA performed worst.

For the breakfast courses, PBA performed best for the hot dog. KBA and PBA performed similarly for the other 2 courses.

Lunch

Of the lunch staple food and main courses, KBA performed best for the rice and the pork chop. Of the 5 vegetables, PBA outperformed KBA for the white shoots and loofah. Of the 4 lunch dishes incorporating 2 or 3 ingredients, PBA outperformed KBA for the dish that included fried bean curd, green pepper, and shredded carrot.

Dinner

Of the dinner courses, PBA also outperformed KBA for bacon spaghetti.

Beverages

All 7 beverage drinks were significantly different. KBA performed best in 1 of the 7 beverages, specifically 490-mL black tea. KBA and PBA performed similarly for the other 6 beverages. GBA performed worst.

Participants' Subjective Responses Using System Usability Scale

Table 4 summarizes the response scores for the 3 aids, showing no significant difference. However, KBA significantly outperformed the other 2 aids in terms of usability score (72.8; *P*=.008). In the two-group comparison, KBA significantly outperformed GBA (*P*=.004). In terms of learnability score, no significant difference was found among the 3 aids.

Table 4. System Usability Scale score.

Score[a,b,c]	KBA[d] (n=42), mean (SD)	PBA[e] (n=41), mean (SD)	GBA[f] (n=41), mean (SD)	*P* value			
				Overall	KBA vs PBA	PBA vs GBA	KBA vs GBA
Overall score	69.6 (11.6)	67.0 (9.1)	64.0 (12.6)	.08	.25	.23	.04
Usability score	72.8 (11.9)	69.4 (9.5)	64.8 (13.1)	.008	.15	.07	.004
Learnability score	56.5 (21.4)	57.0 (22.7)	61.0 (23.9)	.62	.92	.44	.38

[a]Questionnaires were presented in Chinese.

[b]Mean score for the commercial apps was 64.7.

[c]The questionnaire's Cronbach alpha of .71 indicated good internal consistency and reliability.

[d]KBA: key-in–based aid.

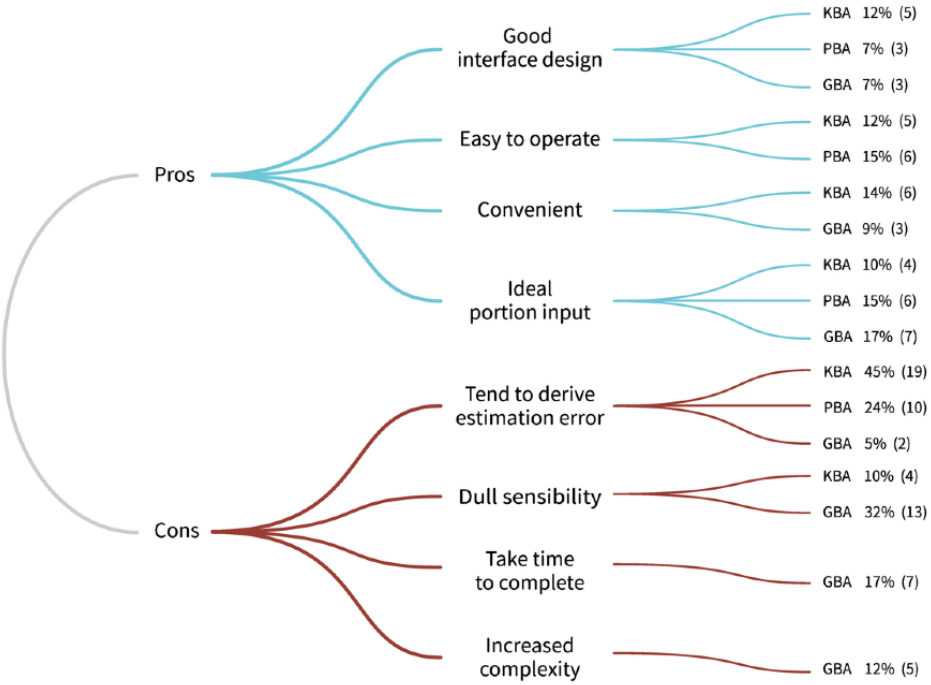
[e]PBA: photo-based aid.

[f]GBA: gesture-based aid.

Open-Ended Questions Response

Figure 6 shows the themes identified from the participant evaluations. Participants in all 3 groups cited advantages, including *good interface design* by KBA (5/42, 12%), PBA (3/41, 7%), and GBA (3/41, 7%) and *ideal portion input* by KBA (4/42, 10%), PBA (6/41, 15%), and GBA (7/41, 17%). KBA (5/42, 12%) and PBA (6/41, 15%) were said to be *easy to operate*, whereas KBA (6/42, 14%) and GBA (3/41, 9%) were *convenient*. In terms of drawbacks, KBA (19/42, 45%) participants were most likely to cite *estimation error*, followed by PBA (10/41, 24%) and GBA (2/41, 5%). GBA (7/41, 17%) users cited *takes time to complete*. All 3 aids were cited for being slow to complete the task, whereas GBA (5/41, 12%) alone was cited for *greater complexity*.

Figure 6. Subjective evaluation. KBA: key-in–based aid; PBA: photo-based aid; GBA: gesture-based aid.



Discussion

Principal Findings

This research investigated user experiences of the 3 aids by 4 metrics: weight error, operation duration, SUS, and open-ended questions. Using a set group of food and beverage items, this research illustrates the capabilities and limitations of each measurement method for young adults under authentic usage conditions. Our evidence-based findings provide a reference for real-time dietary intake reporting apps. The implications of the findings are discussed, along with the suggestions for further system improvement in the mHealth domain.

Analysis of Key-in–Based Aid

Similar to the study by Howes et al [12], foods with amorphous shapes showed the highest percent error. Participants in the KBA group produced widely varying results (see Table 2) potentially because KBA relies on prior knowledge when inputting the food serving amount using serving information for input in terms of unit of volume. One potential reason is that authentic food and beverage servings closer to a standard serving size facilitate more accurate guessing and thus result in a higher measurement accuracy. For example, the realistic weight of a chow mein serving is 104 g, which is close to that of a standard bowl (ie, 100 g); thus, participants could simply input *one* in the bowl field to produce an accurate measurement. However, the serving size of the hot dog was around half the standard serving size, thus simply inputting *one hot dog* is likely to result in a high weight error. KBA also produced relatively high degrees of error for dishes with multiple food ingredients (eg, shredded pork, scrambled eggs, bacon, cabbage, and black fungus). That 45% of the group participants considered KBA to *tend to derive estimation errors* reflects this variation.

KBA also tended to outperform the other 2 approaches for standard size beverage containers (eg, 490 mL and 500 mL). However, the participants seemed to be less familiar with the 390-mL container, resulting in a relatively lower accuracy for 390-mL servings of orange juice and soy milk. In addition, the error rate might be affected by the density of the food item using unit of weight for input. For example (see Table 2), soy milk is considerably more dense than water and thus produced a significantly greater error rate (34.9%) than the same volume of black tea (10.3%). A relatively high number of KBA users (see Multimedia Appendix 5) inputted portion weight measurements within a 10% error range for the 490-, 500-, and 660-mL beverages (>53%); hash browns (42%); and chow mein (83%), but this level of accuracy did not extend to all food items. This would address the widely varying results among participants.

Analysis of Photo-Based Aid

PBA was found to produce relatively higher accuracy rates, although the presentation of the food images differed significantly from the actual food item (Table 2). PBA also performed relatively well for ingredients in the mixed food dishes (eg, shredded pork, bacon, German sausage, cabbage, and black fungus). The research assistant observed that some participants attempted to *dissect* the dish and count the number of chopped vegetables as the basis for choosing an input image. This is an interesting observation worthy of further investigation.

PBA outperformed KBA and GBA for hot dogs potentially because one of its selection items was close to the serving size used in the test. However, selecting the correct representative image does not necessarily result in accurate weight estimation. High weight variety of served food items would raise a substantial challenge for PBA participants. Authentic hash browns, for example, range in weight from 33 g to 64 g, presenting a significant range of error from the 60 g hash brown

in the image. This weight variability caused PBA to underperform KBA for both hash browns and ham. The range of authentic foods used in the test would affect the accuracy of PBA. PBA was considerably less able to differentiate between various beverage container sizes possibly because the reference images showed a 660-mL cup filled to different heights, which differed significantly from the different volume beverage containers used in the actual meals. The image for the 660-mL green tea item, however, was identical to the authentic test item and thus elicited a much higher accuracy rate. PBA tended to outperform for those foods that appeared similar to one of the listed images. Similar results were found in our previous research [29]. Training was suggested to further improve PBA accuracy. Additional training in quantification accuracy from digital images would be necessary [12]. In a study by Lee et al [34], images of adolescents eating were first photographed and reported after 14 hours with the support of 2 portion size estimation aids (ie, 2D images [similar to our PBA] and multiple measurement descriptors [similar to our KBA]). Among the tested foods, sausage links were found to have the highest accuracy for both aids. This is similar to our findings for KBA and PBA. However, Lee et al [34] estimated the portion size after 14 hours of food consumption, and the target group (age ranging from 11 to 18 years) differed from that examined here. Gibson et al [23] tested a measurement aid using fist, thumb, or fingertips in comparison with household measures (eg, cups). Using university staff and students as participants, they found that for foods that closely resemble a comparative reference, for example, finger tips would perform better in weight estimation. This is similar to our findings, in that PBA tends to outperform for foods that closely resemble one of the preselected images.

Analysis of Gesture-Based Aid

GBA requires users to roughly estimate the area and thickness of a food item and then use swiping gestures to describe the approximate food size on the mobile device screen using an accompanying object as a size reference. GBA generally produced relatively inaccurate measurement results. However, the mean estimation error for GBA (Table 2) was relatively consistent for beverages (19.8%-33.2%), vegetables (28.0%-44.7%), and breakfast food items (30.8%-41.5%). GBA also showed a relatively high degree of weight error likely because of the need for the users to determine the length relationship between the food item and the reference object (eg, credit card). GBA errors were also likely because of the need to calculate weight based on food volume, thus requiring users to make 2 independent estimates that compound potential errors. Although GBA suffered from the time spent to complete the measurement, it was considered to be *accurate* based on the subjective opinion. In addition, GBA performed relatively better in some food items when looking at weight accuracy. GBA would be used for some situations when the user was unfamiliar with the food density or the volume size.

Response Time Comparison of Key-in–Based Aid, Photo-Based Aid, and Gesture-Based Aid

In terms of time efficiency, GBA underperformed compared with KBA and PBA, with respective response times in the ranges of 49 to 95, 7 to 13, and 5 to 16.5 seconds. In addition, no correlation was found between the accuracy rate and the response duration for any of the 3 aids. GBA suffered from longer operational time to estimate the area and height of the food. Apparently, GBA required more time for completion. Similar results were found in our previous research [29], in that GBA-like aid performed significantly worse than PBA in terms of response time.

Participants' Perception

Overall SUS scores for KBA and PBA exceeded the average score of 64.80 (Table 4). GBA was marginally close to the average score. Looking at participants' open-ended responses (see Figure 6), participants in each group considered the used aid to be *ideal portion input*, *easy to operate*, and *convenient*. This was consistent with the overall SUS score. User responses to the open-ended question characterized GBA as *takes time to complete* and *more complex*. This was consistent with the quantitative time efficiency evaluation results in Table 3. Furthermore, GBA was considered to be an *ideal portion input* by 17% of respondents. Only 5% of subjects considered the aid to be *prone to estimation error*, as opposed to 45% for KBA. Participants reported concerns of *no timely response*, *slow to use*, and *more complex* for GBA, indicating a need to further improve its user interface. User perceptions of KBA as tending toward increased estimation error conflicted with the testing result shown in Table 2. However, in terms of accuracy and time duration results, KBA performed relatively better for food items that significantly differed. This apparent contradiction should be investigated in future work. Similarly, König et al [35] also showed that participants' perceived accuracy did not match their actual accuracy, raising the need for further education in this type of misperception.

Further Improvements

The design improvement for KBA is a user-friendly design that represents and relates actual food items with the appropriate unit of weight or volume to ease users' concerns of *estimation error*. Further design improvements in PBA would require providing a selection mechanism to reflect the variations in food shapes and containers. For example, images could be provided to allow users to select various types of containers (eg, cup or bottle), followed by volume (eg, 500 mL or 700 mL) and fullness level (eg, half or one third). GBA is an innovative approach that requires active user input to compare the relative food size and describe food volume. Further design improvements in GBA are needed, such as using multiple finger gestures [31] to improve time efficiency and allow users to select rulers (eg, a 20-mm ruler) rather than relying on a credit card as the point of reference. Future work would also need to resolve technology-related issues that present challenges for food reporting. All three methods inherit certain limitations, for example, visual-based attributes, which do not account for variation in food density. Future improvements would require incorporating nonvisual-based methods or technologies to resolve this limitation.

Study Limitations

We allowed the KBA group to key in either units of weight (eg, grams) or volume (eg, pieces). These 2 key-in methods could be tested separately in the future to avoid potentially confounding variables. Further tests are required using different target populations (eg, senior citizens or patients with chronic illnesses) whose results may differ from those of young adults. A wider range of authentic foods and longer testing periods could also be included. Another limitation is that the experiment did not consider the possibility that users may not finish certain dishes. To better reflect the realistic eating situation, future research would need to consider the issue of leftovers, for example, to improve the app's functionality to include the recording of the leftovers.

Conclusions

Experimental results of young adults using 3 prominent aids showed various strengths and weaknesses. KBA was more accurate for common drink containers or food items that approximate standard serving sizes, whereas PBA performed better for irregular shapes, which closely resemble one of the preselected images, and GBA was better suited for unfamiliar containers or dishes but requires additional design improvement to enhance time efficiency. Thus, to optimize performance, different approaches should be used for different conditions. Participants were queried regarding their subjective impressions of the pros and cons of portion size measurement methods. Concerns were raised regarding estimation errors, device sensitivity, and task complexity. However, participants' subjective impressions suggest an unfamiliarity with the distinct portion size estimation capability of each aid, and future designs should take this into account. To deal with the complexity of timely collection of dietary data, more work is needed to generate evidence on the appropriateness of each method under specific eating contexts. The user experience results provide scientific evidence for the continued development of related dietary recording apps. Future work could include improvements to design and functionality and the development of new design innovations to enhance effectiveness and convenience. Further studies involving different evaluation metrics are suggested, such as investigating mental loading during dietary recording or considering broader target groups such as low-literacy populations or different eating environments and meal sets.

Acknowledgments

The authors are grateful for the efforts and cooperation of all study participants at the Food Interaction Design Lab and the Health Informatics Lab of the Chang Gung University in the College of Management. The authors are also grateful to Sherry Yueh-Hsia Chiu (PhD) and Jou-Yin Yeh at the Chang Gung University for statistical support. This work was supported by the Research Fund of Chang Gung Memorial Hospital and Chang Gung University (BMRPD67) and the Ministry of Science and Technology, Taiwan (MOST-106-2221-E-182-042-MY2).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile app based on key-in–based aid.
[MP4 File (MP4 Video), 1391 KB - mhealth_v8i4e14543_app1.mp4]

Multimedia Appendix 2
Mobile app based on photo-based aid.
[MP4 File (MP4 Video), 809 KB - mhealth_v8i4e14543_app2.mp4]

Multimedia Appendix 3
Mobile app based on gesture-based aid.
[MP4 File (MP4 Video), 964 KB - mhealth_v8i4e14543_app3.mp4]

Multimedia Appendix 4
Experimental food samples.
[DOCX File , 754 KB - mhealth_v8i4e14543_app4.docx]

Multimedia Appendix 5
Weight error rates within 10% among the 3 aids.
[DOCX File , 18 KB - mhealth_v8i4e14543_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 31500 KB - mhealth_v8i4e14543_app6.pdf]

References

1. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. Lancet 2010 Nov 20;376(9754):1775-1784. [doi: 10.1016/S0140-6736(10)61514-0] [Medline: 21074255]
2. Allman-Farinelli M. Nutrition promotion to prevent obesity in young adults. Healthcare (Basel) 2015 Sep 7;3(3):809-821 [FREE Full text] [doi: 10.3390/healthcare3030809] [Medline: 27417798]
3. Munt AE, Partridge SR, Allman-Farinelli M. The barriers and enablers of healthy eating among young adults: a missing piece of the obesity puzzle: A scoping review. Obes Rev 2017 Jan;18(1):1-17. [doi: 10.1111/obr.12472] [Medline: 27764897]
4. Simons D, de Bourdeaudhuij I, Clarys P, de Cocker K, Vandelanotte C, Deforche B. A smartphone app to promote an active lifestyle in lower-educated working young adults: development, usability, acceptability, and feasibility study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e44 [FREE Full text] [doi: 10.2196/mhealth.8287] [Medline: 29463491]
5. Smith A. Pew Research Center. 2015 Apr 1. US Smartphone Use in 2015 URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2019-09-01]
6. 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]
7. Podina IR, Fodor LA, Cosmoiu A, Boian R. An evidence-based gamified mHealth intervention for overweight young adults with maladaptive eating habits: study protocol for a randomized controlled trial. Trials 2017 Dec 12;18(1):592 [FREE Full text] [doi: 10.1186/s13063-017-2340-6] [Medline: 29233169]
8. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
9. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
10. Schumer H, Amadi C, Joshi A. Evaluating the dietary and nutritional apps in the Google Play Store. Healthc Inform Res 2018 Jan;24(1):38-45 [FREE Full text] [doi: 10.4258/hir.2018.24.1.38] [Medline: 29503751]
11. Almiron-Roig E, Navas-Carretero S, Emery P, Martínez JA. Research into food portion size: methodological aspects and applications. Food Funct 2018 Feb 21;9(2):715-739. [doi: 10.1039/c7fo01430a] [Medline: 29219156]
12. Howes E, Boushey C, Kerr D, Tomayko E, Cluskey M. Image-based dietary assessment ability of dietetics students and interns. Nutrients 2017 Feb 7;9(2):pii: E114 [FREE Full text] [doi: 10.3390/nu9020114] [Medline: 28178196]
13. 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]
14. Rolls BJ, Morris EL, Roe LS. Portion size of food affects energy intake in normal-weight and overweight men and women. Am J Clin Nutr 2002 Dec;76(6):1207-1213. [doi: 10.1093/ajcn/76.6.1207] [Medline: 12450884]
15. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
16. National Research Council, Division on Earth and Life Studies, Commission on Life Sciences, Food and Nutrition Board, Coordinating Committee on Evaluation of Food Consumption Surveys, Subcommittee on Criteria for Dietary Evaluation. Nutrient Adequacy: Assessment Using Food Consumption Surveys. Washington DC: National Academies Press; 1986.
17. Almiron-Roig E, Solis-Trapala I, Dodd J, Jebb SA. Estimating food portions. Influence of unit number, meal type and energy density. Appetite 2013 Dec;71:95-103 [FREE Full text] [doi: 10.1016/j.appet.2013.07.012] [Medline: 23932948]
18. Waltner G, Schwarz M, Ladstatter S, Weber A, Luley P, Bischof H, et al. MANGO - Mobile Augmented Reality with Functional Eating Guidance and Food Awareness. In: Proceedings of the International Conference on Image Analysis and Processing. 2015 Presented at: ICIAP'15; September 7-11, 2015; Genoa, Italy p. 425-432.
19. Rollo ME, Bucher T, Smith SP, Collins CE. ServAR: An augmented reality tool to guide the serving of food. Int J Behav Nutr Phys Act 2017 May 12;14(1):65 [FREE Full text] [doi: 10.1186/s12966-017-0516-9] [Medline: 28499433]
20. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013 Jun 24;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]
21. Inunganbi S, Seal A, Khanna P. Classification of Food Images through Interactive Image Segmentation. In: Proceedings of the Asian Conference on Intelligent Information and Database Systems. 2018 Presented at: ACIIDS'18; March 19-21, 2018; Dong Hoi City, Vietnam p. 519-528.
22. Robinson E, Oldham M, Cuckson I, Brunstrom JM, Rogers PJ, Hardman CA. Visual exposure to large and small portion sizes and perceptions of portion size normality: Three experimental studies. Appetite 2016 Mar 1;98:28-34 [FREE Full text] [doi: 10.1016/j.appet.2015.12.010] [Medline: 26702602]
23. Gibson AA, Hsu MS, Rangan AM, Seimon RV, Lee CM, Das A, et al. Accuracy of hands v household measures as portion size estimation aids. J Nutr Sci 2016;5:e29 [FREE Full text] [doi: 10.1017/jns.2016.22] [Medline: 27547392]

24. Ball GD, Friedman A. Dice, golf balls, and CDs: assumptions about portion size measurement aids. Can J Diet Pract Res 2010;71(3):146-149. [doi: 10.3148/71.3.2010.146] [Medline: 20825698]
25. Pettitt C, Liu J, Kwasnicki RM, Yang G, Preston T, Frost G. A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. Br J Nutr 2016 Jan 14;115(1):160-167. [doi: 10.1017/S0007114515004262] [Medline: 26537614]
26. Pendergast FJ, Ridgers ND, Worsley A, McNaughton SA. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act 2017 Mar 14;14(1):30 [FREE Full text] [doi: 10.1186/s12966-017-0488-9] [Medline: 28288657]
27. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
28. Liu YC, Chen CH, Tsou YC, Lin YS, Chen HY, Yeh JY, et al. Evaluating mobile health apps for customized dietary recording for young adults and seniors: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e10931 [FREE Full text] [doi: 10.2196/10931] [Medline: 30767906]
29. Liu Y, Chen C, Lee C, Lin Y, Chen H, Yeh J, et al. Design and usability evaluation of user-centered and visual-based aids for dietary food measurement on mobile devices in a randomized controlled trial. J Biomed Inform 2016 Dec;64:122-130 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.001] [Medline: 27720984]
30. Nelson M, Atkinson M, Darbyshire S. Food photography. I: The perception of food portion size from photographs. Br J Nutr 1994 Nov;72(5):649-663. [doi: 10.1079/bjn19940069] [Medline: 7826990]
31. Asakawa DS, Dennerlein JT, Jindrich DL. Index finger and thumb kinematics and performance measurements for common touchscreen gestures. Appl Ergon 2017 Jan;58:176-181. [doi: 10.1016/j.apergo.2016.06.004] [Medline: 27633211]
32. Deng C, Graz J. Generating Randomization Schedules Using SAS Programming. In: Proceedings of the 27th Annual SAS Users Group International Conference. 2002 Presented at: SUGI'27; April 14-17, 2002; Orlando, Florida.
33. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. USA: Morgan Kaufmann; 2016.
34. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol 2012 Mar 1;6(2):428-434 [FREE Full text] [doi: 10.1177/193229681200600231] [Medline: 22538157]
35. König LM, Ziesemer K, Renner B. Quantifying actual and perceived inaccuracy when estimating the sugar, energy content and portion size of foods. Nutrients 2019 Oct 11;11(10):pii: E2425 [FREE Full text] [doi: 10.3390/nu11102425] [Medline: 31614559]

Abbreviations

GBA: gesture-based aid
KBA: key-in–based aid
mHealth: mobile health
PBA: photo-based aid
SUS: System Usability Scale

Edited by G Eysenbach; submitted 08.06.19; peer-reviewed by S Chokshi, G Temiz, I Podina, L König; comments to author 14.08.19; revised version received 18.10.19; accepted 22.01.20; published 29.04.20.

Please cite as:
Liu YC, Wu ST, Lin SJ, Chen CH, Lin YS, Chen HY
Usability of Food Size Aids in Mobile Dietary Reporting Apps for Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e14543
URL: http://mhealth.jmir.org/2020/4/e14543/
doi:10.2196/14543
PMID:32347805

Original Paper

Facilitators and Barriers to Chronic Disease Self-Management and Mobile Health Interventions for People Living With Diabetes and Hypertension in Cambodia: Qualitative Study

Lesley Steinman[1*], MPH, MSW; Hen Heang[2*], MS; Maurits van Pelt[2*], LLM, MSc; Nicole Ide[3], MPH; Haixia Cui[2], PhD; Mayuree Rao[3,4*], MD; James LoGerfo[5,6*], MD, MPH; Annette Fitzpatrick[6,7,8*], MA, PhD

[1]Department of Health Services, University of Washington, Seattle, WA, United States
[2]MoPoTsyo Patient Information Centre, Phnom Penh, Cambodia
[3]Division of General Internal Medicine, University of Washington, Seattle, WA, United States
[4]General Medicine Service, VA Puget Sound Health Care System, Seattle, WA, United States
[5]Department of Medicine, University of Washington, Seattle, WA, United States
[6]Department of Global Health, University of Washington, Seattle, WA, United States
[7]Departments of Family Medicine, University of Washington, Seattle, WA, United States
[8]Department of Epidemiology, University of Washington, Seattle, WA, United States
[*]these authors contributed equally

Corresponding Author:
Lesley Steinman, MPH, MSW
Department of Health Services
University of Washington
1107 NE 45th St Suite 400
Seattle, WA, 98105
United States
Phone: 1 2065439837
Email: lesles@uw.edu

Abstract

Background: In many low- and middle-income countries (LMICs), heart disease and stroke are the leading causes of death as cardiovascular risk factors such as diabetes and hypertension rapidly increase. The Cambodian nongovernmental organization, MoPoTsyo, trains local residents with diabetes to be peer educators (PEs) to deliver chronic disease self-management training and medications to 14,000 people with hypertension and/or diabetes in Cambodia. We collaborated with MoPoTsyo to develop a mobile-based messaging intervention (mobile health; mHealth) to link MoPoTsyo's database, PEs, pharmacies, clinics, and people living with diabetes and/or hypertension to improve adherence to evidence-based treatment guidelines.

Objective: This study aimed to understand the facilitators and barriers to chronic disease management and the acceptability, appropriateness, and feasibility of mHealth to support chronic disease management and strengthen community-clinical linkages to existing services.

Methods: We conducted an exploratory qualitative study using semistructured interviews and focus groups with PEs and people living with diabetes and/or hypertension. Interviews were recorded and conducted in Khmer script, transcribed and translated into the English language, and uploaded into Atlas.ti for analysis. We used a thematic analysis to identify key facilitators and barriers to disease management and opportunities for mHealth content and format. The information-motivation-behavioral model was used to guide data collection, analysis, and message development.

Results: We conducted six focus groups (N=59) and 11 interviews in one urban municipality and five rural operating districts from three provinces in October 2016. PE network participants desired mHealth to address barriers to chronic disease management through reminders about medications, laboratory tests and doctor's consultations, education on how to incorporate self-management into their daily lives, and support for obstacles to disease management. Participants preferred mobile-based voice messages to arrive at dinnertime for improved phone access and family support. They desired voice messages over texts to communicate trust and increase accessibility for persons with limited literacy, vision, and smartphone access. PEs shared similar views and perceived

mHealth as acceptable and feasible for supporting their work. We developed 34 educational, supportive, and reminder mHealth messages based on these findings.

Conclusions: These mHealth messages are currently being tested in a cluster randomized controlled trial (#1R21TW010160) to improve diabetes and hypertension control in Cambodia. This study has implications for practice and policies in Cambodia and other LMICs and low-resource US settings that are working to engage PEs and build community-clinical linkages to facilitate chronic disease management.

(JMIR Mhealth Uhealth 2020;8(4):e13536) doi:10.2196/13536

KEYWORDS

diabetes mellitus; hypertension; chronic disease; noncommunicable diseases; health educators; mHealth; qualitative; disease management; developing countries

Introduction

Background

There is an increasing burden of noncommunicable diseases (NCDs) across the world and particularly in low- and middle-income countries (LMICs) as they go through the epidemiological transition from infectious to chronic diseases that accompanies economic development [1]. In 2016, NCDs accounted for 60% of disability-adjusted life years (DALYs) lost, a common measure of disease burden, and for 71% of all deaths [2]. Most deaths from cardiovascular diseases now occur in LMICs [3], and 80% of global deaths in LMICs are from NCDs [4].

Cambodia is experiencing the same epidemiological transition as other LMICs. Life expectancy is on the rise, increasing from an average of 56 and 60 years in 1990 to 66 and 72 years in 2015 for men and women, respectively [5]. The 2016 Global Burden of Disease Study's visualizations show an increasing burden of NCDs, with NCD DALYs increasing by 55% from 1990 to 2016 for countries such as Cambodia with a low-to-middle sociodemographic index [6]. In Cambodia, diabetes, stroke, and heart disease have demonstrated the largest percentage change in years lived with disability since 1990. Ischemic heart disease has been the leading cause of death since 2005, and cerebrovascular disease became the second leading cause of death in 2015 [5]. A large proportion of deaths due to heart disease, stroke, and diabetes are attributable to metabolic risks such as high blood pressure, cholesterol, glucose, and BMI, all of which are modifiable risk factors that can be managed or prevented outside of the health system.

The 2010 World Health Organization's (WHO) Stepwise Approach to Surveillance survey first found a prevalence of diabetes and hypertension of 2.9% and 11.2%, respectively, in the general Cambodian population [7], but this had rapidly increased to 9.6% and 14.2% in 2016, respectively, when that survey was repeated. These prevalence rates are higher in more vulnerable populations such as among Cambodians living with HIV [8]. Although a variety of methods have been developed to assist patients in managing their diabetes or hypertension and to increase access to care, the use of community- and peer-based interventions seems especially promising for reaching underserved populations [9-11]. In a review by Joshi et al [12] on studies involving task shifting for the management of chronic conditions in LMICs, it was found that nonphysician health workers could successfully screen patients and improve patients' control of blood pressure and glucose levels when compared with usual health care. Lay health workers also have more time to provide behavioral counseling services and patient education as clinicians in LMICs often have limited time to spend with patients [13-17].

Mobile health (mHealth) offers a promising approach for managing chronic conditions in LMICs. mHealth is a subset of electronic health and encompasses a wide range of mobile and wireless technologies to help improve health and health care [18]. mHealth typically supports four functions: (1) health promotion and awareness, (2) remote monitoring and care support, (3) disease surveillance and outbreak detection, and (4) decision support system [19,20]. mHealth interventions can provide opportunities for improved chronic disease management in LMICs, given the limited health system infrastructure [21] and high cell phone coverage (90%) [22,23]. In Cambodia, 94% of the people report owning their own phone, with 99% being reachable by phone [24]. mHealth can also support people living with chronic conditions outside of the health care system where they spend most of their time and can benefit from regular self-management and routine long-term monitoring [25-27].

Although it is promising, more evidence is needed for mHealth for the management of chronic conditions in LMICs. Most mHealth evidence to date is on communicable diseases (eg, tuberculosis) and on maternal and child health [28]. A recent 2016 systematic review found eight randomized controlled trials (RCTs) to date of mHealth interventions to support chronic conditions in LMICs [19], specifically in China, Honduras, India, Malaysia, Mexico, Pakistan, Taiwan, and Uruguay. The mHealth interventions included two studies on health promotion, five studies on remote monitoring and care (eg, appointment reminders and medication adjustments), and one study on a clinical decision support system. The studies targeted people living with diabetes, hypertension, and asthma, and all studies but one reported some improvements in their outcomes of interest, which included clinical measures, symptoms, health service utilization, and self-rated health. As such a diversity of interventions, outcomes, and diseases was targeted by these mHealth interventions, it is hard to generalize these findings.

Objectives

We aimed to contribute to the research and practice about chronic disease management in Cambodia by conducting an exploratory study to identify the facilitators and barriers to

diabetes and hypertension management and user preferences for mHealth-based messages for chronic disease self-management and community-clinical linkages.

Our specific formative research questions were as follows:

> *1. What are people living with diabetes and/or hypertension currently doing to manage their chronic condition?*
>
> *2. What are the facilitators and barriers to managing diabetes and/or hypertension, from the perspectives of people living with diabetes and/or hypertension (patients) including peer educators (PEs)?*
>
> *3. What are the user (patient and PE) preferences for the format and content of cell phone messages to support people living with diabetes and/or hypertension, including PEs?*
>
> *4. What are the facilitators and barriers to using mHealth technology from the perspectives of people living with diabetes and/or hypertension, including PEs?*

We used these findings to develop an mHealth mobile-based messaging intervention to improve chronic disease management in Cambodia. As such, our fifth research question was as follows:

> *5. How acceptable, appropriate, and feasible are the mHealth messages to people living with diabetes and/or hypertension, and what revisions are needed to improve the mHealth messages before the RCT effectiveness-implementation study?*

This qualitative study is the first phase of a larger RCT to test the effectiveness and implementation of the mHealth intervention to improve clinical outcomes via better evidence-based disease management practices (eg, self-management, regular monitoring of blood pressure and glucose, and medication adherence) for people living with diabetes and hypertension in Cambodia.

Methods

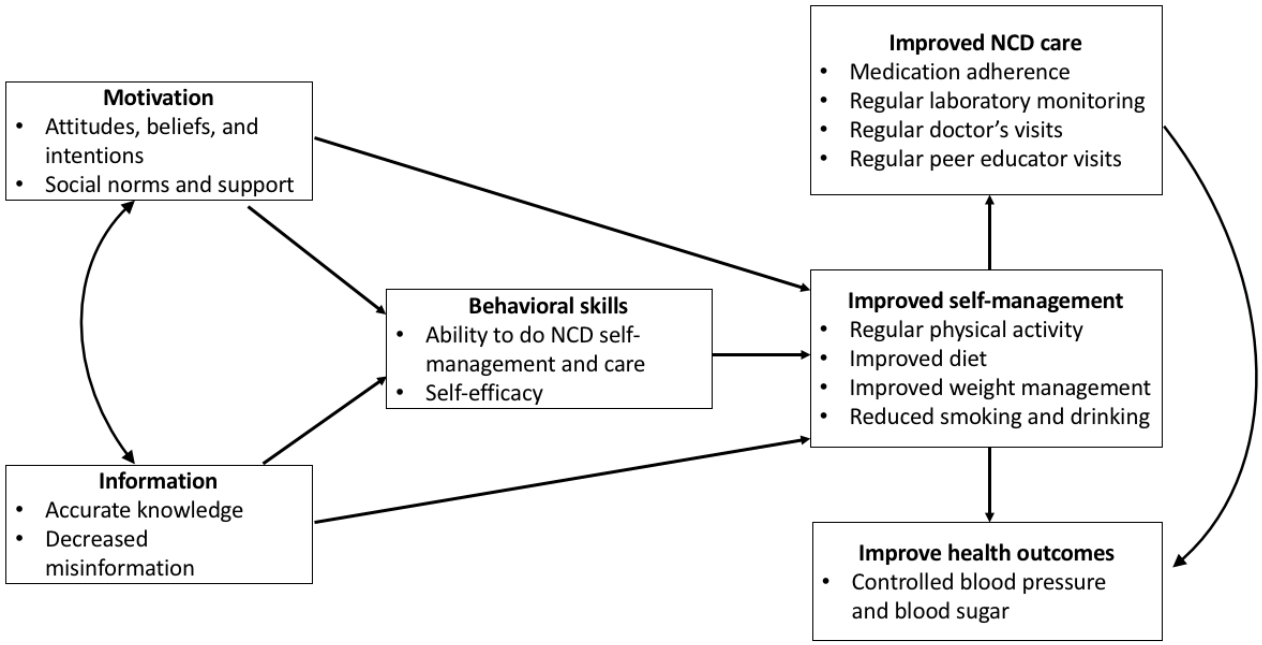
Design

We used an exploratory qualitative study design to better understand patient and PE perspectives on chronic disease management and mHealth. We used the 32-item checklist Consolidated Criteria for Reporting Qualitative Research [29] to guide the design, data collection, analysis, and reporting of our research study. We received institutional review board approval from the University of Washington Human Subjects Division and the Cambodian National Ethics Committee for Health Research.

Theoretical Framework

We used the information-motivation-behavioral (IMB) theoretical framework [30] to guide data collection and mHealth message development. The IMB is a well-established behavior change model that has been applied to other diabetes self-management interventions to improve clinical and care outcomes [31,32]. Figure 1 shows our modified IMB model for this study. The IMB suggests three main determinants of starting and maintaining health behaviors: (1) accurate information (which includes reducing misinformation) that can be easily translated into health behavior changes, (2) personal and social motivation to act on this information, and (3) behavioral skills to implement the health behavior with confidence and effectiveness [30]. Our study will gather information on facilitators and barriers to chronic disease management to craft mHealth messages that can improve knowledge, motivation, and skills for self-management. Improving these three behavioral determinants can then improve chronic disease care (medication adherence, regular monitoring of blood pressure and glucose, regular doctor consultations, and regular visits with PEs) and ultimately clinical outcomes (controlled blood pressure and blood sugar). In addition, improved health outcomes can in turn motivate people to maintain change, which may reduce intervention fatigue and improve the sustainability of the intervention.

Figure 1. Modified information-motivation-behavioral model for improving diabetes and hypertension management through a mobile health intervention in Cambodia. NCD: noncommunicable disease.



Participants and Setting

The study was conducted in close partnership with the Cambodian nongovernmental organization (NGO), MoPoTsyo Patient Information Centre [33-35]. MoPoTsyo was established in 2005 to improve NCD management in Cambodia using the six strategies for improving the quality of chronic disease care, as recommended by the chronic care model [36]: self-management support, decision support, delivery system design, clinical information systems, health care organization, and community resources. MoPoTsyo trains people living with diabetes and/or hypertension to serve as PEs for their community, similar to a lay health worker or community health worker model. These PEs provide education and support for chronic conditions (eg, nutrition guidelines and self-monitoring of blood glucose and pressure) as well as linkages with the health system for regular laboratory profiles, medical consultations, and routine medication dispensing. To facilitate community-clinical linkages and chronic disease management, MoPoTsyo maintains a centralized database that tracks participants' demographics and health outcomes, laboratory results, doctor's visits, and pharmacy invoices when patients pick up medications.

Recruitment

We recruited the study participants from two poor urban areas in the capital, Phnom Penh, and four areas where PE networks cover rural operational districts: Chamkar Leu, Baray-Santuk, Kampong Speu, and Kong Pisey. These six areas were purposefully selected using maximum variation sampling [37] to represent different patient demographics and access to MoPoTsyo's chronic disease management services. We used convenience sampling to select focus group (FG) and interview (IW) participants from these six areas. For FGs with people living with diabetes and/or hypertension (called *patients* by MoPoTsyo), the eligibility criteria included active participation in MoPoTsyo's peer education program and being present for their doctors' consultation on the day of the FG. PEs notified patients of the FG opportunity and explained that they were not required to participate to receive other services from MoPoTsyo. For the IWs, all PEs were invited to participate from these six geographical areas. We obtained written informed consent from both patients and PEs.

Data Collection

The study team developed a semistructured IW and FG guide to better understand patient and PE perspectives on chronic disease management and mHealth. The guide included main questions and follow-up questions and probes depending on how participants answered the main questions. Participants could skip any question that they did not wish to answer. We also conducted a brief demographic survey at the start of each FG and IW. Data were collected by trained qualitative researchers: HH conducted the FGs and half of the IWs in Khmer and LS conducted half of the IWs in the English language with a Khmer interpreter. The FGs lasted 90 min and the IWs lasted 20 to 30 min. Each group included 8 to 12 participants with a similar composition of mixed ages and genders. Refreshments were provided at each group and IW, and the participants were provided a small amount of cash to reimburse them for the personal cost of their time and transportation. The FGs and IWs took place at the community health center or a nearby restaurant where participants came regularly and were familiar with.

Data Analysis

Hard copies and electronic data were stored securely and without any identifiable information. We analyzed the survey data using

Stata 14 [38] for descriptive statistics. FGs and IWs were audio recorded and transcribed, and the recordings were destroyed one month following transcription. We analyzed the transcripts using a deductive thematic analysis, using Atlas.ti for data management. A thematic analysis is a method for identifying, analyzing, organizing, describing, and reporting themes found within a dataset [39]. It applies a well-structured approach to understand and organize key features of the data and for highlighting the similarities and differences across diverse research participants and for generating unanticipated insights [40]. We used the six phases for the thematic analysis articulated by Nowell et al [40]: familiarizing ourselves with the data; coding data using a standardized coding scheme [41,42]; searching for themes from the coded and collated text; reviewing coded text within each theme to add additional codes or themes and collapse themes with insufficient or inconsistent data; defining, naming, and describing the scope and content of each theme and revising after a research team review; and producing a summary report of the findings to provide an account of the data across and within themes and quotes to illustrate each theme. We used this report to create mobile phone voice messages for the chronic disease management mHealth intervention.

We used Lincoln and Guba's [43] criteria for credibility, transferability, dependability, and confirmability to help establish the trustworthiness of our data. For credibility, or goodness of fit between participants' perspectives and how we represented these perspectives, we used prolonged engagement with the data, research triangulation with multiple members of our team, and data collection triangulation with the IW and FG data. For transferability, we provided descriptions of the setting, participants, and their perspectives so that the readers can determine whether the findings are applicable to their context. For dependability and confirmability, we created an audit trail of codes and themes to document decisions made throughout the course of the study.

Mobile Health Message Development

We used findings from IWs and FGs to develop mHealth cell phone messages to improve MoPoTsyo patients' clinical outcomes (eg, blood pressure and glucose). Specifically, the messages were created to improve evidence-based chronic disease management via better access to health care (doctor's consultations, laboratory monitoring, medications, and PEs) and self-management (diet, weight management, physical activity, alcohol, and smoking). We created a matrix with each of these eight categories and grouped current practices and facilitators and barriers to disease management by each of these categories. For each of these categories, we then developed message content aimed at improving knowledge, motivation, and skills to either strengthen the facilitators or address barriers that were identified in the FGs. We used the findings about how mHealth would help or hinder disease management to inform the format of the messages (eg, frequency and duration of messages). The research team reviewed the draft messages together to come up with final messages to record and pilot test.

We pilot tested four messages with 5 MoPoTsyo patients to assess message acceptability, appropriateness, and feasibility using open-ended discussion questions and a brief multiple-choice survey. In the survey, patients were asked to rate on a 5-point Likert scale (1=strongly disagree to 5=strongly agree) whether the messages: got their attention, were believable, were convincing, and were important to them; whether the messages put thoughts in their mind and helped them feel able to change how they were managing their chronic disease; and whether they agreed with or enjoyed the messages. We used input from the discussions and from the ratings to then revise the messages using this feedback before the RCT began.

Results

Participants

The study engaged 70 participants: 59 individuals (*patients*) and 11 PEs living with diabetes, hypertension, or both diabetes and hypertension (Table 1). Patients had a mean age of 55.9 years (SD 9.1), were 63% (37/70) women, completed on average 4.2 years (SD 3.2) of education, had smoked in the last 30 days (9/59, 15%), and had been in MoPoTsyo's program for approximately 3 years (SD 2.82). A majority of these patients were living with both diabetes and hypertension (43/59, 73%), and over 70% (33/46) with diabetes and 75% (42/56) with hypertension) of them felt they were successfully managing their disease. A greater proportion of PEs were men (6/11, 55%), were educated (mean years of school 8.6, SD 4.3), felt they were successfully managing their chronic conditions (over 10/11, 90% with diabetes and 8/8, 100% with hypertension), and enrolled in MoPoTsyo's program (mean 4.9 years, SD 2.3).

A summary of findings on patients' (FG) and PEs' (IW) perspectives on chronic disease management and mHealth support is provided in the following sections, along with illustrative quotes from patients and from PEs living with diabetes, hypertension, or both diabetes and hypertension.

Table 1. Study participants (N=70).

Participant characteristics	MoPoTsyo patients (focus groups)	Peer educators (interviews)	Total
Number of participants, n	59	11	70
Age (years), mean (SD)	55.9 (9.1)	52.8 (6.5)	55.4 (8.8)
Sex, n (%)			
Female	37 (63)	5 (45)	42 (60)
Male	22 (37)	6 (55)	28 (40)
Education			
School completed (years), mean (SD)	4.20 (3.15)	8.55 (4.25)	4.89 (3.67)
DM[a]+HTN[b], n (%)	43 (73)	8 (73)	51 (72)
DM only, n (%)	3 (5)	3 (27)	6 (9)
HTN only, n (%)	13 (22)	0 (0)	13 (19)
Successfully managing DM?, n (%)			
Yes, or more yes than no	33 (72)	10 (91)	43 (75)
No, or more no than yes	13 (28)	1 (9)	14 (25)
Successfully managing HTN?, n (%)			
Yes, or more yes than no	42 (75)	8 (100)	50 (78)
No, or more no than yes	14 (25)	0 (0)	14 (22)
Smoked in the past 30 days, n (%)	9 (15)	2 (18)	11 (16)
Length of time in the peer educator program (years), mean (SD)	2.9 (2.8)	4.9 (2.3)	3.2 (2.8)

[a]DM: diabetes.

[b]HTN: hypertension.

Chronic Disease Management: Knowledge, Attitudes, and Practices

Both patients and PEs were well informed about the recommended diabetes and hypertension management practices. Participants were knowledgeable about the following evidence-based practices: (1) taking daily medications to manage their blood pressure and blood glucose; (2) eating less salt and more fruits and vegetables; (3) exercising or other physical activities; (4) regularly meeting with their PE to receive support and education regarding their self-management activities; to monitor diabetes and hypertension outcomes; to discuss symptoms and complications; to discuss their medical prescription; to monitor their blood pressure, blood glucose, and weight; and to discuss their yearly visits for laboratory and doctor consultations. As one participant shared:

> *We need to do physical exercise and follow a healthy diet to reduce blood glucose and so on. I also take medicines regularly.* [FG3]

Very few participants mentioned traditional remedies for chronic disease management.

Although individuals and PEs knew about evidence-based chronic disease management strategies, there were deficits in IMB skills to carry out these strategies on a regular basis. Knowledge gaps included misunderstandings about what constitutes the appropriate frequency, intensity, and duration of physical activity and how to incorporate dietary changes into their lifestyle. As one patient shared:

> *For me, the most significant content is about physical exercise—how to do physical exercise properly and what are the advantages of doing physical exercise?* [FG6]

Challenges to incorporating recommended dietary changes into their daily routine included that household meals are prepared by other household members and that other household members prefer more salt and sugar for better taste, they work long days and get fatigued when they do not eat their typical foods, and healthier foods are more expensive and harder to access.

MoPoTsyo helps with access to care by providing services as close as possible to where people live and having a fixed low price for medications, laboratory tests, and consultations that patients get through their PE network membership. That said, even with this assistance from MoPoTsyo, resources such as time and money are limited for accessing care as many people live in rural areas and spend half a day travelling to and from and waiting for services as there is a very limited public transportation system in Cambodia. As one PE shared:

> *At the start of the month, their children get salary for working and they have money to come hospital, so there are more patients. But at the end of the month, there are fewer and fewer patients to come because they do not have money for firstly, transportation cost, secondly, consultation fee and thirdly, medicine payment. For the furthest village from Orm Laing, they have to spend about 6 to 7.50 USD for*

> *transportation, this doesn't include expense for food.* [IW9]

Access to quality care is further limited in rural areas where doctors have varying training in chronic disease management and laboratory monitoring is not offered regularly. Patients may also forget to take medicines regularly or may have difficulty taking medicines with food when they are working during the day. As one patient described:

> *My problem is that I am mostly absent-minded. For instance, I am rushing to go to work in the morning! I do not have time to have breakfast, so I plan to have breakfast at workplace, and then I forget to take my medicines.* [FG4]

Family and peer support are important cultural norms that facilitate chronic disease management, as well as having routines and feeling better when they follow recommended strategies. For example:

> *We ask their family to help remind them about consultation appointment; especially the patients with hypertension to see their doctor regarding the appointment because they may not feel badly.* [IW9]

> *Every morning, even when I was very busy, I prepared my medicines on my desk. So, if I walked around in my house, I took a look to see my medicines on my desk, then I recognized immediately, and I took the medicines soon. So I never miss in taking the medicines.* [FG4]

Mobile Health for Chronic Disease Management

Patients and PEs viewed mHealth mobile-based messages as an acceptable, feasible, and appropriate intervention to support chronic disease management. Cell phones are increasingly accessible in Cambodia—one FG participant shared that "Nowadays, Cambodian people are having a mobile phone in every house and a person can have more than one phone" [FG3]. However, participants highlighted that older, poorer, and rural communities have less access or have less phone literacy even if they have access to a phone. mHealth messages are seen not just as ways to educate, support, and remind patients but also as a way to support PEs who can in turn reinforce the messages. Most participants preferred voice messages over text messages, given the small screen size on older phones, limited literacy and low vision (eg, a participant from FG1 shared that she would need her child to read text messages for her), and a preference for a familiar, friendly voice from a trusted organization (MoPoTsyo). Suggestions included "announcing MoPoTsyo before the messages start so they will trust the messages" [IW4] and "having a rhythm of a diabetes song so when I hear the song, I know it's MoPoTsyo!" [FG4]. A few PEs preferred text messages so that they could share them more easily with their patients.

Limited interactivity was desired ("just pressing the answer key and listening" [FG3]) given the low education and digital literacy of many middle-aged and older Cambodians. Participants recommended that messages be simple, interesting, and brief—as one participant requested, "use daily words; do not used technical words! Since we are farmers, please use simple words, then we can understand, if you use the words that we never heard before, it will be difficult to understand!" [FG5]. It was desired that messages be sent two times per week during dinnertime so that patients know when to expect them, could be available (not working or out on errands), have access to their cell phones (since many share phones with spouses or children), and have their family around to support them in listening to and carrying out message recommendations. As one PE described:

> *two times per week is enough; if we send the messages more often, it will be hard and it could annoy their time; the patients would feel boring with the messages and then that would be a problem; or the patients may not have to time for messages.* [IW11]

Participants also wanted to be able to listen to the mHealth messages more than once and did not see each message as a onetime resource or as an individual-level intervention—they saw utility for their whole family and community for promoting health.

Sharing phones with family members was seen as both a facilitator and barrier to using an mHealth intervention for chronic disease management. As described earlier, sharing phones meant that family members would often be around when they received the mHealth messages, providing support, reinforcement, and reminders about what was shared. However, sharing phones meant that the targeted MoPoTsyo participant was not always near the phone through which the message was sent. One PE shared:

> *The difficulty is that the person who handles the phone is not always staying near the patients; the phone is with their children who are working far from home, while their parents who are patients are staying at home! So the message may not be received by the patients but by their children; even though their children will convey the message content, it is not like what the patients will hear the message by them directly!* [IW8]

Other barriers to cell phone access include cell phone numbers changing often for patients to get low promotional SIM cards and limited cell phone literacy. As one participant shared:

> *My husband has one, my child has one, but I don't have one—I don't know how to use the phone!* [FG6]

In addition, participants viewed mHealth as an intervention that could help connect the dots for chronic disease management, supplementing the information, motivation, and skills they receive from their PE as well as linking them to needed medicines, doctor consultations, and laboratory monitoring. One PE shared:

> *The messages through a mobile phone can be an additional support to educate patients besides the verbal explanation by peer educators to patients. The messages could help improve patient's practice lifestyle such as attitude in taking medicines, physical exercise and checking blood pressure and blood glucose! The message is a reminder or a doctor to remind patient to practice a healthy lifestyle. [IW9])*

However, they did not see mHealth as a panacea as it cannot fully solve problems for patients having limited time, money, and other resources to manage their conditions. As one patient described:

> *I don't have any difficulties [using the phone], but on the day I have to come to buy medicines, I don't have much time because I need to look after my grandchild at home. It is hard for me.* [FG1]

Mobile Health Message Development

Table 2 shows the four messages that were pilot tested. During the pilot test, patients strongly agreed with the message's acceptability, agreeability, and feasibility, with a 80%-100% (4/5-5/5) rating of the survey statements as *strongly agree*. Furthermore, the advisory committee suggested shortening and simplifying several of the messages that were too long or complex and changing the voice to a more upbeat, jovial female voice from the more formal, masculine voice that was pilot tested.

Overall, 34 mHealth messages were created to improve chronic disease management for Cambodians living with diabetes, hypertension, or both conditions. Guided by the IMB model, the messages were (1) *informational* (eg, reminders that it was time to pick up medications, get laboratory tests, or go to their doctor's consultations and education about what constitutes healthy food); (2) *motivational* (eg, rationales for why exercising, diet, and medications are important for chronic disease management and supportive messages that acknowledge how hard it can be to manage their chronic condition, such as eating less salt when not in charge of cooking for the family), and (3) *behavioral* (eg, creative strategies for incorporating regular medications, physical activity, and healthy eating into their daily life). Multimedia Appendix 1 provides information on each of the mHealth messages, including the message script, length of message (in seconds), type (IMB framework), chronic disease management topic, and target (whether the message was sent to all MoPoTsyo patients or a subset of MoPoTsyo patients).

In the RCT effectiveness study of the mHealth messaging intervention [44], the messages will be targeted and tailored to MoPoTsyo participants according to clinical and service outcomes. For example, patients due to pick up their medications will receive the following message:

> *It is time to go to the pharmacy with your patient-book to buy your medicine for your prescription [Information and Behavior]. Remember—taking your medicine every day even if you do not feel sick will help prevent serious complications. [Information, Motivation, and Behavior] If you sometimes forget to take medicine, ask your family to help you remember when to take your medicine [Behavior]. They care about you!* [Motivation]

Table 2. Messages for pilot testing.

Message ID	Khmer script	English script
A1		It is time to go to the pharmacy with your patient-book to buy your medicine for your prescription. Remember—taking your medicine every day even if you do not feel sick will help prevent serious complications. If you sometimes forget to take medicine, ask your family to help you remember when to take your medicine. They care about you!
B3		By updating the lab profile at least one time per year the Doctor can see if he is prescribing the right treatment specially for you. Remember to not eat or drink anything except pure water for 8 hours before the blood draw.
C2		It looks like you did not see the doctor for more than 1 year. If you went to another doctor for your diabetes or high blood pressure, press #1, if not, press #2. Going to your doctor is important so that they can prescribe the medications and other things you need for your blood pressure and blood sugar, to stay healthy and feel better. They also can tell you whether there are any medications that you do not need to take any more.
D1		It looks like your blood sugar (blood pressure) is too high which may make you feel worse or have other health problems. Please go to see a Doctor who can prescribe the right medicines for you. You can also contact your PE to get an appointment with a doctor.

Discussion

Summary of Findings

In this study, we identified the facilitators and barriers to diabetes and hypertension management in Cambodia from the perspectives of people living with these chronic conditions, including trained lay health workers (PEs) to support chronic disease management. Not surprisingly, given their involvement in the PE network, many people living with chronic conditions are aware of the best practices for disease management (eg, medications, regular doctor's visits and laboratory monitoring, physical activity and healthy eating, and less smoking and alcohol use). However, challenges remain on how to best incorporate these practices into daily lives. Key barriers include limited time and resources to access medications, clinical support, and recommended guidelines for physical activity (eg, aerobic and muscle strengthening activities) and healthier diets

with reduced sodium intake. Access to effective, quality chronic care remains to be a challenge, particularly in rural areas.

This study also identified the preferred mHealth mobile-based message format and content to inform our RCT of an mHealth intervention to strengthen self-management for diabetes and hypertension and to improve community-clinical linkages to existing services and support systems in Cambodia. Study participants see mHealth as providing opportunities for reminders about medications, laboratory studies, and doctor visits; education about how to include best practices for chronic disease management where they live and work; and support for those barriers that cannot easily be overcome. Many participants have access to cell phones and prefer simple, short voice messages over texting because of limitations in literacy, vision, and cell phone facilities as well as the trust conveyed by a familiar voice. Participants also prefer that messages be delivered at dinnertime to increase the likelihood that they will receive the messages and can engage their family and friends for support in their disease management. We did not find any major differences in perspectives on chronic disease management and mHealth between people living with chronic conditions (FG discussions) and those persons who were trained as PEs (one-on-one IWs).

Comparison With Existing Literature

The literature on chronic disease management in Cambodia is scarce. The few existing peer-reviewed publications highlight a variety of limitations and challenges in screening and diagnosis, service delivery, financing and insurance schemes, health workforce, drug supply, health information systems, and governance [45,46]. These findings are similar to many other studies conducted in LMICs specific to the management of hypertension and diabetes. Other common health service challenges reported in these studies include an insufficient number of trained health care workers coupled with a high patient load, few diagnostic capabilities especially for glycated hemoglobin, low remuneration of health workers often leading to low motivation, no electronic databases for monitoring patient data, limited services available in rural or community settings, and few trained specialists such as ophthalmologists or podiatrists [13,14,16,17,47-50]. Poor access to essential medicines is also reported as a common challenge in LMICs, where the drug stock is often unavailable, particularly in government subsidized facilities as well as rural settings [13,16,51,52]. These challenges in health service delivery prompted the creation of MoPoTsyo in 2005 to empower patients by (1) training and maintaining a lay health workforce that can support chronic disease self-management and (2) creating a financially sustainable system for continuous access to low-cost medications, laboratory monitoring, and doctor consultations that cooperates with Cambodia's public health care system.

As reported in our study, the financial burden for ongoing management of chronic conditions such as hypertension and diabetes are one of the most typical barriers that patients face. Financial barriers can include the cost of clinical care (eg, doctor fees, laboratory tests, medications, and catastrophic health expenditures), higher costs for healthy food, and transport costs for traveling to health facilities or pharmacies [13,14,16,17,47,49,53]. Khatib et al [54] reported that in rural LMICs, expenditures on hypertension medications alone can constitute up to 49% (IQR 20-100) of a household's capacity to pay. Some patients are forced to borrow or take money from friends or relatives, spend less on other household needs, or are unable to renew their required prescriptions in full or on time [50,55]. Although the costs of clinical care and medications within the MoPoTsyo network are more affordable than health services outside the network, participants still reported a cost burden from medications, laboratory tests, and annual visits with a physician. These costs were both direct monetary costs for these services as well as indirect costs from transportation or spending the day to travel for clinical care. This suggests additional structural support and system changes are needed to supplement an mHealth intervention and the access to care provided via MoPoTsyo's negotiated low, fixed prices for health services.

Patients also face a variety of personal challenges in implementing changes to their behaviors and lifestyle, as was confirmed by our study. A variety of sources in LMICs report that the most common self-management practices for people living with diabetes and hypertension are medication use, diet control, and physical activity [55-59]. The literature discussing patient-reported barriers to medication adherence describes many similar challenges reported by participants in our study, including unwanted side effects, dislike of taking medications, the belief that medications are not necessary, and financial and access barriers, as described in the previous paragraph [16,55,58]. Adherence to a recommended diet is also a common patient-reported barrier, both in the literature from LMICs and in the results from our study. Challenges include understanding how to implement diet changes [57]; ease of access to unhealthy foods, particularly foods high in fat, salt, and sugar [17,58]; problems obtaining healthy food at home if they are not the primary cook in the household [49,58]; and still feeling hungry after eating the recommended diet [55]. Other patient barriers reported in the literature include cultural norms that are not conducive to healthy behaviors [48], not having access to devices for regular self-monitoring of blood glucose or blood pressure [49,60], and a preference of traditional remedies over Western medications [16,17]. These barriers were not cited frequently by either MoPoTsyo participants or the PEs.

Social support emerged as a key factor for both chronic disease management and for using mHealth to support disease management. The MoPoTsyo PE network [33] is built on the evidence that trained members of the patient community can provide social support (along with education, identification, and referral) for people living with chronic conditions [61]. Family members have often been identified as the main source of social support available to people living with diabetes and hypertension in LMICs [62], yet this can be both a facilitator and barrier, as was identified in this study. Specifically, although family members can provide instrumental support (eg, driving patients to appointments) and emotional support (eg, helping patients cope with their disease), recommended lifestyle changes can conflict with regular family routines and family members may provide misinformation or increase the stress levels of

patients [63]. Future mHealth interventions may benefit from engaging family members more directly given their key role in both cell phone access and usability and from facilitating recommended disease management strategies around diet and exercise and obtaining medications and accessing health care providers.

As described earlier, it is too early to conclude whether mHealth messaging interventions for chronic conditions in developing countries are effective or not [19], with studies showing mixed results [64-71]. Although it is not exactly clear what contributes to the success of mobile messaging programs, it is likely that features such as the specific content, frequency, timing, personalization, interactive capabilities, and usability can affect outcomes. For example, most interventions with successful outcomes reported sending participants messages at least once a week, whereas daily messages were typically viewed as intrusive by participants and sending messages less than once a week was too infrequent to have an impact. Similarly, according to the opinions of participants in our study, most preferred to receive two messages per week.

From the available literature, it appears that most mHealth messaging programs utilized only text messaging; few incorporated voice messages into the intervention [66,72,73]. However, our results showed that participants often preferred voice over text messages because of low literacy (both limited ability to read and to use technology), poor vision, limited familiarity with sending and receiving text messages, higher trust in a familiar voice, and the ability to replay messages if desired. These findings are not unique in low-income settings; for example, a study in India found that 89% of participants preferred receiving voice reminders for their medications over text messages [74]. Even with voice messages, more targeted and tailored dissemination efforts may be needed to reach older and poorer patients with less cell phone access and lower technological literacy.

Strengths and Limitations

The primary strength of this study is our conducting of formative research with our target audience to identify the facilitators and barriers to chronic disease management to develop an mHealth messaging intervention for improving self-management and access to care. We worked in close partnership with a successful Cambodian NGO (MoPoTsyo) to design an mHealth intervention for a real-world setting and patient population. Our MoPoTsyo coauthors were involved in each phase of the study and received training in research methods to build the capacity for future research studies. The lead author (LS) has over 15 years of experience conducting community-engaged, cross-cultural, qualitative research with patients and providers. The second and third authors (HH and MP) lead MoPoTsyo's work including engagement, training, monitoring, and evaluation. For the past 15 years, the last two authors (JL and AF) have worked closely with MoPoTsyo and separately with the lead author on health promotion research and practice with low-resource settings in the United States. This was the lead author's first project with MoPoTsyo and in Cambodia; as such, partnering with the study coauthors and the MoPoTsyo team was imperative to yield relevant findings to improve NCD outcomes in Cambodia. Over the course of the study, we also consulted with an advisory committee of key stakeholders (see Acknowledgements) to ensure the messages were designed for dissemination and sustainability and were aligning with national policies and priorities such as the *National Strategic Plan for Prevention and Control of Noncommunicable Diseases 2013-2020*.

The limitations of the study are common to formative research studies. We used a convenience sample of participants in an existing PE network; as such, our findings may not be generalizable to other contexts. FGs as a data collection method may also foster group thinking in which only some perspectives are shared. In this study, the participants knew each other from participating together in MoPoTsyo's peer networks, which may have influenced what responses were shared. In addition, although the use of the IMB framework is well established for designing individual behavior change interventions, newer models such as Opoku et al's mHealth realist framework [75] can be useful for evaluating the implementation of the mHealth intervention to provide a more systematic appraisal of contextual factors that support or hinder intervention implementation, sustainability, and scale-up [76]. Their realist framework combines two models to articulate mechanisms between individual and contextual factors and access to care: Anderson's Behavioral Model of Health Services Utilization [77] calls out three key factors that determine access to care: predisposing characteristics (eg, age and beliefs), enabling resources (eg, availability of providers), and need (eg, disease burden). Davis's Technology Acceptance Model [78] adds perceived usefulness and perceived ease of use as two additional factors that lead to use and acceptance of technology.

Conclusions

This formative study adds to the growing literature on using mHealth to support the self-management of diabetes and hypertension in LMICs. It provides a model for developing culturally appropriate mHealth NCD interventions in LMICs that are based on patient-identified factors. Specifically, this study identified the facilitators and barriers to chronic disease management and to using an mHealth messaging intervention to improve disease self-management and community-clinical linkages in a PE network in the Cambodian context. Our formative research informed the creation of 34 IMB mHealth messages that are currently being tested for effectiveness and implementation in a cluster RCT of 75 health center coverage areas in Cambodia. Our findings align with the existing literature on the facilitators and barriers to chronic disease management and mHealth messaging interventions in LMICs, suggesting that there may be implications for other settings. This study will help build the needed evidence for successful mHealth messaging interventions for NCDs in LMICs.

Acknowledgments

The authors thank MoPoTsyo's PEs and patients who participated in the IWs and FGs. They also thank the MoPoTsyo staff who facilitated participant recruitment. The authors also wish to thank the in-country study advisory committee: Dr Chhea Chhorvann (Chair); Dr Kol Hero, Director of Preventive Medicine Department at Ministry of Health (MOH); Ms Channe Suy, Innovative Support for Emergencies, Diseases, and Disasters iLab Southeast Asia; Dr Sok Po, Hospital Services Department, MOH; Dr Touch Khun, Preah Kossamak National Hospital; Dr Mam Sovatha, National Center for HIV/AIDS, Dermatology, and STD at the MOH; Dr Khim Sam Ath, NCDs and Health through the Life-Course, WHO, Phnom Penh. Funding for this study was provided by the National Institutes of Health Fogarty International Center and National Institute of Biomedical Imaging and Bioengineering mHealth: Technology and Outcomes in LMICs (R21) 2016-2018 (AF, principal investigator). The sponsor was not involved in the review or approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
mHealth message script, length, information-motivation-behavioral content, chronic disease management topic, and target audience (N=34).
[XLSX File (Microsoft Excel File), 22 KB - mhealth_v8i4e13536_app1.xlsx]

References

1. Islam SM, Purnat TD, Phuong NT, Mwingira U, Schacht K, Fröschl G. Non-communicable diseases (NCDs) in developing countries: a symposium report. Global Health 2014 Dec 11;10:81 [FREE Full text] [doi: 10.1186/s12992-014-0081-9] [Medline: 25498459]
2. Roser M, Ritchie H. Our World in Data. 2020. Burden of Disease URL: https://ourworldindata.org/burden-of-disease [accessed 2020-02-09] [WebCite Cache ID 75fr4R8uE]
3. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease study and the preventable burden of NCD. Glob Heart 2016 Dec;11(4):393-397. [doi: 10.1016/j.gheart.2016.10.024] [Medline: 27938824]
4. Dans A, Ng N, Varghese C, Tai ES, Firestone R, Bonita R. The rise of chronic non-communicable diseases in southeast Asia: time for action. Lancet 2011 Feb 19;377(9766):680-689. [doi: 10.1016/S0140-6736(10)61506-1] [Medline: 21269677]
5. Institute for Health Metrics Evaluation (IHME). 2017. Country Profiles: Cambodia URL: http://www.healthdata.org/cambodia [accessed 2020-02-09]
6. GBD 2016 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
7. Oum S, Prak PR, Khuon EM, Mey V, Aim S, Bounchan Y, et al. World Health Organization. 2010 Sep. Prevalence of Non-Communicable Disease Risk Factors in Cambodia: STEPS Survey, Country Report URL: https://www.who.int/ncds/surveillance/steps/2010_STEPS_Report_Cambodia.pdf [accessed 2019-01-24] [WebCite Cache ID 75fqDVEhY]
8. Chhoun P, Tuot S, Harries AD, Kyaw NT, Pal K, Mun P, et al. High prevalence of non-communicable diseases and associated risk factors amongst adults living with HIV in Cambodia. PLoS One 2017;12(11):e0187591 [FREE Full text] [doi: 10.1371/journal.pone.0187591] [Medline: 29121661]
9. Fisher EB, Boothroyd RI, Coufal MM, Baumann LC, Mbanya JC, Rotheram-Borus MJ, et al. Peer support for self-management of diabetes improved outcomes in international settings. Health Aff (Millwood) 2012 Jan;31(1):130-139 [FREE Full text] [doi: 10.1377/hlthaff.2011.0914] [Medline: 22232103]
10. Tsolekile LP, Abrahams-Gessel S, Puoane T. Healthcare professional shortage and task-shifting to prevent cardiovascular disease: implications for low- and middle-income countries. Curr Cardiol Rep 2015 Dec;17(12):115. [doi: 10.1007/s11886-015-0672-y] [Medline: 26482758]
11. Park PH, Wambui CK, Atieno S, Egger JR, Misoi L, Nyabundi JS, et al. Improving diabetes management and cardiovascular risk factors through peer-led self-management support groups in western Kenya. Diabetes Care 2015 Aug;38(8):e110-e111 [FREE Full text] [doi: 10.2337/dc15-0353] [Medline: 26207058]
12. Joshi R, Alim M, Kengne AP, Jan S, Maulik PK, Peiris D, et al. Task shifting for non-communicable disease management in low and middle income countries--a systematic review. PLoS One 2014;9(8):e103754 [FREE Full text] [doi: 10.1371/journal.pone.0103754] [Medline: 25121789]
13. Alberti H, Boudriga N, Nabli M. Primary care management of diabetes in a low/middle income country: a multi-method, qualitative study of barriers and facilitators to care. BMC Fam Pract 2007 Nov 9;8:63 [FREE Full text] [doi: 10.1186/1471-2296-8-63] [Medline: 17996084]

14. Beran D. The impact of health systems on diabetes care in low and lower middle income countries. Curr Diab Rep 2015 Apr;15(4):20. [doi: 10.1007/s11892-015-0591-8] [Medline: 25721248]
15. Mannava P, Abdullah A, James C, Dodd R, Annear PL. Health systems and noncommunicable diseases in the Asia-Pacific region: a review of the published literature. Asia Pac J Public Health 2015 Mar;27(2):NP1-N19. [doi: 10.1177/1010539513500336] [Medline: 24097936]
16. Mendenhall E, Norris SA. Diabetes care among urban women in Soweto, South Africa: a qualitative study. BMC Public Health 2015 Dec 26;15:1300 [FREE Full text] [doi: 10.1186/s12889-015-2615-3] [Medline: 26706228]
17. Risso-Gill I, Balabanova D, Majid F, Ng KK, Yusoff K, Mustapha F, et al. Understanding the modifiable health systems barriers to hypertension management in Malaysia: a multi-method health systems appraisal approach. BMC Health Serv Res 2015 Jul 3;15:254 [FREE Full text] [doi: 10.1186/s12913-015-0916-y] [Medline: 26135302]
18. Abaza H, Marschollek M. mHealth application areas and technology combinations*. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 8;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
19. 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]
20. Howitt P, Darzi A, Yang G, Ashrafian H, Atun R, Barlow J, et al. Technologies for global health. Lancet 2012 Aug 4;380(9840):507-535. [doi: 10.1016/S0140-6736(12)61127-1] [Medline: 22857974]
21. World Health Organization. World Health Statistics 2010. Geneva: World Health Organization; 2010.
22. International Telecommunication Union (ITU). The World in 2013: ICT Facts and Figures URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf [accessed 2019-01-24] [WebCite Cache ID 75fsBGJSb]
23. Kaplan WA. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? Global Health 2006 May 23;2:9 [FREE Full text] [doi: 10.1186/1744-8603-2-9] [Medline: 16719925]
24. Phong K, Solá J. The Asia Foundation. 2014 Oct. Mobile Phones in Cambodia 2014: Research Study URL: https://asiafoundation.org/resources/pdfs/MobilephonesinCB.pdf [accessed 2019-01-24] [WebCite Cache ID 75ftO5Q1F]
25. Shippee ND, Shah ND, May CR, Mair FS, Montori VM. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. J Clin Epidemiol 2012 Oct;65(10):1041-1051 [FREE Full text] [doi: 10.1016/j.jclinepi.2012.05.005] [Medline: 22910536]
26. Kadirvelu A, Sadasivan S, Ng SH. Social support in type II diabetes care: a case of too little, too late. Diabetes Metab Syndr Obes 2012;5:407-417 [FREE Full text] [doi: 10.2147/DMSO.S37183] [Medline: 23226028]
27. Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. Br Med J 2005 Mar 19;330(7492):644-648 [FREE Full text] [doi: 10.1136/bmj.330.7492.644] [Medline: 15774996]
28. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15:33 [FREE Full text] [doi: 10.1186/s12913-014-0654-6] [Medline: 25609559]
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Fisher JD, Fisher WA. The information-motivation-behavioral skills model. In: DiClemente RJ, Crosby RA, Kegler MC, editors. Emerging Theories in Health Promotion Practice and Research: Strategies for Improving Public Health. San Francisco, CA: Jossey-Bass; 2002:40-70.
31. Osborn CY, Egede LE. Validation of an information-motivation-behavioral skills model of diabetes self-care (IMB-DSC). Patient Educ Couns 2010 Apr;79(1):49-54 [FREE Full text] [doi: 10.1016/j.pec.2009.07.016] [Medline: 19699601]
32. Gao J, Wang J, Zhu Y, Yu J. Validation of an information-motivation-behavioral skills model of self-care among Chinese adults with type 2 diabetes. BMC Public Health 2013 Feb 4;13:100 [FREE Full text] [doi: 10.1186/1471-2458-13-100] [Medline: 23379324]
33. van Olmen J, Eggermont N, van Pelt M, Hen H, de Man J, Schellevis F, et al. Patient-centred innovation to ensure access to diabetes care in Cambodia: the case of MoPoTsyo. J Pharm Policy Pract 2016;9:1 [FREE Full text] [doi: 10.1186/s40545-016-0050-1] [Medline: 26798483]
34. Taniguchi D, LoGerfo J, van Pelt M, Mielcarek B, Huster K, Haider M, et al. Correction: Evaluation of a multi-faceted diabetes care program including community-based peer educators in Takeo province, Cambodia, 2007-2013. PLoS One 2018;13(4):e0196192 [FREE Full text] [doi: 10.1371/journal.pone.0196192] [Medline: 29664971]
35. MoPoTsyo %. URL: https://www.mopotsyo.org/ [accessed 2020-02-12]
36. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
37. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]

38. StataCorp. College Station, TX: StataCorp; 2015. Stata Statistical Software: Release 14 URL: https://www.stata.com/stata14/ [accessed 2020-02-12]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1):1-13 [FREE Full text] [doi: 10.1177/1609406917733847]
41. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Fourth Edition. Thousand Oaks, CA: Sage; 2014.
42. King N. Using templates in the thematic analysis of text. In: Cassel C, Symon G, editors. Essential Guide To Qualitative Methods In Organizational Research. London, UK: Sage; 2004:257-270.
43. Lincoln YS, Guba E. Naturalistic Inquiry. Thousand Oaks, CA: Sage; 1985.
44. Fitzpatrick AL, van Pelt M, Heang H, Steinman L, Ide N, Chhea C, et al. Using targeted mHealth messages to address hypertension and diabetes self-management in Cambodia: protocol for a clustered randomized controlled trial. JMIR Res Protoc 2019 Mar 19;8(3):e11614 [FREE Full text] [doi: 10.2196/11614] [Medline: 30888330]
45. Bigdeli M, Jacobs B, Men CR, Nilsen K, van Damme W, Dujardin B. Access to treatment for diabetes and hypertension in rural Cambodia: performance of existing social health protection schemes. PLoS One 2016;11(1):e0146147 [FREE Full text] [doi: 10.1371/journal.pone.0146147] [Medline: 26815916]
46. Jacobs B, Hill P, Bigdeli M, Men C. Managing non-communicable diseases at health district level in Cambodia: a systems analysis and suggestions for improvement. BMC Health Serv Res 2016 Jan 27;16:32 [FREE Full text] [doi: 10.1186/s12913-016-1286-9] [Medline: 26818827]
47. van Minh H, Do YK, Bautista MA, Anh TT. Describing the primary care system capacity for the prevention and management of non-communicable diseases in rural Vietnam. Int J Health Plann Manage 2014;29(2):e159-e173. [doi: 10.1002/hpm.2179] [Medline: 23553675]
48. Krishnan AG, Kahandaliyanage A, Garg R. Hypertension in the South-East Asia Region: An Overview. Regional Health Forum 2013;17(1):7-14 [FREE Full text]
49. Abdulrehman MS, Woith W, Jenkins S, Kossman S, Hunter GL. Exploring cultural influences of self-management of diabetes in Coastal Kenya: an ethnography. Glob Qual Nurs Res 2016;3:2333393616641825 [FREE Full text] [doi: 10.1177/2333393616641825] [Medline: 28462335]
50. Beran D, Higuchi M. Delivering diabetes care in the Philippines and Vietnam: policy and practice issues. Asia Pac J Public Health 2013 Jan;25(1):92-101. [doi: 10.1177/1010539511412177] [Medline: 21727084]
51. Metta E, Haisma H, Kessy F, Geubbels E, Hutter I, Bailey A. 'It is the medicines that keep us alive': lived experiences of diabetes medication use and continuity among adults in Southeastern Tanzania. BMC Health Serv Res 2015 Mar 18;15:111 [FREE Full text] [doi: 10.1186/s12913-015-0768-5] [Medline: 25890162]
52. Beran D, Ewen M, Laing R. Constraints and challenges in access to insulin: a global perspective. Lancet Diabetes Endocrinol 2016 Mar;4(3):275-285. [doi: 10.1016/S2213-8587(15)00521-5] [Medline: 26857998]
53. Kankeu HT, Saksena P, Xu K, Evans DB. The financial burden from non-communicable diseases in low- and middle-income countries: a literature review. Health Res Policy Syst 2013 Aug 16;11:31 [FREE Full text] [doi: 10.1186/1478-4505-11-31] [Medline: 23947294]
54. Khatib R, McKee M, Shannon H, Chow C, Rangarajan S, Teo K, PURE study investigators. Availability and affordability of cardiovascular disease medicines and their effect on use in high-income, middle-income, and low-income countries: an analysis of the PURE study data. Lancet 2016 Jan 2;387(10013):61-69. [doi: 10.1016/S0140-6736(15)00469-9] [Medline: 26498706]
55. Bhandari P, Kim M. Self-care behaviors of Nepalese adults with type 2 diabetes: a mixed methods analysis. Nurs Res 2016;65(3):202-214. [doi: 10.1097/NNR.0000000000000153] [Medline: 27124256]
56. Mogre V, Abanga ZO, Tzelepis F, Johnson NA, Paul C. Adherence to and factors associated with self-care behaviours in type 2 diabetes patients in Ghana. BMC Endocr Disord 2017 Mar 24;17(1):20 [FREE Full text] [doi: 10.1186/s12902-017-0169-3] [Medline: 28340613]
57. Rahman AR, Wang J, Kwong GM, Morales DD, Sritara P, Sukmawan R, all members of the Asian Cardiovascular Expert Forum Committee. Perception of hypertension management by patients and doctors in Asia: potential to improve blood pressure control. Asia Pac Fam Med 2015;14(1):2 [FREE Full text] [doi: 10.1186/s12930-015-0018-3] [Medline: 25729324]
58. Shima R, Farizah MH, Majid HA. A qualitative study on hypertensive care behavior in primary health care settings in Malaysia. Patient Prefer Adherence 2014;8:1597-1609 [FREE Full text] [doi: 10.2147/PPA.S69680] [Medline: 25484577]
59. Worku A, Abebe SM, Wassie MM. Dietary practice and associated factors among type 2 diabetic patients: a cross sectional hospital based study, Addis Ababa, Ethiopia. Springerplus 2015;4:15 [FREE Full text] [doi: 10.1186/s40064-015-0785-1] [Medline: 25635244]
60. Camara A, Baldé NM, Sobngwi-Tambekou J, Kengne AP, Diallo MM, Tchatchoua AP, et al. Poor glycemic control in type 2 diabetes in the South of the Sahara: the issue of limited access to an HbA1c test. Diabetes Res Clin Pract 2015 Apr;108(1):187-192. [doi: 10.1016/j.diabres.2014.08.025] [Medline: 25697633]

61. Alaofè H, Asaolu I, Ehiri J, Moretz H, Asuzu C, Balogun M, et al. Community health workers in diabetes prevention and management in developing countries. Ann Glob Health 2017;83(3-4):661-675 [FREE Full text] [doi: 10.1016/j.aogh.2017.10.009] [Medline: 29221543]
62. Lall D, Engel N, Devadasan N, Horstman K, Criel B. Models of care for chronic conditions in low/middle-income countries: a 'best fit' framework synthesis. BMJ Glob Health 2018;3(6):e001077 [FREE Full text] [doi: 10.1136/bmjgh-2018-001077] [Medline: 30687524]
63. Baig AA, Benitez A, Quinn MT, Burnet DL. Family interventions to improve diabetes outcomes for adults. Ann N Y Acad Sci 2015 Sep;1353:89-112 [FREE Full text] [doi: 10.1111/nyas.12844] [Medline: 26250784]
64. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
65. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-text adherence support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 9;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
66. 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 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
67. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]
68. Islam SM, Niessen LW, Ferrari U, Ali L, Seissler J, Lechner A. Effects of mobile phone sms to improve glycemic control among patients with type 2 diabetes in Bangladesh: a prospective, parallel-group, randomized controlled trial. Diabetes Care 2015 Aug;38(8):e112-e113. [doi: 10.2337/dc15-0505] [Medline: 26207059]
69. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
70. Van Olmen J, Kegels G, Korachais C, de Man J, Van Acker K, Kalobu JC, et al. The effect of text message support on diabetes self-management in developing countries - a randomised trial. J Clin Transl Endocrinol 2017 Mar;7:33-41 [FREE Full text] [doi: 10.1016/j.jcte.2016.12.005] [Medline: 29067248]
71. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, GISMAL group. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
72. Rodrigues R, Shet A, Antony J, Sidney K, Arumugam K, Krishnamurthy S, et al. Supporting adherence to antiretroviral therapy with mobile phone reminders: results from a cohort in South India. PLoS One 2012;7(8):e40723 [FREE Full text] [doi: 10.1371/journal.pone.0040723] [Medline: 22952574]
73. Kunutsor S, Walley J, Katabira E, Muchuro S, Balidawa H, Namagala E, et al. Using mobile phones to improve clinic attendance amongst an antiretroviral treatment cohort in rural Uganda: a cross-sectional and prospective study. AIDS Behav 2010 Dec;14(6):1347-1352. [doi: 10.1007/s10461-010-9780-2] [Medline: 20700644]
74. DeSouza SI, Rashmi MR, Vasanthi AP, Joseph SM, Rodrigues R. Mobile phones: the next step towards healthcare delivery in rural India? PLoS One 2014;9(8):e104895 [FREE Full text] [doi: 10.1371/journal.pone.0104895] [Medline: 25133610]
75. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017 Feb 6;15(1):24 [FREE Full text] [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
76. Opoku D, Busse R, Quentin W. Achieving sustainability and scale-up of mobile health noncommunicable disease interventions in Sub-Saharan Africa: views of policy makers in Ghana. JMIR Mhealth Uhealth 2019 May 3;7(5):e11497 [FREE Full text] [doi: 10.2196/11497] [Medline: 31066706]
77. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [doi: 10.2307/2137284] [Medline: 7738325]
78. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

Abbreviations

DALY: disability-adjusted life year
FG: focus group
IMB: information-motivation-behavioral
IW: interview

LMIC: low- and middle-income country
mHealth: mobile health
MOH: Ministry of Health
NCD: noncommunicable disease
NGO: nongovernmental organization
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.01.19; peer-reviewed by D Opoku, J Chen, R Rodrigues; comments to author 25.04.19; revised version received 04.07.19; accepted 24.09.19; published 24.04.20.

Please cite as:
Steinman L, Heang H, van Pelt M, Ide N, Cui H, Rao M, LoGerfo J, Fitzpatrick A
Facilitators and Barriers to Chronic Disease Self-Management and Mobile Health Interventions for People Living With Diabetes and Hypertension in Cambodia: Qualitative Study
JMIR Mhealth Uhealth 2020;8(4):e13536
URL: http://mhealth.jmir.org/2020/4/e13536/
doi:10.2196/13536
PMID:32329737

Original Paper

Patterns of Mobile Phone Ownership and Use Among Pregnant Women in Southern Tanzania: Cross-Sectional Survey

Lavanya Vasudevan[1,2,3], PhD, MPH, CPH; Jan Ostermann[2,3,4,5], PhD; Sara Marwerwe Moses[6], BA; Esther Ngadaya[6], MD, PGDHE, PhD; Sayoki Godfrey Mfinanga[6,7,8], MD, PhD

[1]Department of Family Medicine and Community Health, School of Medicine, Duke University, Durham, NC, United States

[2]Duke Global Health Institute, Duke University, Durham, NC, United States

[3]Center for Health Policy and Inequalities Research, Duke University, Durham, NC, United States

[4]Department of Health Services, Policy & Management, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[5]South Carolina SmartState Center for Healthcare Quality, University of South Carolina, Columbia, SC, United States

[6]Muhimbili Research Centre, National Institute for Medical Research, Dar-es-Salaam, United Republic of Tanzania

[7]Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, United Republic of Tanzania

[8]School of Life Sciences and Bioengineering, Nelson Mandela African Institution of Science and Technology, Arusha, United Republic of Tanzania

Corresponding Author:
Lavanya Vasudevan, PhD, MPH, CPH
Department of Family Medicine and Community Health
School of Medicine
Duke University
DUMC 104006
Durham, NC, 27710
United States
Phone: 1 9196131423
Email: lavanya.vasudevan@duke.edu

Abstract

Background: There is a paucity of subnational data on patterns of mobile phone ownership and use in Tanzania to inform the development of digital health interventions.

Objective: The aim of this study is to assess patterns of mobile phone ownership and use in pregnant women to inform the feasibility and design of digital health interventions for promoting timely uptake of childhood vaccines in southern Tanzania.

Methods: Between August and November 2017, pregnant women in their third trimester were enrolled at health facilities and from surrounding communities, and asked about their patterns of mobile phone ownership and use in an interviewer administered survey.

Results: Of 406 women, only 3 had never used a phone. Most women (>98%) could make and receive phone calls. Compared to urban women, rural women reported higher mobile phone use rates but were less likely to be sole owners of phones, and less likely to send or receive SMS, transact money, browse the internet, or use social media via mobile phones.

Conclusions: The findings suggest high feasibility for digital health interventions delivered via mobile phones to pregnant women in southern Tanzania. The feasibility of smartphone-based interventions or strategies relying on the use of social media or the internet is limited.

(JMIR Mhealth Uhealth 2020;8(4):e17122) doi:10.2196/17122

KEYWORDS

digital health; mobile health; pregnant women; Tanzania

Introduction

In recent years, many efforts have leveraged increasing mobile-cellular subscription rates in low- and middle-income countries (LMIC) as a mechanism to promote childhood vaccinations [1,2]. Examples of “digital health” strategies for promoting childhood vaccinations in LMICs include text message–based delivery of educational content, appointment reminders, conditional financial transfers, and tools to support

health care providers in vaccination delivery [1,2]. The feasibility, successful implementation, and scale-up of these digital health strategies depends on the availability of mobile-cellular infrastructure and patterns of mobile phone ownership and use in target populations. Mobile-cellular infrastructure and patterns of mobile use also provide insights about the feasibility and potential success of next-generation interventions and technologies (eg, smartphone apps, wearables).

The government of Tanzania has developed a detailed investment road map for the use of digital health interventions to strengthen the performance of the national health system [3]. Despite ranking 123rd among 143 countries in its Network Readiness Index, a measure summarizing the extent that countries benefit from the opportunities provided by information and communication technologies, Tanzania currently has the third highest mobile cellular subscription rate in East Africa at 77 per 100 people [4,5]. In addition, several large text messaging programs have been successfully implemented to deliver health information to target populations via mobile phones. Examples of such text messaging programs include the *Wazazi Nipendeni* multimedia campaign and the *mobile for reproductive health* educational messages for reproductive health [6,7] Yet, there is little subnational data on mobile phone ownership and use that are publicly available to inform the design and feasibility of digital health interventions.

In support of diverse research and intervention studies in urban and rural areas in Tanzania, we sought to develop a *mobile phone-assisted reminder and incentive system (mPARIS)*, a digital health system capable of, among other things, sending reminders and conditional financial transfers to mothers of newborn children as a means of promoting timely uptake of childhood vaccines [8,9]. To inform the feasibility and design of digital health interventions using *mPARIS*, we used a structured survey to assess pregnant women's mobile phone ownership and use in the Mtwara region in southern Tanzania. The findings of the survey are reported below.

Methods

Mobile phone ownership and use were evaluated to inform research on the feasibility and potential efficacy of SMS reminders and conditional financial incentives for improving the timeliness of childhood vaccinations in southern Tanzania. The study protocol for the parent study, including objectives, context, and methods, was previously published [8]. The protocol was registered in ClinicalTrials.gov (Protocol NCT03252288) and approved by the Institutional Review Boards at Duke University (Protocol 2017-0591) and the University of South Carolina (facilitated review, Pro00051213) in the United States, and the National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/2194). Methods pertinent to the analysis of mobile phone ownership and use are presented below.

The study was implemented in one rural and one urban district in the Mtwara region in southern Tanzania. In 2015, Mtwara ranked 13th in the Human Development Index among Tanzania's 21 regions. Between August and November 2017, pregnant women were recruited from 4 urban and 8 rural health facilities and their surrounding communities. Eligibility was limited to health facilities that regularly provide childhood vaccinations.

To participate in the study, women had to meet the following inclusion criteria: be 16 years or older, be in the third trimester of pregnancy, have access to a mobile phone, and provide informed consent. Eligible women receiving antenatal care at participating facilities were approached by trained study personnel and offered enrollment in the study. Participating women and local community leaders were asked to identify other pregnant women in their community, who, if eligible, were also offered enrollment in the study. Informed consent was obtained from all participating women. Upon enrollment, a trained research assistant administered a structured survey comprised of closed-ended questions to collect data on sociodemographic characteristics, reproductive and antenatal care history, and mobile phone ownership and use. Data were collected electronically using the Qualtrics (Provo, UT) survey platform installed on tablet devices.

Data were imported into the Stata version 16 (StataCorp, College Station, TX) statistical software for analysis. Descriptive summaries (means, standard deviations, and percentages) were calculated for relevant survey responses and are presented below. Differences between rural and urban women were evaluated using the 2-tailed Student's *t* test for continuous variables and chi-square tests for categorical variables.

Results

A total of 406 pregnant women were enrolled in the study. Table 1 summarizes their demographic characteristics, and Table 2 details their mobile phone ownership and use. Enrolled women were 28 years of age, on average. A majority of the women were married, had at least standard 7 schooling, and were either employed or self-employed. Education and employment distributions differed between rural and urban women. A majority of women had started their antenatal care in the second trimester and most had previously given birth.

Table 1. Characteristics of pregnant women in Southern Tanzania, 2017.

Characteristics	All (N=406)	Urban (n=212)	Rural (n=194)	*P* value[a]
Age (years), mean (SD)	27.9 (7.2)	27.5 (6.5)	28.4 (7.9)	.19
Marital status, n (%)				.42
Married	331 (81.5)	174 (82.1)	157 (80.9)	
Widowed	2 (0.5)	1 (0.5)	1 (0.5)	
Divorced/separated	20 (4.9)	7 (3.3)	13 (6.7)	
Never married	53 (13.1)	30 (14.2)	23 (11.9)	
Employment, n (%)				<.001
Unemployed/housewife	165 (40.6)	93 (43.9)	72 (37.1)	
Self-employed	209 (51.5)	94 (44.3)	115 (59.3)	
Employed	25 (6.2)	22 (10.4)	3 (1.5)	
Other (eg, student, casual laborer)	7 (1.7)	3 (1.4)	4 (2.1)	
Education, n (%)				<.001
None	103 (25.4)	24 (11.3)	79 (40.7)	
Standard 1-6	31 (7.6)	13 (6.1)	18 (9.3)	
Standard 7	205 (50.5)	123 (58.0)	82 (42.3)	
Form 1-4	54 (13.3)	40 (18.9)	14 (7.2)	
Form 5 or higher	13 (3.2)	12 (5.7)	1 (0.5)	
First antenatal care visit, n (%)				.86
First trimester	159 (39.2)	83 (39.2)	76 (39.2)	
Second trimester	219 (53.9)	114 (53.8)	105 (54.1)	
Third trimester	11 (2.7)	7 (3.3)	4 (2.1)	
Don't know	17 (4.2)	8 (3.8)	9 (4.6)	
Reproductive history, n (%)				.77
First birth	91 (23.6)	48 (24.2)	43 (23.0)	
Prior births	294 (76.4)	150 (75.8)	144 (77.0)	

[a]Denotes the statistical significance of differences between rural and urban participants.

Table 2. Mobile phone ownership and use among pregnant women in Southern Tanzania, 2017.

Survey Questions	All (N=406), n (%)	Urban (n=212), n (%)	Rural (n=194), n (%)	*P* value[a]
How often do you use a mobile phone?				<.001
Never	3 (0.7)	0 (0.0)	3 (1.5)	
Less than once a week	325 (80.0)	195 (92.0)	130 (67.0)	
At least once a week	52 (12.8)	12 (5.7)	40 (20.6)	
Everyday	26 (6.4)	5 (2.4)	21 (10.8)	
Who owns the mobile phone that you use?[b]				<.001
Woman only	232 (57.6)	148 (69.8)	84 (44.0)	
Father of the child to be born only	85 (21.1)	31 (14.6)	54 (28.3)	
Woman and father of the child, jointly	27 (6.7)	13 (6.1)	14 (7.3)	
Woman and other relative	8 (2.0)	6 (2.8)	2 (1.0)	
Others	51 (12.7)	14 (6.6)	37 (19.4)	
Have you ever used a mobile phone to:[b]				
Make phone calls	398 (98.8)	210 (99.1)	188 (98.4)	.57
Receive phone calls	400 (99.3)	210 (99.1)	190 (99.5)	.62
Send or receive SMS	311 (77.2)	182 (85.8)	129 (67.5)	<.001
Receive money	300 (74.4)	172 (81.1)	128 (67.0)	.001
Send money	248 (61.5)	154 (72.6)	94 (49.2)	<.001
Browse the internet	32 (7.9)	30 (14.2)	2 (1.0)	<.001
Use Facebook	37 (9.2)	31 (14.6)	6 (3.1)	<.001
Use WhatsApp	34 (8.4)	31 (14.6)	3 (1.6)	<.001
None of the above	1 (0.2)	1 (0.5)	0 (0.0)	.34
If an important message were to be delivered to you on a mobile phone, how soon would you receive it?[b]				.10
Same day	390 (96.8)	208 (98.1)	182 (95.3)	
Next day	12 (3.0)	3 (1.4)	9 (4.7)	
2-3 days	0 (0.0)	0 (0.0)	0 (0.0)	
4-7 days	0 (0.0)	0 (0.0)	0 (0.0)	
>7 days	1 (0.2)	1 (0.5)	0 (0.0)	
In the past month, how often have you had problems with charging phones?[b]				.17
Less than once per week	312 (77.4)	172 (81.1)	140 (73.3)	
One or more times per week	82 (20.3)	36 (17.0)	46 (24.1)	
Most days	9 (2.2)	4 (1.9)	5 (2.6)	
In the past month, how often have you had connection problems?[b]				.25
Less than once per week	361 (89.6)	193 (91.0)	168 (88.0)	
One or more times per week	40 (9.9)	19 (9.0)	21 (11.0)	
Most days	2 (0.5)	0 (0.0)	2 (1.0)	
How much do you spend per week on phone charges?[b]				<.001
0-499 TSH[c]	41 (10.2)	20 (9.4)	21 (11.0)	
500-999 TSH	185 (45.9)	78 (36.8)	107 (56.0)	
1000-1999 TSH	145 (36.0)	91 (42.9)	54 (28.3)	
2000-4999 TSH	28 (6.9)	20 (9.4)	8 (4.2)	

Survey Questions	All (N=406), n (%)	Urban (n=212), n (%)	Rural (n=194), n (%)	*P* value[a]
5000-9999 TSH	4 (1.0)	3 (1.4)	1 (0.5)	
Do you use any bundles?[b]				
Daily	261 (64.8)	136 (64.2)	125 (65.4)	.79
Weekly	169 (41.9)	106 (50.0)	63 (33.0)	<.001
Monthly	6 (1.5)	4 (1.9)	2 (1.0)	.49
None	50 (12.4)	16 (7.5)	34 (17.8)	.002

[a]Denotes the statistical significance of differences between rural and urban participants.

[b]Question skipped for 3 women who said they had never used a phone.

[c]TSH: Tanzanian shillings.

Only 3 women, all residing in rural areas, had never used a mobile phone. However, rural women reported more frequent mobile phone use than urban women. Most women either owned a mobile phone or could access one through family members including the father of the child to be born. Notably, fewer rural women reported being sole owners of phones compared to urban women. Similar percentages of urban and rural women reported using phones for making phone calls. However, compared to urban women, fewer rural women reported being able to communicate via text messaging and transact mobile money. Fewer than 10% of both urban and rural women reported using social media or the internet on phones.

Most urban and rural women said that important messages delivered to a mobile phone would reach them the same day. Problems related to phone charging and network connectivity were reported to be infrequent (less than once per week) by a majority of the women. Most women reported spending 500-2000 Tanzanian shillings (TSH; US $1 corresponded to approximately 2200 TSH at the time of the study) per week on phone-related charges.

Discussion

Study findings suggest high mobile phone access among pregnant women in the Mtwara region. Our data suggests the feasibility of text- or voice-based interventions delivered via mobile phones to pregnant women in the Mtwara region of southern Tanzania, but there was less readiness for smartphone-based interventions or strategies relying on the use of social media or the internet. High rates of mobile phone use for financial transactions (sending or receiving money) and the widespread use of bundles, both of which require menu-based interactions with phones, suggest that Unstructured Supplementary Service Data (USSD) assessments or interventions may be feasible. Potential applications could include short questionnaires for data collection or monitoring. Due to observed variations in phone access and ownership, it is likely that any mobile phone–based interventions will be delivered to shared phones for some participants. This finding may be of relevance to studies involving sensitive health topics (eg, HIV/AIDS and other infectious diseases; intimate partner violence).

According to our data, the population in the Mtwara region are relatively smartphone-naive, compared to some other Tanzanian cities like Dar-es-Salaam and Arusha, where 4G network infrastructure has been previously reported [10]. Smartphone-based interventions may be possible in Mtwara, but may require larger financial investments for the provision of smartphones to participants, as well as higher training needs to facilitate their use. We did not formally assess whether mobile networks in the Mtwara region would be able to support broad use of smartphones.

This study is subject to several limitations. First, only women who had access to a mobile phone were enrolled for the survey. Our study, therefore, excludes women from families that do not have access to a mobile phone. Literature suggests that women from socioeconomically disadvantaged families are more likely to lack mobile phone access [10]. Such women may be more vulnerable and could potentially benefit more from interventions promoting health information and health service use. Second, the selected facilities were within 20 kilometers of Mtwara, the urban regional capital and commercial center. Generalizability of the study findings is limited to the specific study area in late 2017, as mobile phone use and ownership characteristics may vary in other regions and is likely to increase over time. Third, the study provides only contextual information for the design of interventions; qualitative work concerning end user perceptions on feasibility and acceptability should complement the design and implementation of specific digital health interventions.

In conclusion, study findings suggest high feasibility of text-, voice-, or USSD-based interventions delivered via mobile phones to pregnant women in the Mtwara region of southern Tanzania. The use of social media and internet among pregnant women is limited. Future studies may use this study's findings to track growth in mobile phone ownership and changes in use patterns among pregnant women in the region.

Acknowledgments

The authors are grateful to the women who participated in the study. The authors would like to acknowledge the research assistants from the National Institute for Medical Research Muhimbili Medical Research Centre who supported data collection activities. The authors would like to thank the regional and district medical offices of Mtwara and their staff for their support of the study.

This study was supported by grants from the Fogarty International Center of the National Institutes of Health under Award Number R21TW010262 to Duke University, and the Maternal, Adolescent and Child Health (MACH) working group of the Duke Global Health Institute. LV was funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number 1KL2TR002554. The development of the mobile phone–based appointment reminder and incentive system (mParis) was supported by the National Institute of Mental Health under Award Number R01MH106388. The funding bodies have no role in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

LV, JO, EN, and SGM conceived the study; JO and LV designed the study protocol; EN and SMM led study implementation and data collection; and LV and JO carried out analysis and interpretation of these data. LV drafted the manuscript; JO, EN, SMM, and SGM critically revised the manuscript for intellectual content. All authors read and approved the final manuscript. LV and JO are guarantors of the paper.

Conflicts of Interest

None declared.

References

1. 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 Oct 04;5(10):e148 [FREE Full text] [doi: 10.2196/mhealth.7792] [Medline: 28978495]
2. Watterson JL, Walsh J, Madeka I. Using mHealth to improve usage of antenatal care, postnatal care, and immunization: a systematic review of the literature. Biomed Res Int 2015;2015:153402 [FREE Full text] [doi: 10.1155/2015/153402] [Medline: 26380263]
3. PATH.org. 2016 Dec. Data use partnership: the journey to better data for better health in Tanzania URL: https://tinyurl.com/rnm2leh [accessed 2020-03-05]
4. World Economic Forum. reports.weforum.org/. 2019. Network readiness index URL: https://tinyurl.com/rbqcno3 [accessed 2019-10-23]
5. WorldBank.org. 2019. Mobile cellular subscriptions (per 100 people) URL: http://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2019-10-23]
6. mHealth Tanzania. CDCfoundation.org. 2013 May 10. Wazazi Nipendeni: how text messages are supporting mothers in tanzania URL: https://www.cdcfoundation.org/blog-entry/wazazi-nipendeni [accessed 2020-03-05]
7. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
8. Ostermann J, Vasudevan L, Baumgartner JN, Ngadaya E, Mfinanga SG. Do mobile phone-based reminders and conditional financial transfers improve the timeliness of childhood vaccinations in Tanzania? Study protocol for a quasi-randomized controlled trial. Trials 2019 Jul 04;20(1):397 [FREE Full text] [doi: 10.1186/s13063-019-3430-4] [Medline: 31272487]
9. Ostermann J, Njau B, Hobbie A, Mtuy T, Thielman N. Pin127 - Heterogeneous preferences for HIV/AIDS counseling and testing among two high risk populations in Northern Tanzania - results from a discrete choice experiment. In: Value in Health. 2018 Oct Presented at: ISPOR Europe 2018: New Perspectives for Improving 21st Century Health Systems; 19-23 May 2018; Baltimore, MD p. S242 URL: https://www.sciencedirect.com/science/article/pii/S1098301518362508?via%3Dihub [doi: 10.1016/j.jval.2018.09.2875]
10. GSMA Connected Women. 2019. The Mobile Gender Gap Report 2019 URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf

Abbreviations

LMIC: low- and middle-income countries
mPARIS: mobile phone-assisted reminder and incentive system
TSH: Tanzanian shilling
USSD: Unstructured Supplementary Service Data.

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by T Scherr, S Zaidi, J Mwaisaka; comments to author 30.12.19; revised version received 04.02.20; accepted 05.02.20; published 08.04.20.

Please cite as:
Vasudevan L, Ostermann J, Moses SM, Ngadaya E, Mfinanga SG
Patterns of Mobile Phone Ownership and Use Among Pregnant Women in Southern Tanzania: Cross-Sectional Survey
JMIR Mhealth Uhealth 2020;8(4):e17122
URL: https://mhealth.jmir.org/2020/4/e17122
doi:10.2196/17122
PMID:32267240

Original Paper

An Automated Blood Pressure Display for Self-Measurement in Patients With Chronic Kidney Disease (iHealth Track): Device Validation Study

Victoria Mazoteras-Pardo[1*], MSc, PhD; Ricardo Becerro-De-Bengoa-Vallejo[1*], BSc, MLIS, DPM, PhD, DHL; Marta Elena Losa-Iglesias[2*], PhD, BSc, MsC, DP; Daniel López-López[3*], PhD, BSc, MsC, DP; César Calvo-Lobo[1*], PhD, PT, MSc; David Rodríguez-Sanz[1*], PhD, MsC, PT, DP; Eva María Martínez-Jiménez[4*], PhD, BSc, MsC, DP; Patricia Palomo-López[5*], PhD, MsC, DP

[1]Facultad de Enfermería, Fisioterapia y Podología, Universidad de Complutense de Madrid, Madrid, Spain

[2]Faculty of Health Sciences, Universidad Rey Juan Carlos, Alcorcón, Spain

[3]Research, Health and Podiatry Group, Faculty of Nursing and Podiatry, Departament of Health Sciences, Universidade da Coruña, Ferrol, Spain

[4]Facultad de Fisioterapia y Enfermería, Departamento de Enfermería, Universidad Castilla la Mancha, Toledo, Spain

[5]University Center of Plasencia, Universidad de Extremadura, Plasencia, Spain

[*]all authors contributed equally

Corresponding Author:
Daniel López-López, PhD, BSc, MsC, DP
Research, Health and Podiatry Group
Faculty of Nursing and Podiatry, Departament of Health Sciences
Universidade da Coruña
Campus Universitario de Esteiro S/N
Ferrol, 15403
Spain
Phone: 34 981337400 ext 3546
Email: daniellopez@udc.es

Abstract

Background: Hypertension is a global public health issue and is closely related to chronic kidney disorder (CKD). In people with CKD, strict monitoring of blood pressure is an important part of therapy.

Objective: The aim of this research was to validate the iHealth Track blood pressure monitoring device for patients with CKD according to the European Society of Hypertension International Protocol 2010 (ESH-IP2).

Methods: In total, 33 patients who received hemodialysis in Plasencia participated in the study. There were 9 successive measurements made, which conformed to the ESH-IP2. We calculated the differences between the standard reference device (Omron M3 Intellisense) and the test device (iHealth Track) for blood pressure and heart rate values. For 99 total comparisons of paired measurements, we classified differences into various categories (≤5 mmHg, ≤10 mmHg, and ≤15 mmHg for blood pressure; ≤3, ≤5, and ≤8 beats per minute for heart rate).

Results: In 90 of 99 systolic blood pressure and 89 of 99 diastolic blood pressure comparisons between the devices, measurement differences were within 5 mmHg. In 81 of 99 heart rate comparisons between the devices, measurement differences were within 3 beats per minute. The mean differences between the test and reference standard measurements were 3.27 (SD 2.99) mmHg for systolic blood pressure, 3.59 (SD 4.55) mmHg for diastolic blood pressure, and 2.18 (SD 2.75) beats per minute for heart rate. We also observed that for both systolic and diastolic blood pressure, 31 of 33 participants had at least two of three comparisons between the devices with measurement differences less than 5 mmHg. For heart rate, 28 of 33 patients had at least two of three comparisons between the devices with measurement differences less than 3 beats per minute.

Conclusions: To our knowledge, this is the first study to show that iHealth Track meets the requirements of the ESH-IP2 in patients with CKD. Therefore, the iHealth Track is suitable for use in renal patients.

(JMIR Mhealth Uhealth 2020;8(4):e14702) doi:10.2196/14702

KEYWORDS

iHealth Track; validation; blood pressure; heart rate; International Protocol

Introduction

Chronic kidney disorder (CKD) is a syndrome defined by persistent alterations in renal function or structure that cause complications in a patient's health. Some of the structural anomalies may be tumors, cysts, malformations, or atrophies. In addition, renal dysfunction can be manifested through alterations in the output or grade of urine, increased risks of intellectual disabilities in children, edema, and hypertension [1,2].

In fact, the diseases most related with CKD are hypertension and diabetes, especially in high- and middle-income countries [3,4]. Hypertension may simply be a consequence of CKD [5,6] or both a cause and consequence of CKD [7,8]. Hypertension may be due to hypervolemia or activation of the renin-angiotensin system or neurohumoral (catecholamine and aldosterone) axis. In addition, sometimes high blood pressure (BP) originates from calcineurin or corticosteroid inhibitors used to treat underlying kidney disorders [9].

The interaction between CKD and hypertension is complex and increases the probability of cerebrovascular and cardiovascular problems [9-11]. In several studies, cardiovascular events and deaths from any cause were reduced when systolic BP was <120 mmHg (compared to <140 mmHg) in patients with CKD and hypertension but not diabetes [11-13]. Therefore, strict control of BP is important for CKD therapy [14].

Monitoring of BP should be done with devices that are easy to use and accurate [11,15-18]. These devices must be tested and validated by independent experts (eg, the British Hypertension Society [19], the Association for the Advancement of Medical Instrumentation [20], and the European Society of Hypertension [21,22]) with protocols validated and designed expressly for BP monitoring [23]. The purpose of this study was to validate the iHealth Track BP monitoring device for self-measurement in patients with CKD, according to the European Society of Hypertension International Protocol 2010 (ESH-IP2). Therefore, the hypothesis of this study was that iHealth Track would be valid for the self-measurement of BP and heart rate (HR) in renal patients according to the ESH-IP2.

Methods

Ethical Information

The study protocol was approved by the Institutional Local Research and Ethical Committee (Universidad de Extremadura, Badajoz, Spain; record number 152/2019). In conducting this study, we complied with the ethical principles of the Declaration of Helsinki [24], including any emendations between 2000 and 2013. All participants provided signed informed consent prior to participating in this study.

The Devices

Omron M3 Intellisense

The standard device we used for reference was the Omron M3 Intellisense (Omron Healthcare, Kyoto, Japan), which has been validated according to the International Protocol for the general population [25] as well as CKD patients [26]. We purchased the Omron M3 Intellisense monitor from a local marketplace. The Omron M3 Intellisense is an oscillometric and automated upper-arm device for home BP monitoring. The device's standard arm cuff is 22 to 32 cm around, and a large cuff is also available for arm circumferences of 32 to 42 cm. The device uses IntelliSense technology to produce comfortable, controlled inflation without the need for pressure presetting or reinflation.

iHealth Track

The test device was the iHealth Track automatic appliance with serial number KN-550BT (iHealthLabs Europe, Paris, France), which registers brachial BP with the oscillometric protocol. It detects BP between the range of 0 mmHg to 300 mmHg (measuring precision ±3 mm Hg) and HRs within the range of 40 to 180 beats per min (measurement precision ±5%). The device's arm cuff is 22 to 42 cm around.

The device's liquid crystal display screen shows the measured systolic (S) BP, diastolic (D) BP, and HR values. The device unit has enough memory for 99 recordings. Additionally, this device unit can be used with Apple Bluetooth 4.0 devices and certain Android Bluetooth 4.0 cellular phones through an application named Health MyVitals. This means that BP and HR data can be stored on wireless devices connected to iHealth Track and then displayed graphically.

Patients and Recruitment

We recruited patients with CKD who attended the Fresenius Medical Care dialysis clinic in Plasencia, Spain. A total of 33 patients who met the selection criteria participated. The inclusion criteria were adults 25 years of age or older that received hemodialysis. We sought at least 10 male participants and 10 female participants. The exclusion criteria, which were created according to the ESH-IP2 [21,22], were sustained arrhythmia, circulatory problems where use of the cuff is contraindicated, and pregnancy.

Research Protocol

The professional validation team consisted of 2 nurses with senior experience (more than 6 years) in BP measurement. The measurement area was correctly conditioned to a suitable temperature, and factors that could affect the records, such as noise or distractions, were removed [21,22]. All measurements were made in the same room. The color of the room was white.

After dialysis, each patient first reported information regarding their sex, age, height, and dry weight. In addition, we calculated participants' BMI using Quetelet´s index in kg/m^2. The circumference of the patient's arm was measured to ensure that the cuff size was adequate.

Next, patients sat in the measurement room and BP measurements were started after a 10- to 15-minute rest period. Each patient was seated in a standard-size plastic chair with a backrest and armrests.

In total, 9 consecutive measurements were made on each participant with the Omron M3 Intellisense (5) and the iHealth Track (4) as follows [21,22]:

- (BPA): input BP, by the standard device unit
- (BPB): device BP detection by the test device unit
- (BP1): standard device unit
- (BP2): test device unit
- (BP3): standard device unit
- (BP4): test device unit
- (BP5): standard device unit
- (BP6): test device unit
- (BP7): standard device unit

During the measurements, participants remained calm, quiet, seated, and still. Participants kept their backs straight and feet on the floor in a parallel position (ie, without crossing their legs). They rested their arms on a flat surface with their palms facing upwards and their elbows slightly flexed so that their fists were at the height of their hearts.

BP records were made at heart level on the right arm in 31 participants and on the left arm in 2 participants (because of an arteriovenous fistula on the right arm). The standard cuff size (22-32 cm) for the Omron M3 Intellisense was used for all men (20). For women, the standard cuff (22-32 cm) was used for 11 participants, and the large cuff (32-42 cm) was used for 2 participants. Since the iHealth Track has only one cuff size (22-42 cm), all measurements were taken with it. The interval between one measurement and the next was 30 to 60 seconds [22].

All measurements were made for each participant during their hemodialysis appointment, after the dialysis was complete. The relative values were then used to calculate the mean difference between the reference device readings and the test device readings. All participants were receiving hemodialysis in the Fresenius Medical Care dialysis clinic in Plasencia.

Data Analysis

We analyzed the data with the software SPSS Statistics, version 19.0 (IBM, Armonk, New York). We reported the findings as mean (SD).

According to the ESH-IP2, the accuracy of a device is based on a comparison between the test device (iHealth Track) and the standard reference device (Omron M3 Intellisense) measurements. For each participant, we first compared measurements BP2, BP4, and BP6 with measurements BP1, BP3, and BP5, respectively, and then compared measurements BP3, BP5, and BP7 with each other. The most favorable comparisons were used.

In our comparisons, we classified differences for both SBP and DBP, separately, by whether paired values were within 5, 10, or 15 mmHg [22], and whether paired values for HR were within 3, 5, or 8 beats per minute (BPM). We determined whether the test device passed the ESH-IP validation protocol. Part 1 of the validation process concerns the number of differences allowed in the specified ranges of each measure (SBP, DBP, and HR) for comparisons of individual measurements between devices (99 measurements) [22]. Part 2 concerns the comparisons between devices of each measure for individual participants (33) [22].

Moreover, we produced Bland-Altman plots [27,28] to display the agreement between the two devices (the iHealth Track and the Omron M3 Intellisense). These plots show the difference between each pair of measurements on the y-axis against the mean of each pair of measurements on the x-axis (for SBP, DBP, and HR).

Results

Participants

Of the 34 participants we recruited, 33 completed the study successfully (one was excluded for device failure). The 33 participants included 20 men and 13 women. Table 1 shows a summary of their biometric characteristics.

Table 1. Participants' biometric characteristics.

Variables	Total sample (N=33)		Males (N=20)		Females (N=13)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age (years)	71.24 (11.76)	47.0-88.0	70.25 (11.42)	47.0-85.0	72.77 (12.57)	47.0-88.0
Weight (kg)	70.48 (15.87)	46.0-101.0	70.18 (12.95)	47-100.0	70.94 (20.16)	46.0-101.0
Height (cm)	162.24 (9.87)	141.0-180.0	167.10 (4.58)	160.0-180.0	154.77 (11.24)	141.0-174.0
BMI (kg/m^2)	26.95 (6.72)	18.0-44.0	25.07 (4.08)	18.0-33.0	29.84 (8.90)	19.0-44.0
Arm circumference (mm)	265.0 (33.12)	220.0-350.0	264.75 (26.13)	220.0-320.0	265.0 (33.12)	220.0-350.0

Blood Pressure Measurements

The iHealth Track BP device validation results were taken in accordance with the ESH-IP2. The mean differences between the reference standard and test devices were 3.27 (SD 2.99) mmHg for SBP and 3.59 (SD 5.28) mmHg for DBP. In 90 out of 99 SBP and 89 out of 99 DBP comparisons between the devices, measurement differences were within 5 mmHg, exceeding the ESH-IP thresholds (>72 comparisons for SBP and >64 comparisons for DBP). Additionally, in 95 out of 99 SBP and 94 out of 99 DBP comparisons between the devices, measurement differences were within 10 mmHg, also exceeding

the ESH-IP thresholds (>86 comparisons for SBP and >80 comparisons for DBP). Moreover, in 98 out of 99 SBP and 94 out of 99 DBP comparisons between the devices, measurement differences were within 15 mmHg, which again surpasses the ESH-IP thresholds (>95 comparisons for SBP and >92 comparisons for DBP). Therefore, the iHealth Track passed part 1 of the validation protocol for BP monitoring.

Regarding part 2 of the ESH-IP2, 31 out of 33 participants had at least two of the three comparisons between devices with measurement differences within 5 mmHg for both SBP and DBP, exceeding the ESH-IP threshold (>23 participants). One participant had all three comparisons for both SBP and DBP with measurement differences greater than 5 mmHg, which is less than the ESH-IP maximum of 3 participants. Given these results, the iHealth Track also passed part 2 of the validation protocol for BP. Thus, because the iHealth Track passed parts 1 and 2 of the BP validation protocol, it passed part 3 of the protocol, overall validation.

Heart Rate Measurements

The validation findings for the iHealth Track HR monitoring device were taken according to the ESH-IP2. The mean difference between the reference standard and test devices was 2.18 BPM (SD 2.75). In comparisons between devices, 81 out of 99 pairs of measurements were within 3 BPM, 91 out of 99 were within 5 BPM, and 96 out of 99 were within 8 BPM. These results indicate that the iHealth Track passed part 1 of the validation protocol for HR.

Regarding part 2 of the ESH-IP2, 28 out of 33 patients had at least two of three comparisons between devices with measurement differences within 3 BPM, which exceeds the ESH-IP threshold (>23 participants). Only 2 participants had all three HR comparisons with measurement differences greater than 3 BPM, which is less than the ESH-IP maximum of 3 participants. Therefore, the iHealth Track passed part 2 of the HR validation protocol and, consequently, part 3 (overall validation) as well.

The Bland-Altman graphs (Figures 1-3) give further information on the performance of the iHealth Track device. The graphs show that the measurement differences between the devices were fairly constant across the ranges of SBP, DBP, and HR.

Figure 1. Bland-Altman graph of systolic blood pressure differences between the iHealth Track and the Omron M3.

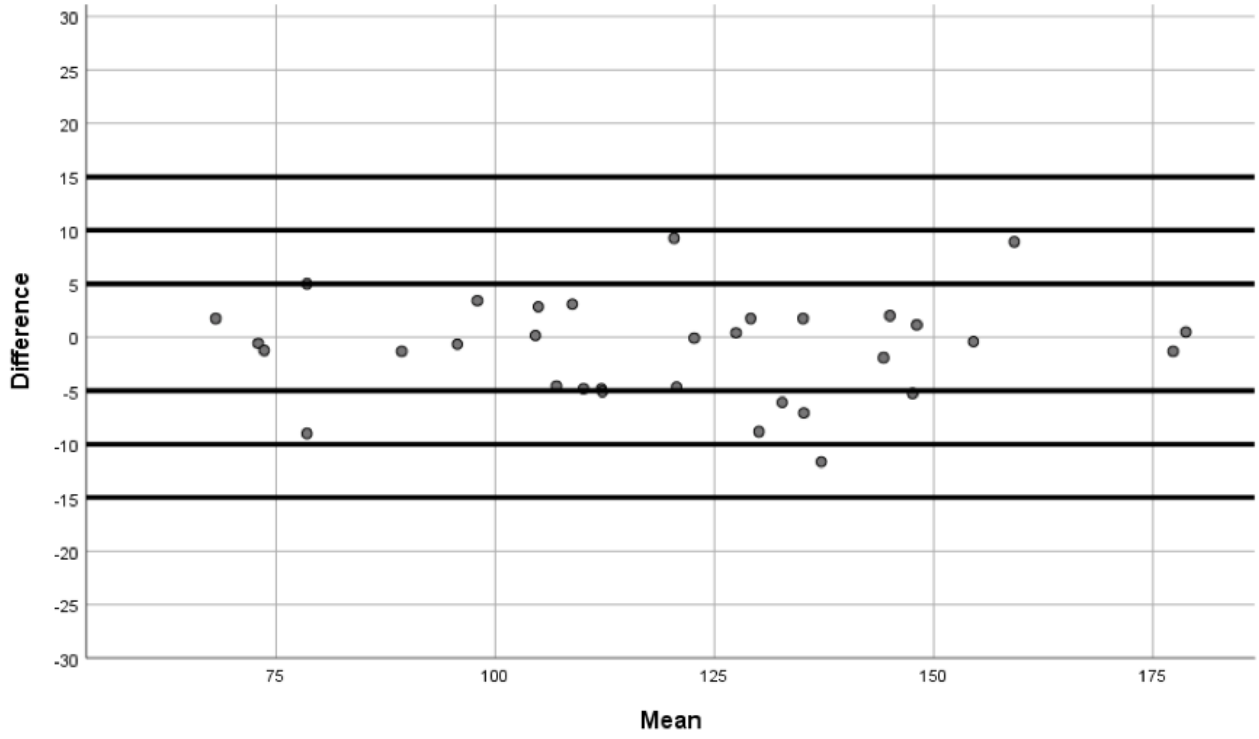


Figure 2. Bland-Altman graph for diastolic blood pressure differences between the iHealth Track and the Omron M3.

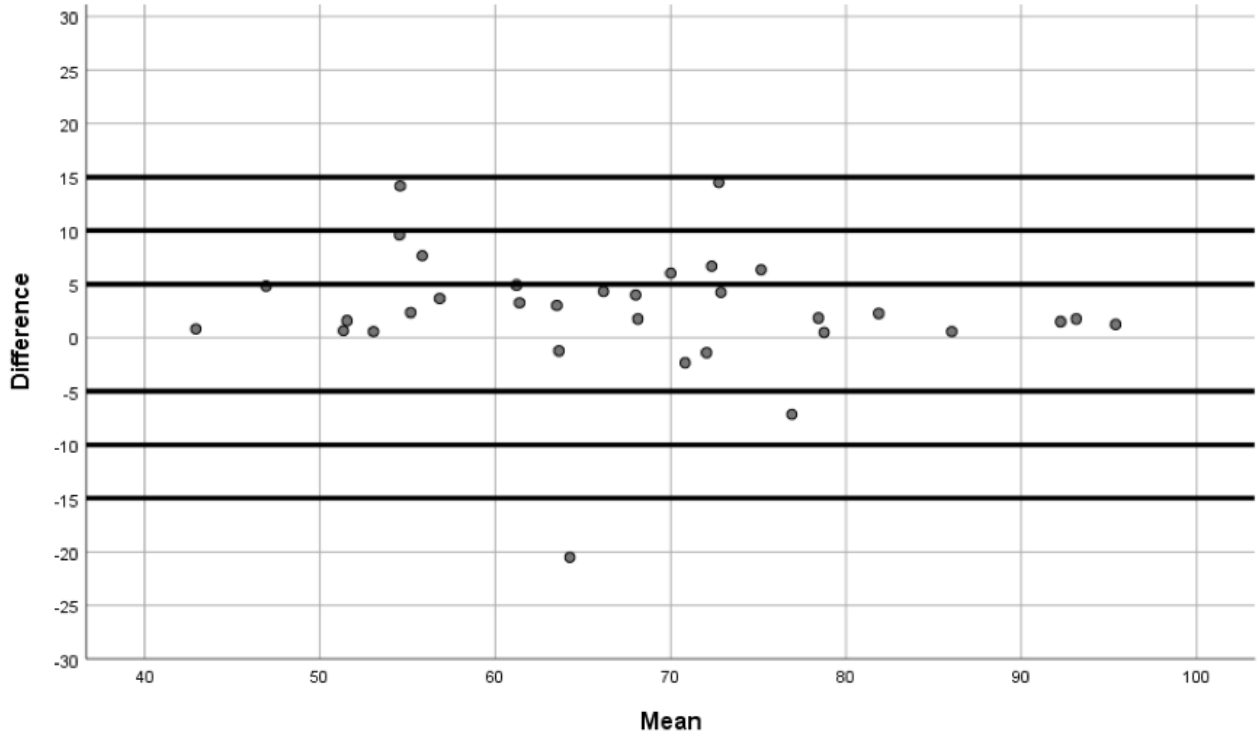
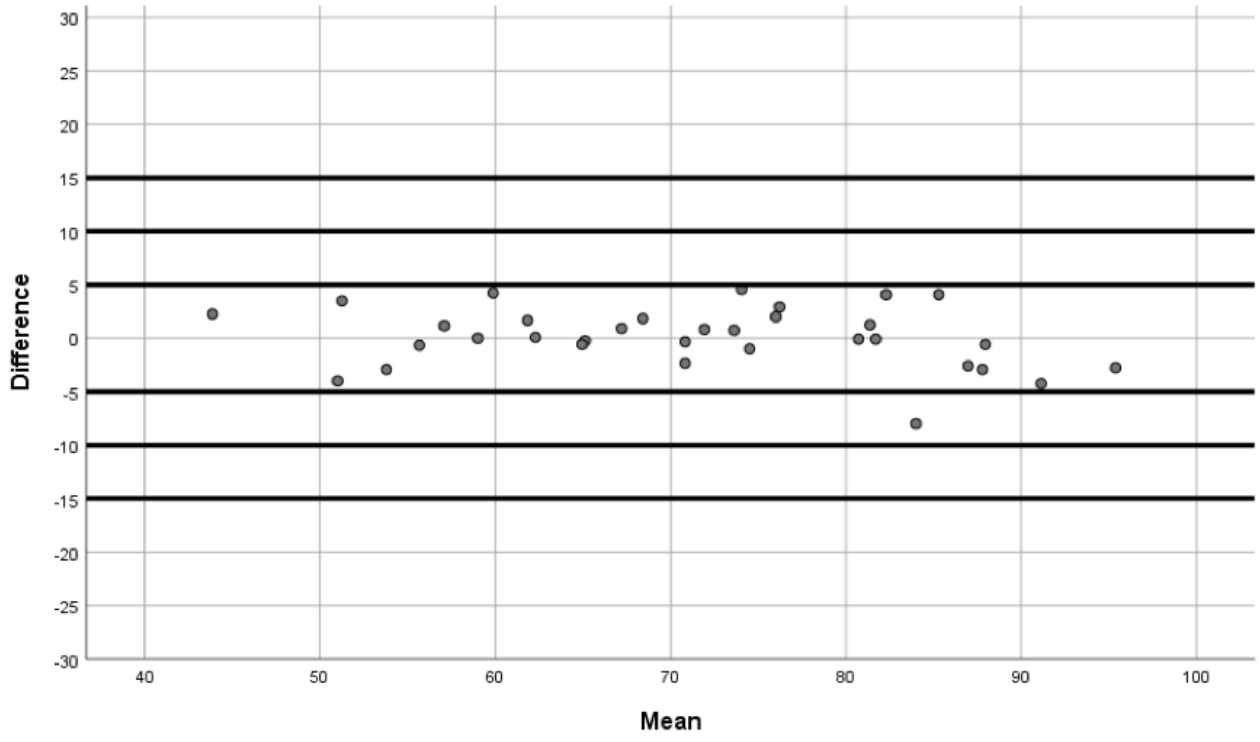


Figure 3. Bland-Altman graph for heart rate differences between the iHealth Track and the Omron M3.



Discussion

This study is the first to validate the accuracy of the iHealth Track device for BP and HR recordings in patients with CKD. The results indicate that the iHealth Track, as used in renal patients, passed the ESH-IP2 validation requirements. We previously validated the iHealth Track device for the general population following ESH-IP2 [29].

This study showed two limitations. Although the iHealth Track has been validated in the general population and now in patients with CKD, we cannot necessarily extrapolate our results to other specific populations. In addition, patients with CKD have stiffer arteries than other people [9,11,14]. We did not investigate arterial stiffness, but it would be useful to assess it in future validation studies.

The ESH-IP2 for home BP monitoring highlights the need for specific validation in patients with end-stage renal disease [14,30] as strict control of hypertension is required in these patients [9,14].

However, there are few devices that have been validated in patients with renal disease [26,31-34]. Akpolat et al [26] validated Omron M3 HEM-7051 in patients with CKD according to the ESH-IP2 revision. They used the mercury sphygmomanometer as their standard reference device and included 66 participants, rather than 33. The results were similar to ours, since both studies passed the ESH-IP revision's two phases of validation. However, the number of differences included in the category of 5 mmHg according to the ESH-IP2 were better for the iHealth Track for both SBP and DBP (ie, iHealth Track achieved higher differences than Omron). The differences obtained with the HRs cannot be contrasted as HR was not measured by Akpolat et al. Likewise, our findings cannot be compared with the rest of the validation studies found [31-34], since none followed the ESH-IP2 validation requirements. We believe that more validation studies for BP monitoring devices are necessary for patients with CKD.

The purpose of the ESH-IP2 was to simplify the previous protocols of the British Hypertension Society (BHS) [19] and the Association for the Advancement of Medical Instrumentation (AAMI) [20]. However, the protocol does have some shortcomings. First, the major limitation is that it is underpowered with only 33 participants (99 measures) required rather than the 85 participants (255 measures) required by the previous AAMI and BHS validation protocols [19,20]. Second, the ESH-IP2 does not indicate the number of validation studies needed to establish the accuracy of a device. According to some experts, at least two validation studies should be performed in different centers and in different populations before accepting the device as accurate [35]. Therefore, it is valuable to evaluate BP devices in diverse specific populations before they are used widely in clinics and homes. Third, the ESH-IP2 imposes certain gender requirements and limits validation studies to individuals older than 25 years who have BPs within specific ranges. Therefore, device accuracy remains unknown in children, adolescents, young adults, and patients with extreme BP values. Finally, the ESH-IP2 does not mention explicit criteria for validation in specific populations. Following the start of our study, in March 2019, the AAMI/ESH/ISO Universal Standard was published as the recommended standard for validation of BP measuring devices. [36]. This standard includes criteria for the validation of BP devices in specific populations. This will be considered in our future validations.

Our study is the first to show that the iHealth Track device meets the requirements of ESH-IP2 in patients with CKD. Future versions of the ESH-IP should include explicit criteria for validation in specific populations. Validation of this device would be valuable in other specific populations such as pregnant women, older adults, and patients with arrhythmias.

Acknowledgments

We thank the patients and medical staff from the Fresenius Medical Care clinic in Plasencia, especially Dr Alejandro Cives, nephrologist.

Conflicts of Interest

None declared.

References

1. Levin A, Stevens P, Bilous R, Coresh J, de Francisco ALM, de Jong PE, et al. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013 Jan. [doi: 10.1038/kisup.2012.73]
2. Zoccali C, Vanholder R, Massy ZA, Ortiz A, Sarafidis P, Dekker FW, et al. The systemic nature of CKD. Nat Rev Nephrol 2017 Jun;13(6):344-358. [doi: 10.1038/nrneph.2017.52] [Medline: 28435157]
3. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. The Lancet 2013 Jul;382(9888):260-272. [doi: 10.1016/S0140-6736(13)60687-X]
4. Stanifer JW, Kilonzo K, Wang D, Su G, Mao W, Zhang L, et al. Traditional medicines and kidney disease in low- and middle-income countries: opportunities and challenges. Semin Nephrol 2017 May;37(3):245-259. [doi: 10.1016/j.semnephrol.2017.02.005] [Medline: 28532554]
5. Eriksen BO, Stefansson VT, Jenssen TG, Mathisen UD, Schei J, Solbu MD, et al. Elevated blood pressure is not associated with accelerated glomerular filtration rate decline in the general non-diabetic middle-aged population. Kidney Int 2016 Aug;90(2):404-410. [doi: 10.1016/j.kint.2016.03.021] [Medline: 27188503]
6. Freedman BI, Cohen AH. Hypertension-attributed nephropathy: what's in a name? Nat Rev Nephrol 2016 Jan 10;12(1):27-36. [doi: 10.1038/nrneph.2015.172] [Medline: 26553514]
7. Muntner P, Anderson A, Charleston J, Chen Z, Ford V, Makos G, et al. Hypertension awareness, treatment, and control in adults with CKD: results from the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2010 Mar;55(3):441-451 [FREE Full text] [doi: 10.1053/j.ajkd.2009.09.014] [Medline: 19962808]
8. Agarwal R, Pappas MK, Sinha AD. Masked uncontrolled hypertension in CKD. JASN 2015 Jul 10;27(3):924-932. [doi: 10.1681/asn.2015030243]

9. Romagnani P, Remuzzi G, Glassock R, Levin A, Jager KJ, Tonelli M, et al. Chronic kidney disease. Nat Rev Dis Primers 2017 Nov 23;3:17088. [doi: 10.1038/nrdp.2017.88] [Medline: 29168475]
10. Flynn JT, Mitsnefes M, Pierce C, Cole SR, Parekh RS, Furth SL, et al. Blood pressure in children with chronic kidney disease: a report from the Chronic Kidney Disease in Children study. Hypertension 2008 Oct;52(4):631-637 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.108.110635] [Medline: 18725579]
11. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. J Hypertens 2018 Oct;36(10):1953-2041. [doi: 10.1097/HJH.0000000000001940] [Medline: 30234752]
12. Sim JJ, Shi J, Kovesdy CP, Kalantar-Zadeh K, Jacobsen SJ. Impact of achieved blood pressures on mortality risk and end-stage renal disease among a large, diverse hypertension population. J Am Coll Cardiol 2014 Aug 12;64(6):588-597 [FREE Full text] [doi: 10.1016/j.jacc.2014.04.065] [Medline: 25104529]
13. Cheung AK, Rahman M, Reboussin DM, Craven TE, Greene T, Kimmel PL, et al. Effects of intensive BP control in CKD. J Am Soc Nephrol 2017 Sep 22;28(9):2812-2823 [FREE Full text] [doi: 10.1681/ASN.2017020148] [Medline: 28642330]
14. Parati G, Ochoa JE, Bilo G, Agarwal R, Covic A, Dekker FW, et al. Hypertension in chronic kidney disease part 2. Hypertension 2016 Jun;67(6):1102-1110. [doi: 10.1161/hypertensionaha.115.06896]
15. Uhlig K, Patel K, Ip S, Kitsios GD, Balk EM. Self-measured blood pressure monitoring in the management of hypertension: a systematic review and meta-analysis. Ann Intern Med 2013 Aug 06;159(3):185-194. [doi: 10.7326/0003-4819-159-3-201308060-00008] [Medline: 23922064]
16. Sharman JE, Howes FS, Head GA, McGrath BP, Stowasser M, Schlaich M, et al. Home blood pressure monitoring: Australian Expert Consensus Statement. J Hypertens 2015 Sep;33(9):1721-1728 [FREE Full text] [doi: 10.1097/HJH.0000000000000673] [Medline: 26136205]
17. Imai Y, Kario K, Shimada K, Kawano Y, Hasebe N, Matsuura H, et al. The Japanese Society of Hypertension guidelines for self-monitoring of blood pressure at home (second edition). Hypertens Res 2012 Aug;35(8):777-795. [doi: 10.1038/hr.2012.56] [Medline: 22863910]
18. Mancia G, Fagard R, Narkiewicz K, Redón J, Zanchetti A, Böhm M, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). J Hypertens 2013 Jul;31(7):1281-1357. [doi: 10.1097/01.hjh.0000431740.32696.cc] [Medline: 23817082]
19. O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman DG, et al. An outline of the revised British Hypertension Society protocol for the evaluation of blood pressure measuring devices. J Hypertens 1993 Jun;11(6):677-679. [doi: 10.1097/00004872-199306000-00013] [Medline: 8397248]
20. ANSI/AAMI/ISO 81060- 2:2013.: Association For The Advancement Of Medical Instrumentation Non-invasive sphygmomanometers — Part 2: Clinical investigation of automated measurement type URL: http://my.aami.org/aamiresources/previewfiles/8106002_1306_preview.pdf
21. O'Brien E, Pickering T, Asmar R, Myers M, Parati G, Staessen J, et al. Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults. Blood Press Monit 2002 Feb;7(1):3-17. [doi: 10.1097/00126097-200202000-00002] [Medline: 12040236]
22. O'Brien E, Atkins N, Stergiou G, Karpettas N, Parati G, Asmar R, et al. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. Blood Press Monit 2010 Feb;15(1):23-38. [doi: 10.1097/MBP.0b013e3283360e98] [Medline: 20110786]
23. Stergiou GS, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, et al. A universal standard for the validation of blood pressure measuring devices. Journal of Hypertension 2018;36(3):472-478. [doi: 10.1097/hjh.0000000000001634]
24. Asociación Médica Mundial. 2013 Oct. Declaración de Helsinki de la Amm - principios éticos para las investigaciones médicas en seres humanos URL: https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/ [accessed 2019-03-04]
25. Asmar R, Khabouth J, Topouchian J, El Feghali R, Mattar J. Validation of three automatic devices for self-measurement of blood pressure according to the International Protocol: The Omron M3 Intellisense (HEM-7051-E), the Omron M2 Compact (HEM 7102-E), and the Omron R3-I Plus (HEM 6022-E). Blood Press Monit 2010 Feb;15(1):49-54. [doi: 10.1097/MBP.0b013e3283354b11] [Medline: 20032779]
26. Akpolat T, Erdem E, Aydogdu T. Validation of the Omron M3 Intellisense (HEM-7051-E) upper arm blood pressure monitor, for self-measurement, according to the European Society of Hypertension International Protocol revision 2010 in a stage 3-5 chronic kidney disease population. Kidney Blood Press Res 2012;35(2):82-88 [FREE Full text] [doi: 10.1159/000330719] [Medline: 21912183]
27. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]
28. Bland J, Altman D. Comparing methods of measurement: why plotting difference against standard method is misleading. The Lancet 1995 Oct;346(8982):1085-1087. [doi: 10.1016/s0140-6736(95)91748-9]

29. Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, López-López D, Palomo-López P, Rodríguez-Sanz D, et al. Validation in the general population of the iHealth track blood pressure monitor for self-measurement according to the European Society of Hypertension International Protocol Revision 2010: descriptive investigation. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e131377 [FREE Full text] [doi: 10.2196/13137] [Medline: 30888331]
30. Parati G, Stergiou GS, Asmar R, Bilo G, de Leeuw P, Imai Y, et al. European Society of Hypertension practice guidelines for home blood pressure monitoring. J Hum Hypertens 2010 Dec 3;24(12):779-785. [doi: 10.1038/jhh.2010.54] [Medline: 20520631]
31. Thompson AM, Eguchi K, Reznik ME, Shah SS, Pickering TG. Validation of an oscillometric home blood pressure monitor in an end-stage renal disease population and the effect of arterial stiffness on its accuracy. Blood Pressure Monitoring 2007;12(4):227-232. [doi: 10.1097/mbp.0b013e328108f544]
32. Peixoto AJ, Gray TA, Crowley ST. Validation of the SpaceLabs 90207 ambulatory blood pressure device for hemodialysis patients. Blood Pressure Monitoring 1999;4(5):217-221. [doi: 10.1097/00126097-199910000-00003]
33. Semret M, Zidehsarai M, Agarwal R. Accuracy of oscillometric blood pressure monitoring with concurrent auscultatory blood pressure in hemodialysis patients. Blood Press Monit 2005 Oct;10(5):249-255. [doi: 10.1097/01.mbp.0000172713.28029.84] [Medline: 16205443]
34. Czarkowski M, Staszków M, Kostyra K, Shebani Z, Niemczyk S, Matuszkiewicz-Rowińska J. Determining the accuracy of blood pressure measurement by the Omron HEM-907 before and after hemodialysis. Blood Press Monit 2009 Oct;14(5):232-238. [doi: 10.1097/mbp.0b013e328331d5b5] [Medline: 19938337]
35. Topouchian J, Agnoletti D, Blacher J, Youssef A, Ibanez I, Khabouth J, et al. Validation of four automatic devices for self-measurement of blood pressure according to the international protocol of the European Society of Hypertension. Vasc Health Risk Manag 2011;7:709-717 [FREE Full text] [doi: 10.2147/VHRM.S27193] [Medline: 22174581]
36. Stergiou GS, Palatini P, Asmar R, Ioannidis JP, Kollias A, Lacy P, et al. Recommendations and Practical Guidance for performing and reporting validation studies according to the Universal Standard for the validation of blood pressure measuring devices by the Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO). J Hypertens 2019 Mar;37(3):459-466. [doi: 10.1097/HJH.0000000000002039] [Medline: 30702492]

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
BHS: British Hypertension Society
BP: blood pressure
BPM: beats per minute
CKD: chronic kidney disorder
D: diastolic
ESH-IP2: Hypertension International in European Society protocol 2010
HR: heart rate
ISO: International Organization for Standardization
S: systolic

Edited by G Eysenbach; submitted 13.05.19; peer-reviewed by E Navarro Flores, I Sanz, H Beltrán Alacreu, S Omboni, S Melville, J Bayod; comments to author 28.09.19; revised version received 08.10.19; accepted 27.01.20; published 02.04.20.

Please cite as:
Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, López-López D, Calvo-Lobo C, Rodríguez-Sanz D, Martínez-Jiménez EM, Palomo-López P
An Automated Blood Pressure Display for Self-Measurement in Patients With Chronic Kidney Disease (iHealth Track): Device Validation Study
JMIR Mhealth Uhealth 2020;8(4):e14702
URL: https://mhealth.jmir.org/2020/4/e14702
doi:10.2196/14702
PMID:32238337

Original Paper

Behavioral Pain Assessment Implementation in Long-Term Care Using a Tablet App: Case Series and Quasi-Experimental Design

Mahnoor Zahid[1], BSc; Natasha L Gallant[1], MA; Thomas Hadjistavropoulos[1], PhD; Eleni Stroulia[2], PhD

[1]Department of Psychology and Centre on Aging and Health, University of Regina, Regina, SK, Canada

[2]Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Thomas Hadjistavropoulos, PhD
Department of Psychology and Centre on Aging and Health
University of Regina
3737 Wascana Parkway
Regina, SK, S4S 0A2
Canada
Phone: 1 306 585 4457
Email: hadjistt@uregina.ca

Abstract

Background: Pain is often underassessed and undertreated among long-term care (LTC) residents living with dementia. When used regularly, the Pain Assessment Checklist for Seniors With Limited Ability to Communicate (PACSLAC) scales have been shown to have beneficial effects on pain assessment and management practices and stress and burnout levels in frontline staff in LTC facilities. Such scales, however, are not utilized as often as recommended, which is likely to be related to additional record-keeping and tracking over time involved with their paper-and-pencil administration.

Objective: Using implementation science principles, we assessed the introduction of the PACSLAC-II scale by comparing two methods of administration—a newly developed tablet app version and the original paper-and-pencil version—with respect to the frequency of pain assessment and facility staff feedback.

Methods: Using a case series approach, we tracked pain-related quality indicators at baseline, implementation, and follow-up periods. A quasi-experimental design was used to evaluate the effect of the method of administration (ie, *paper-and-pencil only* [n=18], *tablet only* [n=12], *paper-and-pencil followed by tablet app* [n=31], and *tablet app followed by paper-and-pencil* [n=31]) on pain assessment frequency and frontline staff stress and burnout levels. Finally, semistructured interviews were conducted with frontline staff to obtain perspectives on each method of administration.

Results: The implementation effort resulted in a great increase in pain assessment frequency across 7 independent LTC units, although these increases were not maintained during the follow-up period. Frontline staff reported lower levels of workload in the *paper-and-pencil followed by tablet app condition* than those in the *paper-and-pencil only* (P<.001) and *tablet app followed by paper-and-pencil* (P<.001) conditions. Frontline staff also reported lower levels of workload in the *tablet-only* condition than those in the *paper-and-pencil only* condition (P=.05). Similarly, lower levels of emotional exhaustion were reported by frontline staff in the *paper-and-pencil followed by tablet app* condition than those in the *paper-and-pencil only* (P=.002) and *tablet app followed by paper-and-pencil* (P=.002) conditions. Finally, frontline staff reported higher levels of depersonalization in the *paper-and-pencil only* condition than those in the *tablet app only* (P=.008), *paper-and-pencil followed by tablet app* (P<.001), and *tablet app followed by paper-and-pencil* (P<.001) conditions. Furthermore, narrative data from individual interviews with frontline staff revealed a preference for the tablet app over the paper-and-pencil method of administration.

Conclusions: This study provides support for the use of either the tablet app or the paper-and-pencil version of the PACSLAC-II to improve pain-related quality indicators, but a reported preference for and lower levels of stress and burnout with the use of the tablet app method of administration suggests that the use of the tablet app may have more advantages compared with the paper-and-pencil method of administration.

(JMIR Mhealth Uhealth 2020;8(4):e17108) doi:10.2196/17108

KEYWORDS

pain measurement; long-term care; nursing; technology Alzheimer disease; mHealth

Introduction

Pain in Long-Term Care

Pain is very prevalent among older adults, and its prevalence is expected to increase as Canada's population continues to age [1]. Among older adults residing in long-term care (LTC) facilities, and especially among residents with cognitive impairments, pain is often underassessed and undertreated [2,3]. Although pain levels do not differ among older adults with and without cognitive impairment, those with cognitive impairments are less likely to report pain compared with their cognitively intact counterparts [4-6]. As a result, higher rates of pain persist among older adults with cognitive impairment because pain is not as readily addressed as it is for their cognitively intact counterparts.

Importantly, underassessed pain can have dire consequences for this population. Among residents with dementia, for example, pain can result in behavioral disturbances and, if the pain is not assessed properly, such disturbances may be misattributed to a psychiatric condition [7]. Thus, residents with dementia experiencing pain are frequently treated with psychotropic medications, such as benzodiazepines, rather than analgesic medication [8]. The increased use of benzodiazepines to treat behavioral disturbances because of undermanaged pain in this population can result in negative consequences such as an increased risk of falls [9].

Self-report pain assessments are typically considered a valid means for evaluating the subjective nature of pain [10,11]. However, as previously mentioned, assessment is more complicated when assessing pain in individuals with moderate-to-severe dementia who are often unable to accurately self-report their pain. Observational behavioral pain assessment checklists are, therefore, important tools that are used in such cases [11]. The Pain Assessment Checklist for Seniors With Limited Ability to Communicate (PACSLAC) scales are examples of evidence-based behavioral pain assessment checklists that can be used for residents with dementia [4,12]. The clinical usefulness of these scales is strongly supported by the research literature. For example, the PACSLAC, when used regularly, has been shown to improve the use of analgesic medication and, consequently, can lead to reduced pain levels observed by staff [13]. Frontline staff using the PACSLAC scales also reported reduced stress and burnout levels compared with those who did not use the PACSLAC. Finally, regular use of PACSLAC scales reduced the use of benzodiazepines, which could help address the problem of polypharmacy in LTC [11].

Although observational pain assessment tools such as the PACSLAC scales have been implemented in LTC facilities, they are not used as frequently as recommended (eg, so that each resident is assessed for pain using a standardized tool a minimum of once per week) [14]. Regular use of these tools can be hampered because of a variety of factors, including limited time, staffing issues, and extra workload associated with paper-and-pencil administration of the tool. The paper-and-pencil version of the PACSLAC-II, for example, involves manual addition of checklist scores. It is the nurses' responsibility to diarize pain scores over time for pain-related pattern identification because PACSLAC-II scores should always be considered in relation to previous scores [15]. In addition, tools such as the PACSLAC-II should be administered a minimum of once per week for each non-self-reporting resident and, if moderate-to-severe pain is identified, it should be administered once again within 24 hours, after treatment has been implemented [14]. These guidelines require frontline staff to complete a considerable amount of paperwork requiring proper charting.

Aims and Objectives

This study was aimed at comparing a newly developed tablet app version of the PACSLAC-II with the original paper-and-pencil version. All the potential complexities associated with the paper-and-pencil version of the PACSLAC-II, as noted above, were addressed in the development of the tablet app method of administration. The tablet app, for example, can automatically add up the scores for each administration. The tablet app can also graph pain assessment scores over time. Finally, the tablet app compiles all the pain assessments and reassessments into a PDF containing the time and date of administration, the name of the assessor, the total score on the PACSLAC-II, and all the items from the PACSLAC-II that were endorsed to result in that total score. This method of administration allows for easier charting and record-keeping.

The tablet app version of the PACSLAC-II would be considered a type of mobile health (mHealth). Although evidence for the efficacy of mHealth is still mixed, a recent meta-analysis demonstrated that mHealth interventions are more effective than comparable non-mHealth interventions in improving health outcomes with a small overall weighted effect size [16]. Moreover, mHealth has been shown to lead to improvements in symptoms, hospitalizations, and death for a variety of health conditions [17]. The use of mHealth to assist in the management of pain [18-21] and the provision of care for aging populations [22-24] is supported by recent literature. Thus, given the recent findings regarding mHealth, the development and evaluation of the tablet version of the PACSLAC-II are warranted.

The first objective of this study was to evaluate whether pain assessment frequency improved with the use of the tablet app compared with that for the paper-and-pencil method of administration of the PACSLAC-II. For this primary objective, a case series design—that is, a descriptive approach that follows groups exposed to the same intervention over a specified period [25]—was employed. Each of the 7 independent LTC units was therefore separately evaluated using this design. Using a quasi-experimental design, the second objective was to evaluate the impact of each method of administration of the PACSLAC-II on frontline staff stress and burnout levels with the use of self-report questionnaires. The third objective was to obtain the perspectives of frontline staff on each method of administration using semistructured individual interviews.

Methods

Participants

Participants who completed self-report questionnaires included nurses and care aides working at participating LTC units who were fluent in both written and verbal English. The original sample included 121 frontline staff with females comprising 91.7% (111/121) of the sample between the ages of 21 and 70 years (mean 42.21, SD 10.85). The staff reported up to 36 years of experience working in LTC (mean 10.31, SD 8.47) and comprised 27.3% (33/121) of nurses and 72.7% (88/121) of special care aides. Because of limitations regarding the availability of frontline staff during their shifts, only 92 of these participants completed the stress and burnout questionnaires following the implementation period. A total of 43 participants completed interviews that were used in our qualitative analysis. All participants provided informed consent before participating in this study. An additional consent form was used for the audio recording of individual interviews. Participants were provided with a Can $10 (US $7.54) gift card after completing a set of questionnaires. In addition, participants who completed an individual interview (n=43) were compensated with an additional Can $10 (US $7.54) gift card. Personal information and individual responses collected as part of the study were kept confidential. Approval for this study was granted by our institutional ethics review board. Quality indicators comprised anonymized data about pain assessment frequency (and timeliness on related intervention) from each participating LTC unit (see *Setting* section for more details).

Setting

A total of 7 independent LTC units took part in this study. Units A through E were located in a mid-sized metropolitan area. Units A through D belonged to the same facility, but they were treated as independent units because no overlap between the frontline staff was established. Unit E was a separate LTC facility. Units F and G were separate LTC facilities located in a rural area near the mid-sized metropolitan area. Units consisted of 32, 28, 39, 40, 90, 30, and 45 beds, respectively.

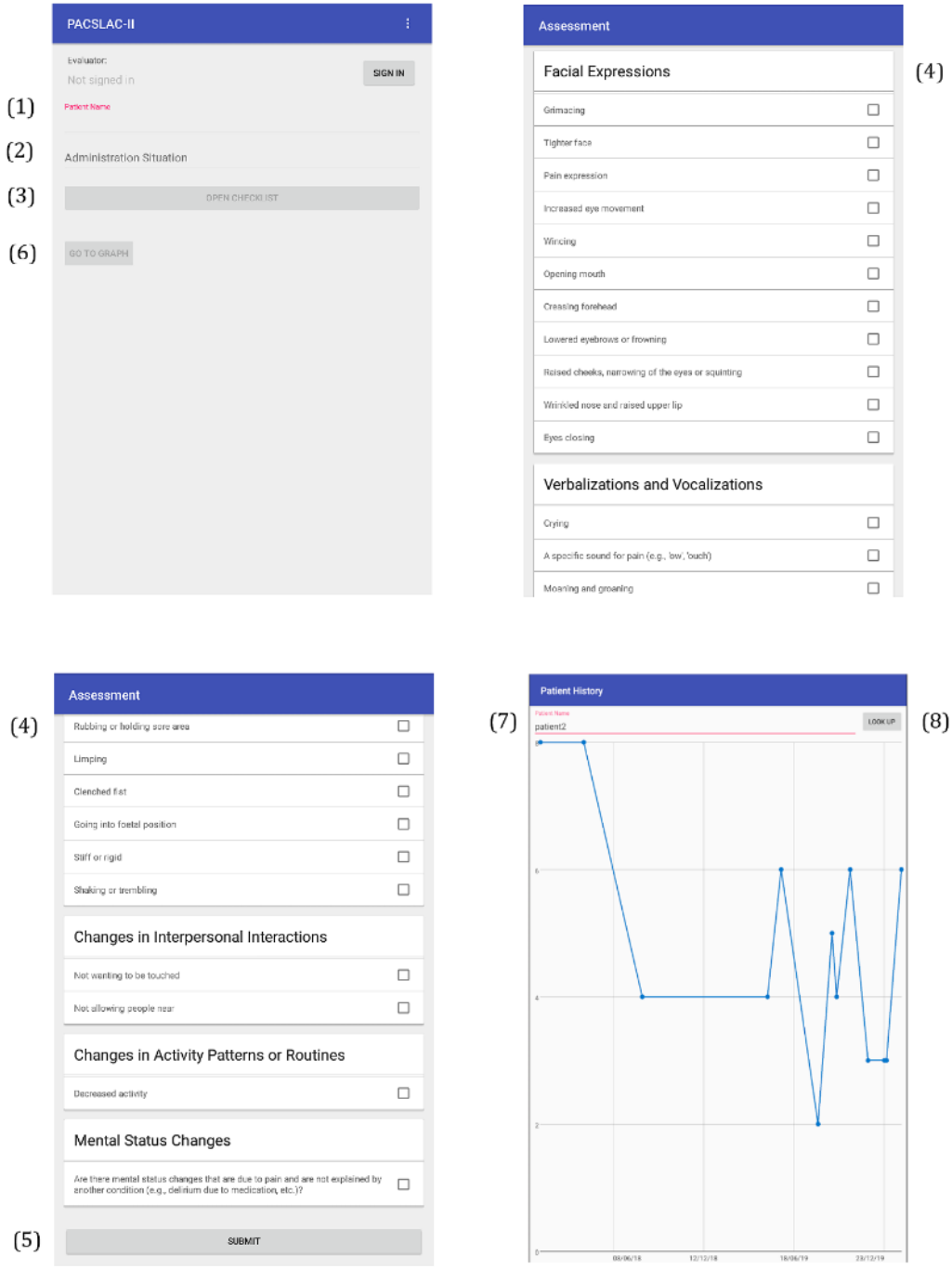
Materials

Tablet App Version of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate-II

For this study, we used Samsung Galaxy Tab A 7.0 devices with a 7-inch display screen and a weight of 283 g. The tablet app version of the PACSLAC-II was downloaded onto each of these tablets. A document providing instructions on how to use the PACSLAC-II app on the tablet was distributed to the staff of the participating LTC units. The tablet app version of the PACSLAC-II was meant to be a simple and literal interpretation of the paper-and-pencil version of the pain assessment tool. After signing in, the staff was presented with the same set of items in the same linear order and under the same headings as the paper-and-pencil version. Patient names and demographic identifiers were not entered into the app to protect confidentiality. Data collected from the app were encrypted through https to be transmitted and stored in a back-end repository that was secured behind a firewall.

To use the PACSLAC-II app, frontline staff would click on the PACSLAC-II icon in the apps folder on the tablet device's home page. Once the PACSLAC-II app was open, staff would sign in using a username and password that they had created for themselves. To complete a PACSLAC-II assessment (see Figure 1 for corresponding screenshots), staff would complete the following steps: (1) enter the resident's identification number and initials in the *patient name* box, (2) enter the location of administration in the *administration situation* box, (3) proceed to the PACSLAC-II observational checklist by clicking the *open checklist* button, (4) indicate the items by clicking the box adjacent to the applicable items for that resident, (5) click the *submit* button at the bottom of the PACSLAC-II observational checklist and to access the graph showing PACSLAC-II results over time, staff would (6) click the *go to graph* button on the bottom left of the screen, (7) enter the resident's identification number and initials in the *look up* box, and (8) click the *patient PDF* button.

Figure 1. Screenshots of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate (PACSLAC)-II tablet app with numbers corresponding with the instructions provided to participants. The screenshot on the bottom right of the figure shows the graph for a sample participant with the date of the PACSLAC-II administration on the horizontal axis and the PACSLAC-II score on the vertical axis.



Web-Based Training Program

In collaboration with an instructional designer and a Web developer, an interactive and dynamic Web-based training program pertaining to pain assessment in LTC with a focus on the use of the PACSLAC-II for persons with dementia was developed. The Web-based training program included 6 core modules that were each designed to be completed within 10 to 15 min. An additional optional module was created to provide staff with the opportunity to practice using the PACSLAC-II. The Web-based training program has been deemed to be useful by nursing staff and to increase the frequency of pain assessments in LTC when the paper-and-pencil method of administration was used [26].

Measures

Quality Indicators

Quality indicators measured each LTC unit's performance with regard to pain assessment frequency and follow-up for moderate-to-severe pain. These quality indicators were adapted from a consensus protocol for pain assessment and management developed following consultation with a group of public policy and geriatric pain experts [27]. This consensus protocol has been successfully implemented in LTC facilities [2]. The protocol incorporated the following quality indicators: (1) percentage of new residents who were assessed for pain with the PACSLAC-II on admission; (2) percentage of current residents assessed with the PACSLAC-II a minimum of once per week; (3) for residents with PACSLAC-II findings of moderate-to-severe pain, percentage of residents with a documented treatment plan within 24 hours of pain identification; (4) percentage of residents reassessed with a standardized pain assessment tool within 24 hours of treatment implementation; and (5) percentage of residents assessed for side effects within 24 hours of treatment implementation. Quality indicators were reported on a weekly basis by a pain champion from each LTC unit for the period of the study.

Demographic Questionnaire

Participants completed a demographic information sheet that included questions about their age, gender, years of experience working in LTC, and professional title (nurse or care aide).

Nursing Stress Scale

The Nursing Stress Scale (NSS) [28] is a 34-item scale comprising situations that could be perceived as stressful by frontline staff. Subscales include death and dying, conflict with supervisor, inadequate preparation, lack of support, conflict with coworker, workload, and uncertainty concerning treatment. Each item is rated on a 4-point Likert scale according to the frequency with which each item is felt to be stressful, that is, ranging from 0 (never) to 3 (very frequently). A higher total score indicates higher levels of perceived stress. For the purposes of this study, the original NSS was administered to nurses, and a modified version of the NSS was administered to care aides, that is, for items of the original NSS in which physicians were mentioned, the modified version of the NSS mentioned both physicians and nurses (eg, criticism by a physician in the original NSS was changed to criticism by a nurse or physician in the modified NSS). The NSS total score has demonstrated good test-retest reliability (ie, r=.81) and internal consistency (ie, Cronbach alpha=.89) [28]. In the same study, test-retest reliability was adequate for all subscales except for inadequate preparation (r=.42), lack of support (r=.65), and uncertainty concerning treatment (r=.68) [28]. Similarly, with regard to internal reliability, most subscales were adequately reliable, but conflict with supervisor, lack of support, and conflict with coworker had Cronbach alpha values of .68, .64, and .68 [28], respectively. In our study, internal consistency for total scores on the NSS was excellent (Cronbach alpha=.94). The death and dying, conflict with supervisor, inadequate preparation, lack of support, conflict with coworker, workload, and uncertainty concerning treatment subscales had Cronbach alpha values of .81, .67, .70, .73, .78, .83, and .79, respectively. All subscales, except for the conflict with supervisor subscale, had satisfactory reliability in our study. The reliability of the conflict with supervisor subscale was marginal.

Maslach Burnout Inventory

The Maslach Burnout Inventory (MBI) [29] is a 22-item scale comprising statements describing different experiences or situations encountered on the job for health care professionals. Subscales include emotional exhaustion, depersonalization, and personal accomplishment. Each item is rated on a 7-point Likert-type scale ranging from 0 (never) to 6 (every day) to assess the frequency of perceived burnout. A higher score is indicative of higher perceived burnout. With regard to test-retest reliability, correlations of each MBI subscales from the first and second administrations are between .53 and .69 [29]. The MBI has also demonstrated adequate internal consistency with a Cronbach alpha value of .84 for the total score and .86, .72, and .74, respectively, for emotional exhaustion, depersonalization, and personal accomplishment subscales [29]. In our study, internal consistency for total scores on the MBI was satisfactory (Cronbach alpha=.79). The emotional exhaustion, depersonalization, and personal accomplishment subscales had Cronbach alpha values of .89, .70, and .66, respectively. All subscales, except for the personal accomplishment subscale, had satisfactory reliability in our study. The internal consistency for personal accomplishment subscale was marginal.

Interviews

Semistructured interviews were conducted with the help of an interview guide designed specifically for this study (Textbox 1). The interview guide consisted of a series of questions regarding the attitudes toward pain assessment practices in older adults with dementia, barriers to effective pain management practices in LTC, and frontline staff experiences of using the tablet app version of the PACSLAC-II during the implementation period.

Textbox 1. Moderator guide for interviews with frontline staff.

Assessing pain among long-term care residents with dementia

- What were some of the pain assessment tools that you had previously used to assess pain among long-term care residents?
- How would you describe your workload and stress levels when assessing pain in long-term care residents with dementia?
- What role did you play in assessing pain among residents with dementia in your facility?
- In what ways do you believe pain to be adequately or inadequately addressed for residents with dementia?

Implementation of Pain Assessment Checklist for Seniors With Limited Ability to Communicate-II (PACSLAC-II)

- Why would you or why would you not consider using a tablet app version of the PACSLAC-II?
- In what ways would it or would it not be feasible to use the tablet app version of the PACSLAC-II in your facility?
- How did the tablet app/paper/both versions of the PACSLAC-II affect your workload levels?
- How would you describe your experience using the tablet app version of the PACSLAC-II?
- Would you prefer the paper or the tablet app version of the PACSLAC-II to assess pain?

Pain assessment practices

- What are particular barriers or challenges to changing or improving pain assessment practices within your facility?
- What are the supporting factors that assist in changing or improving pain assessment practices within your facility?
- How are decisions about changes in pain assessment practices made within your facility?
- What role do you play in the decisions made about pain assessment practices?

Procedure

Before the start of the study, a pain champion was appointed for each of the participating LTC units. This pain champion was responsible for overseeing the implementation of the standardized pain assessment protocol for the implementation period. This protocol comprised of the criteria assessed by the quality indicators (see *Quality Indicators* section in *Methods*). For this study, participants from each LTC unit took part in a baseline, training, implementation, and a follow-up period (Textbox 2).

Textbox 2. Description of the study's methodology.

Baseline (3 weeks)

- Quality indicators collected for each week.

Training (9 weeks)

- Web-based training program and/or in-person training.
- Training about the tablet app as needed.

Implementation (6 weeks)

- Assigned to use the tablet app, paper-and-pencil, or both methods of administration.
- Quality indicators collected for each week.
- Questionnaires completed at the end of the period.
- Interviews completed at the end of the period.

Follow-up (3 weeks)

- Quality indicators collected for each week.

Baseline Period

During the 3-week baseline period, participating LTC units collected the weekly quality indicators without making any changes to their pain assessment practices.

Training Period

Frontline staff in all participating LTC units were then trained on how to adequately assess pain using the PACSLAC-II and implement the standardized pain assessment protocol. A Web-based training program (see *Web-Based Training Program* section) in conjunction with in-person training was employed. The training was provided in both formats to ensure that we

were able to provide the necessary information to the staff who was going to use the PACSLAC-II as part of this study. Our research team was able to provide direct training to 89 frontline staff members (including 18 members who completed the Web-based training program). Moreover, the pain champion at each facility was available as a resource for all staff who had questions or uncertainties about the assessment process and the pain protocol. For those LTC units randomized to use the tablet app (see *Implementation Period* section), in-person training on how to use different features of the tablet app was provided by a member of the research team to all frontline staff and printed instructions on how to use the tablet app were distributed to these LTC units (see *Tablet App Version of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate-II* section).

Implementation Period

Following the training period, LTC units were randomized to use the paper-and-pencil, tablet app, or a combination of both methods of administration. Two tablets were used in the implementation period at each unit, and LTC units collected quality indicator data. The pain champion encouraged the use of the PACSLAC-II in accordance with the implementation protocol. The research team visited the LTC units on a regular basis to ensure that the implementation protocol was being adhered to throughout the implementation period. Tablets were made available to each unit during each of the tablet periods. During the 6-week implementation period, frontline staff completed the demographic information sheet as well as the MBI and NSS. Finally, following the implementation period, semistructured individual interviews were completed with interested frontline staff.

Follow-Up Period

During the 3-week follow-up period, participating LTC units collected the weekly quality indicators while continuing with the standardized pain assessment protocol. Regular visits and support from the research team did not take place during this period.

Results

Quality Indicators

For each quality indicator, percentages were averaged for both the baseline and implementation periods (Table 1). Consistent with a case series design [25], percentages were averaged separately for each participating LTC unit. Overall, as shown in Table 1, for units that had at least one admission during the implementation period, the average percentage of pain assessments with specialized assessment tools completed on admission was at 100% for all except one of the LTC units. The average percentage of current residents who underwent weekly pain assessments increased from the baseline to the implementation period. Improvements in treatment plan documentation and reassessment following the treatment plan implementation were found for all except 1 of the LTC units. Assessments of side effects following treatment plan implementation were not consistently observed across facilities. Furthermore, during the follow-up period, improvements in pain assessment practices identified during the implementation period were not maintained. Finally, an examination of the method of administration during the implementation period did not reveal apparent differences in averaged percentages for any of the quality indicators.

Table 1. Quality indicators from baseline, implementation, and follow-up periods averaged across the 3 weeks from each period.

Long-term care unit	QI[a] 1[b], % (n/N)	QI 2[c], % (n/N)	QI 3[d], % (n/N)	QI 4[e], % (n/N)	QI 5[f], % (n/N)
Unit A					
Baseline	—[g]	0 (0.00/31.67)[h]	73 (2.67/3.67)	9 (0.33/3.67)	0 (0.00/3.67)
Tablet app	—	100 (30.67/30.67)	100 (3.00/3.00)	100 (3.00/3.00)	0 (0.00/0.00)
Tablet app (continued)	100 (1.00/1.00)	100 (29.67/29.67)	100 (1.00/1.00)	100 (1.00/1.00)	0 (0.00/0.00)
Follow-up	0 (0.00/1.00)	2 (0.67/30.67)	100 (1.50/1.50)	100 (1.50/1.50)	67 (1.00/1.50)
Unit B					
Baseline	0 (0.00/1.50)	9 (2.33/25.67)	91 (3.33/3.67)	0 (0.00/3.67)	46 (1.67/3.67)
Paper-and-pencil	—	100 (18.00/18.00)	100 (1.00/1.00)	100 (1.00/1.00)	0 (0.00/1.00)
Tablet app	100 (1.00/1.00)	100 (19.33/19.33)	100 (1.00/1.00)	100 (1.00/1.00)	0 (0.00/1.00)
Follow-up	—	15 (3.33/22.00)	—	—	—
Unit C					
Baseline	—	0 (0.00/39.00)	100 (8.00/8.00)	4 (0.33/8.00)	8 (0.67/8.00)
Paper-and-pencil	—	100 (39.00/39.00)	100 (5.00/5.00)	100 (5.00/5.00)	0 (0.00/5.00)
Paper-and-pencil (continued)	—	100 (38.33/38.33)	100 (2.00/2.00)	100 (2.00/2.00)	0 (0.00/2.00)
Follow-up	—	0 (0.00/39.00)	100 (2.00/2.00)	100 (2.00/2.00)	0 (0.00/2.00)
Unit D					
Baseline	0 (0.00/1.00)	0 (0.00/38.33)	100 (1.50/1.50)	100 (1.50/1.50)	100 (1.50/1.50)
Paper-and-pencil	—	98 (38.33/39.00)	100 (1.00/1.00)	100 (1.00/1.00)	0 (0.00/1.00)
Tablet app	—	100 (38.33/38.33)	100 (1.50/1.50)	100 (1.50/1.50)	0 (0.00/1.50)
Follow-up	—	0 (0.00/39.00)	100 (1.00/1.00)	100 (1.00/1.00)	0 (0.00/1.00)
Unit E					
Baseline	0 (0.00/2.00)	5 (4.67/90.00)	93 (4.33/4.67)	14 (0.67/4.67)	0 (0.00/4.67)
Paper-and-pencil	0 (0.00/4.33)	65 (58.33/90.00)	100 (1.33/1.33)	50 (0.67/1.33)	100 (1.33/1.33)
Tablet app	0 (0.00/3.50)	10 (8.67/90.00)	—	—	—
Follow-up	0 (0.00/2.00)	8 (7.33/90.00)	—	—	—
Unit F					
Baseline	—	0 (0.00/30.00)	100 (2.00/2.00)	100 (2.00/2.00)	100 (2.00/2.00)
Tablet app	—	100 (30.00/30.00)	100 (4.00/4.00)	100 (4.00/4.00)	100 (4.00/4.00)
Paper-and-pencil	—	100 (30.00/30.00)	100 (4.33/4.33)	100 (4.33/4.33)	100 (4.33/4.33)
Follow-up	100 (1.00/1.00)	4 (1.33/30.00)	100 (1.67/1.67)	100 (1.67/1.67)	100 (1.67/1.67)
Unit G					
Baseline	100 (1.00/1.00)	0 (0.00/45.00)	84 (9.00/10.67)	0 (0.00/10.67)	69 (7.33/10.67)
Paper-and-pencil	—	51 (23.00/45.00)	8 (0.67/8.33)	4 (0.33/8.33)	0 (0.00/8.33)
Paper-and-pencil (continued)	100 (1.00/1.00)	29 (13.00/45.00)	0 (0.00/3.00)	0 (0.00/3.00)	0 (0.00/3.00)
Follow-up	—	16 (7.33/45.00)	30 (1.00/3.33)	10 (0.33/3.33)	10 (0.33/3.33)

[a]QI: quality indicator.

[b]Number and percentage of new residents assessed using the PACSLAC-II on admission averaged across the total number of weeks from each period.

[c]Number and percentage of current residents assessed using the PACSLAC-II at least once per week averaged across the total number of weeks from each period.

[d]Number and percentage of residents with moderate-to-severe pain with a treatment plan within 24 hours averaged across the total number of weeks from each period.

[e]Number and percentage of residents with moderate-to-severe pain with a reassessment within 24 hours averaged across the total number of weeks from

each period.

[f]Number and percentage of residents with moderate-to-severe pain assessed for side effects within 24 hours averaged across the total number of weeks from each period.

[g]The symbol "—" indicates that there were no new admissions or no residents with moderate-to-severe pain during for that period.

[h]The n and N values in this table have decimals as they represent total numbers collected on a weekly basis averaged over a period of 3 weeks.

Questionnaires

Descriptive statistics and frequencies for demographic characteristics for each participating LTC unit are reported in Table 2.

The means and standard deviations for stress and burnout subscale scores for each condition are presented in Tables 3 and 4, respectively. Before conducting our principal analyses, we examined the relationship between the demographic characteristics of experience, age, gender, and professional title (ie, nurses or care aide) and subscale scores on stress and burnout measures using Pearson correlations. Of all the correlations, age was significantly related to conflict with physicians subscale (r=−.22, P=.04), and job title was significantly related to death and dying (r=.25, P=.02), conflict with physicians (r<.29, P=.006), and workload (r=.28, P=.008) subscales, with nurses scoring higher than care aides. Therefore, these demographic characteristics were included as covariates in subsequent analyses involving the corresponding subscale scores. To test the significance of mean differences among conditions, a series of analyses of variance (ANOVA) and covariance (ANCOVA) with conditions (ie, *paper-and-pencil only*, *paper-and-pencil followed by tablet app*, *tablet app followed by paper-and-pencil*, and *tablet app only*) as the between-subjects factor were conducted. The first series of ANOVAs and ANCOVAs examined stress subscale scores (ie, death and dying, conflict with physicians, inadequate preparation, lack of staff support, conflict with other nurses, workload, and uncertainty concerning treatment). The second series of ANOVAs examined burnout subscale scores (ie, emotional exhaustion, depersonalization, and personal achievement) at the end of the implementation period. It was expected that stress and burnout levels would be lower in conditions in which frontline staff had most recently used the tablet app method of administration rather than the paper-and-pencil method of administration.

With regard to stress subscale scores, the ANCOVA for workload (with nurse vs care aide as covariates) revealed a significant between-subjects effect ($F_{3,86}$=8.29, P<.001, partial η^2=0.22). Follow-up pairwise comparisons showed that frontline staff in the *paper-and-pencil followed by tablet app* condition reported significantly lower levels of workload compared with the *paper-and-pencil only* (P<.001) and *tablet app followed by paper-and-pencil* (P<.001) conditions. Frontline staff also reported lower levels of workload in the *tablet only* condition compared with the *paper-and-pencil only* condition (P=.05). No other differences for workload were found. The ANOVA and ANCOVA analyses for death and dying, conflict with physicians, inadequate preparation, lack of staff support, conflict with other nurses, and uncertainty concerning treatment did not show any significant effects.

With regard to burnout subscale scores, the ANOVA for emotional exhaustion revealed a significant between-subjects effect ($F_{3,88}$=4.86, P=.004, partial η^2=0.14). Follow-up pairwise comparisons showed that frontline staff in the *paper-and-pencil followed by tablet app* condition reported significantly lower levels of emotional exhaustion compared with the *paper-and-pencil only* (P=.002) and *tablet app followed by paper-and-pencil* (P=.002) conditions. No other significant differences in emotional exhaustion were found. The ANOVA for depersonalization revealed a significant between-subjects effect ($F_{3,88}$=9.54, P<.001, partial η^2=0.25). Follow-up pairwise comparisons showed that frontline staff in the *paper-and-pencil only* condition reported significantly higher levels of depersonalization compared with the *tablet app only* (P=.008), *paper-and-pencil followed by tablet app* (P<.001), and *tablet app followed by paper-and-pencil* (P<.001) conditions. No other significant differences in depersonalization were found. The ANOVA for personal achievement did not show any significant effects.

Table 2. Demographic characteristics for each participating long-term care unit.

Long-term care unit	Participants, N	Experience (years), mean (SD)	Age (years), mean (SD)	Gender (female), n (%)	Title (nurses), n (%)
Unit A	16	10.25 (6.73)	44.81 (11.35)	12 (75)	3 (19)
Unit B	21	10.52 (8.00)	40.70 (9.68)	18 (86)	6 (29)
Unit C	16	10.38 (7.86)	39.60 (9.60)	16 (100)	5 (31)
Unit D	20	8.35 (7.96)	41.26 (10.07)	18 (90)	8 (40)
Unit E	24	9.08 (8.45)	44.08 (12.07)	23 (96)	7 (29)
Unit F	17	13.65 (10.53)	43.31 (10.06)	17 (100)	3 (18)
Unit G	7	11.43 (11.47)	39.71 (16.01)	7 (100)	1 (14)

Table 3. Stress subscale scores for each condition.

Condition	Count, n	Stress subscale scores, mean (SD)						
		Death and dying	Conflict with supervisor	Inadequate preparation	Lack of support	Conflict with coworker	Workload	Uncertainty concerning treatment
Paper-and-pencil only	18	7.44 (3.57)	4.53 (2.65)	3.39 (2.00)	1.72 (2.05)	4.50 (3.15)	8.82 (4.45)	4.11 (2.78)
Tablet app only	12	8.33 (2.96)	4.42 (2.58)	3.08 (0.79)	2.25 (1.49)	4.08 (1.83)	6.17 (2.52)	4.33 (2.46)
Paper-and-pencil followed by tablet app	31	6.55 (2.59)	3.21 (1.66)	2.42 (1.06)	1.84 (1.32)	2.97 (2.17)	4.87 (2.59)	2.87 (2.16)
Tablet app followed by paper-and-pencil	31	6.87 (3.67)	3.60 (2.54)	2.65 (1.38)	2.48 (1.96)	4.45 (3.29)	8.48 (4.05)	4.26 (2.80)

Table 4. Burnout subscale scores for each condition.

Condition	Count, n	Burnout subscale scores, mean (SD)		
		Emotional exhaustion	Depersonalization	Personal accomplishment
Paper-and-pencil only	18	24.22 (11.47)	9.50 (6.51)	36.44 (7.32)
Tablet app only	12	18.50 (9.41)	5.17 (4.65)	35.92 (6.50)
Paper-and-pencil followed by tablet app	31	14.48 (7.82)	2.94 (3.12)	39.07 (5.09)
Tablet app followed by paper-and-pencil	31	23.23 (12.49)	4.10 (3.43)	38.81 (7.00)

Interviews

We interviewed a total of 43 participants. Saturation was achieved, given that we observed that the addition of the last several interviews was not yielding new information. Responses from interviews were analyzed using thematic content analysis [30]. By identifying common themes in the data set, researchers can capture and make sense of the meanings and experiences of participants in relation to the questions being asked. In accordance with the outline provided by Braun and Clarke [30], the primary researcher began the data analysis by transcribing all the interviews. Because a deductive approach was taken, the themes were constructed around the 3 main content areas of the interview guide (ie, user-friendliness, feasibility, and overall impression of the PACSLAC-II). Slight variations in main content areas across the LTC units, however, were observed. That is, LTC units randomized to use both versions of the PACSLAC-II were asked about all of the main content areas, LTC units randomized to use the tablet app version of the PACSLAC-II were asked about all of the main content areas with a focus on user-friendliness, and LTC units randomized to use the paper-and-pencil version of the PACSLAC-II were asked about all of the main content areas except for user-friendliness. Following the identification of the main content areas, the researcher completed constant reading, rereading, and highlighting of important and reoccurring aspects of the transcripts to become more familiar with the textual data for proper coding. Codes (ie, units of text that represents a meaningful idea relating to the identified content areas) were identified using NVivo (QSR International, Melbourne, Australia). The identified codes from the text were then reviewed and organized into themes. A second coder then organized the narrative responses of 21% (9/43) of the participants into these thematic categories to ensure consistency of the coding and trustworthiness of themes. Disagreements were resolved by consensus or if consensus could not be reached, reorganization of thematic categories. Representative quotes from each theme are shown in Textbox 3.

Textbox 3. Quotes representative of each identified theme.

User-friendliness

Benefits of the visual features

- "It was nice because you can go back and see the ups and downs from each of the graphs. It was nice, it was visual. Whereas paper, it's 10/10 pain or 7/10 pain and it's like where are you at. But with the graph, you could see, it's visual."
- "I think a visual graph when you can see it, and you see the peak times, and from there you can implement probably a medication that would work better or start something and see."

Faster and easier access of data

- "Yes, with tablet I think once it's filled out it's collated, cause with paper form, I have to do the photocopy and make sure that we have another copy in the chart and then the follow-up, right, and within 24 hours, you have to do another PACSLAC if they are in pain."
- "When using paper, you have to write, and it would take time… But for the tablet, all you have to do is tap. It's much faster… It's time-wise when using tablet."

Preference for the tablet app version of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate-II (PACSLAC-II)

- "Myself, tablet version because, like I said, we can go back and see the graphs and how they are responding and not responding, where the pain lies, do they need something extra. Paper version, definitely if something like that goes out, we have the option of using the paper always, that's always nice to have to keep track. Just for me, visually, I like using the tablet way better."
- "For me, it would be better if we had the tablet version. Firstly, it saves money. Imagine printing. But if you have it on tablet, you can just click, click. Storing is better than having the paper. It would appear cheaper in the long run."

Feasibility

Barriers associated with the tablet app version of the PACSLAC-II

- "Sometimes the tablet is not working because of the connection but it's easier to use the tablet…"
- "I think the tablet version is a great idea. I think that if you were going to implement it, it would just be getting its own Wi-Fi connection because obviously we struggle with the Wi-Fi connection."

Barriers associated with changes in pain assessment practices

- "Like I said, we look after these people, we're here more often with these people than we are with our own families. So, we know these people inside and out and so when we say that there's an issue or this person's off or they look like they're having a lot more pain, trust us….the doctor's only here once a week and he spends not very much time with these people and he comes in and he does his two minute assessment and says, 'they look fine today, no let's hold off.' Really, now we have to go another seven whole days of more documentation for him to say, 'well, I really don't know, we'll bump them up a little bit.' So, you know what I'm saying, it's the frustration of not being heard."
- "I think we should try to get everyone on the same page for care and management because it just seems that some nurses will do certain things and others won't."

Overall impression of the PACSLAC-II

No significant increase in workload levels with implementation of PACSLAC-II

- "It means more paperwork, but it's the paperwork that will help us in the end, and it doesn't take too long, maybe three or four minutes."
- "It honestly didn't take them that long. It was very quick, and I don't think it was that time-consuming."

Overall positive impression of the PACSLAC-II

- "I guess, overall experience using the PACSLAC, I think it was a win-win."
- "I think it kind of gets everybody on the same page, you know, we can discuss it afterwards and the continuing care aides would talk about it too and it kind of makes you think."

User-Friendliness

As shown in Textbox 3, 3 themes emerged pertaining to the user-friendliness of the tablet app version of the PACSLAC-II. The first theme focused on the benefits of the visual features of the tablet app version. That is, frontline staff spoke of the value of being able to identify visual patterns in pain scores over time using a graph instead of numbers. They noted that this feature was helpful in terms of pain assessment and pain management. The second theme related to faster and easier access to data; the consensus among the frontline staff was that the paper-and-pencil version of the PACSLAC-II resulted in extra paperwork that affected routine pain assessment practices. In contrast, the tablet app version provided a faster and easier way to store and access data. The third theme centered around a preference for the tablet app version of the PACSLAC-II. Most frontline staff preferred the tablet app version, although some

of the frontline staff noted hesitations about the tablet app version related to connectivity issues.

Feasibility

Regarding feasibility, 2 themes were recognized (Textbox 3). The first theme identified related to barriers associated with the tablet app version of the PACSLAC-II and the second theme related to barriers associated with changes in pain assessment practices. The primary barrier associated with the use of the tablet app version was connectivity. The tablet was connected to a wireless internet connection that was, occasionally, unreliable. Thus, significant challenges in consistently using the tablet app version when needed were noted.

Given that pain assessments were only required a minimum of once a week, we were advised that staff completed the pain assessments the following day if the technology did not work as planned. In the very small number of instances that there was a pressing need for an immediate pain assessment and the app was not working (eg, because of a connectivity problem), the PACSLAC-II was completed on paper, and the information was entered into the app the following day.

The primary barrier associated with changes in pain assessment practices identified by frontline staff was miscommunication across disciplines (ie, physicians, nurses, and care aides). Specifically, care aides noted that they sometimes felt dismissed when they would report to a nurse that one of the LTC residents was in pain. Another challenge was having physicians agree with nurses on the need for pain relief for LTC residents.

Overall Impression of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate-II

The overall impression of frontline staff of the implementation period was another main area of focus for the interviews, and 2 main themes emerged that pertained to the overall impression of the frontline staff (Textbox 3). With regard to the first theme, the majority of frontline staff reported that they did not notice a significant increase in workload levels and that, even if an increase in workload were to occur, that it would be worthwhile. However, some frontline staff—most notably the nurse manager or pain champion—indicated additional workload stemming from paperwork and management of the implementation because of their unit's participation in the research aspects of the study and the tracking of the quality indicators. With regard to the second theme, the majority of frontline staff reported that their experience with using the PACSLAC-II during the implementation period was positive. Of note, some participants noted that the use of the PACSLAC-II provided a common language for staff across disciplines to talk about pain.

Discussion

Principal Findings

The implementation of the PACSLAC-II in this study resulted in improvements in pain assessment frequency regardless of the medium used for the PACSLAC-II. Importantly, the percentage of new residents assessed on admission and current residents assessed on a weekly basis increased to 100% for all except 1 participating LTC unit regardless of the medium of administration that was employed. The tablet app version of the PACSLAC-II was also associated with lower levels of reported stress and burnout among staff. For example, frontline staff who were in the *paper-and-pencil only* and *tablet app to paper-and-pencil* conditions experienced significantly higher levels of emotional exhaustion and workload compared with frontline staff who were in the *paper-and-pencil to tablet app* condition. Perhaps the benefit of a reduced workload because of the tablet (when it was preceded by paper and pencil administrations) became more evident and was better appreciated by those who saw the change by experiencing the paper-and-pencil version first. Similarly, those who started with the tablet version would have reported a greater workload following the paper and pencil version because of the contrast between the two conditions. Unsurprisingly, frontline staff in the *tablet app only* condition also reported lower levels of workload compared with their counterparts in the *paper-and-pencil only* condition. Furthermore, those in the *paper-and-pencil only* condition experienced significantly higher levels of depersonalization than all other conditions.

Finally, frontline staff responses during interviews suggested an overall preference for the tablet app version. This overall preference is a very important theme because, through the identification of preferences, we can recognize the features of the tablet app that prove to be beneficial in improving pain assessment practices. It also shows potential for the PACSLAC-II to be utilized on a more regular basis if the tablet app version is employed by LTC facilities. We are hoping to incorporate additional features in a future version of the app that could automatize the tracking of quality indicators (eg, percentage of residents assessed at least once a week), facilitating the monitoring of quality improvement programs without extra burden for staff. Another possible feature would be to incorporate the capability for generating reminders for staff about which residents are due for an assessment.

Based on the qualitative analysis, we were very encouraged by staff who observed that the PACSLAC-II gave them a common language to discuss pain across disciplines. Diarized PACSLAC-II scores could also assist physicians who visit the facility intermittently and who would otherwise not have an ongoing index of the patients' day-to-day pain.

Limitations and Future Directions

Improvements in pain assessment frequency were not maintained during the follow-up period. Improvement in aspects of treatment follow-up for residents with moderate-to-severe pain (ie, treatment plan implemented within 24 hours, reassessed for pain within 24 hours of the treatment plan, and assessment for side effects of treatment) was also not consistently found across LTC facilities. Also, differences in pain assessment frequency and treatment follow-up as a function of the tablet app and paper-and-pencil versions of the PACSLAC-II were not identified. Furthermore, although an overall preference for the tablet app was identified, one issue that would need to be resolved to increase the uptake of this method of administration is improved connectivity.

Several directions regarding the standardized pain assessment protocol should be considered for future research. This study

was carried out over a 12-week period. Future research should focus on the study of the different modes of administration of the PACSLAC-II over the longer term. To ensure the feasibility of integrating the use of the tablet app version in LTC facilities, future research could aim for longer implementation and follow-up periods, especially given the trend of decreased pain assessment frequency and follow-up treatment during the follow-up periods in this study. Furthermore, many of the residents who experienced moderate-to-severe pain did not have documentation in their charts to indicate whether or not they were checked for side effects 24 hours following treatment. Moreover, in this study, we measured whether a treatment plan was implemented when moderate-to-severe pain was identified, but we did not measure the types of treatment plans that were implemented (eg, medications or repositioning). Research into these aspects of treatment follow-up should be explored in future studies.

Other avenues for future research may include the addition and study of new PACSLAC-II app features. We plan to use lessons learned from this study as we develop the next version of our app. Such features could include automatic detection of significant changes in residents' regular pattern of scores, access to educational resources and tools, and ability to track average facility-wide pain levels. This latter feature could allow for the use of facility-wide PACSLAC-II scores as a quality indicator with respect to progress in managing pain. Moreover, although the tablet app method of administration was meant to be a simple and literal implementation of the paper-and-pencil method of administration, the inclusion of a user experience/interface design could have nevertheless resulted in a tablet app version of the PACSLAC-II that was better designed for the user.

In this study, many of our participants were trained in pain assessment and use of the PACSLAC-II through a Web-based training program that we had previously validated using a paper and pencil version of the app. We are planning to integrate specific app training in a future version of our training program and then study the contribution of the app-specific component on app use and benefits. Another area for future research would be the development of a protocol for steps to take when technical issues arise.

Conclusions

Given that observational pain assessment tools are not utilized as frequently in LTC facilities as they should be, it is important to introduce innovative ways in which pain assessment practices could be improved. One such innovation is the use of technology to facilitate regular pain assessments. Findings from this study support the use of either version of the PACSLAC-II when implementing a standardized pain assessment protocol as both versions were found to improve pain assessment frequency and some aspects of treatment follow-up. However, the tablet app version of the PACSLAC-II was preferred by frontline staff and seemed to have resulted in lower levels of reported emotional exhaustion, depersonalization, and workload. More research regarding the use of technological innovations for pain assessments in LTC facilities, therefore, has the potential to lead to improved pain assessments, and an associated increase in quality of life, for LTC residents.

Acknowledgments

This work was supported, in part, through funding from the AGE-WELL Networks of Centers of Excellence.

Authors' Contributions

TH played a major role in the write up of the grant application that funded the study and its conceptualization. He also guided and supervised every aspect of this investigation, including a major contribution to the write up and interpretation of the results. MZ assisted with the development of the data collection protocol, data collection, data analysis/interpretation, write up, and literature review. NG played a major role in data analysis, interpretation, and write up. ES played a major role in the development of the tablet app and the description of this tablet app in the write up. All authors discussed the results and read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Reid MC, Eccleston C, Pillemer K. Management of chronic pain in older adults. Br Med J 2015 Feb 13;350:h532 [FREE Full text] [doi: 10.1136/bmj.h532] [Medline: 25680884]
2. Hadjistavropoulos T, Williams J, Kaasalainen S, Hunter PV, Savoie ML, Wickson-Griffiths A. Increasing the frequency and timeliness of pain assessment and management in long-term care: knowledge transfer and sustained implementation. Pain Res Manag 2016;2016:6493463 [FREE Full text] [doi: 10.1155/2016/6493463] [Medline: 27445619]
3. Takai Y, Yamamoto-Mitani N, Okamoto Y, Koyama K, Honda A. Literature review of pain prevalence among older residents of nursing homes. Pain Manag Nurs 2010 Dec;11(4):209-223. [doi: 10.1016/j.pmn.2010.08.006] [Medline: 21095596]
4. Chan S, Hadjistavropoulos T, Williams J, Lints-Martindale A. Evidence-based development and initial validation of the pain assessment checklist for seniors with limited ability to communicate-II (PACSLAC-II). Clin J Pain 2014 Sep;30(9):816-824. [doi: 10.1097/AJP.0000000000000039] [Medline: 24281294]

5. Kunz M, Mylius V, Scharmann S, Schepelman K, Lautenbacher S. Influence of dementia on multiple components of pain. Eur J Pain 2009 Mar;13(3):317-325. [doi: 10.1016/j.ejpain.2008.05.001] [Medline: 18562225]
6. Zwakhalen SM, Hamers JP, Abu-Saad HH, Berger MP. Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. BMC Geriatr 2006 Jan 27;6:3 [FREE Full text] [doi: 10.1186/1471-2318-6-3] [Medline: 16441889]
7. Cipher DJ, Clifford PA, Roper KD. Behavioral manifestations of pain in the demented elderly. J Am Med Dir Assoc 2006 Jul;7(6):355-365. [doi: 10.1016/j.jamda.2005.11.012] [Medline: 16843236]
8. Balfour JE, O'Rourke N. Older adults with Alzheimer disease, comorbid arthritis and prescription of psychotropic medications. Pain Res Manag 2003;8(4):198-204. [doi: 10.1155/2003/105459] [Medline: 14679414]
9. Woolcott JC, Richardson KJ, Wiens MO, Patel B, Marin J, Khan KM, et al. Meta-analysis of the impact of 9 medication classes on falls in elderly persons. Arch Intern Med 2009 Nov 23;169(21):1952-1960. [doi: 10.1001/archinternmed.2009.357] [Medline: 19933955]
10. Catananti C, Gambassi G. Pain assessment in the elderly. Surg Oncol 2010 Sep;19(3):140-148. [doi: 10.1016/j.suronc.2009.11.010] [Medline: 20006488]
11. Hadjistavropoulos T, Herr K, Prkachin KM, Craig KD, Gibson SJ, Lukas A, et al. Pain assessment in elderly adults with dementia. Lancet Neurol 2014 Dec;13(12):1216-1227. [doi: 10.1016/S1474-4422(14)70103-6] [Medline: 25453461]
12. Fuchs-Lacelle S, Hadjistavropoulos T. Development and preliminary validation of the pain assessment checklist for seniors with limited ability to communicate (PACSLAC). Pain Manag Nurs 2004 Mar;5(1):37-49. [doi: 10.1016/j.pmn.2003.10.001] [Medline: 14999652]
13. Fuchs-Lacelle S, Hadjistavropoulos T, Lix L. Pain assessment as intervention: a study of older adults with severe dementia. Clin J Pain 2008 Oct;24(8):697-707. [doi: 10.1097/AJP.0b013e318172625a] [Medline: 18806535]
14. Hadjistavropoulos T, Marchildon GP, Fine PG, Herr K, Palley HA, Kaasalainen S, et al. Transforming long-term care pain management in north america: the policy-clinical interface. Pain Med 2009 Apr;10(3):506-520. [doi: 10.1111/j.1526-4637.2009.00566.x] [Medline: 19254336]
15. Hadjistavropoulos T, Fitzgerald TD, Marchildon GP. Practice guidelines for assessing pain in older persons with dementia residing in long-term care facilities. Physiother Can 2010;62(2):104-113 [FREE Full text] [doi: 10.3138/physio.62.2.104] [Medline: 21359040]
16. Yang Q, van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
17. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
18. Lee RR, Rashid A, Ghio D, Thomson W, Cordingley L. 'Seeing pain differently': a qualitative investigation into the differences and similarities of pain and rheumatology specialists' interpretation of multidimensional mobile health pain data from children and young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2019 Jul 2;7(7):e12952 [FREE Full text] [doi: 10.2196/12952] [Medline: 31267979]
19. Shaballout N, Aloumar A, Neubert T, Dusch M, Beissner F. Digital pain drawings can improve doctors' understanding of acute pain patients: survey and pain drawing analysis. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e11412 [FREE Full text] [doi: 10.2196/11412] [Medline: 30632970]
20. Shaballout N, Neubert T, Boudreau S, Beissner F. From paper to digital applications of the pain drawing: systematic review of methodological milestones. JMIR Mhealth Uhealth 2019 Sep 5;7(9):e14569 [FREE Full text] [doi: 10.2196/14569] [Medline: 31489841]
21. Sun Y, Jiang F, Gu JJ, Wang YK, Hua H, Li J, et al. Development and testing of an intelligent pain management system (IPMS) on mobile phones through a randomized trial among chinese cancer patients: a new approach in cancer pain management. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e108 [FREE Full text] [doi: 10.2196/mhealth.7178] [Medline: 28743681]
22. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 9;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
23. Ferreira PH, Maximiano M, Reis CI. Care4Value: A Clinical Instrument Data Collection Platform for Long-Term Healthcare Units. In: Proceedings of the 2019 IEEE 6th Portuguese Meeting on Bioengineering. 2019 Presented at: ENBENG'19; February 22-23, 2019; Lisbon, Portugal URL: https://ieeexplore.ieee.org/document/8692494 [doi: 10.1109/enbeng.2019.8692494]
24. van Maurik IS, Visser LN, Pel-Littel RE, van Buchem MM, Zwan MD, Kunneman M, et al. Development and usability of ADappt: Web-based tool to support clinicians, patients, and caregivers in the diagnosis of mild cognitive impairment and alzheimer disease. JMIR Form Res 2019 Jul 8;3(3):e13417 [FREE Full text] [doi: 10.2196/13417] [Medline: 31287061]
25. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. J Bone Joint Surg Am 2009 May;91(Suppl 3):21-26. [doi: 10.2106/JBJS.H.01573] [Medline: 19411496]

26. Gallant NL. An Online Pain Assessment/Management Training Program for Long-Term Care Staff. In: Proceedings of AGE-WELL's 2nd Annual Conference. 2017 Presented at: AGE-WELL'17; October 18-19, 2017; Winnipeg, Manitoba URL: http://agewell-nce.ca/wp-content/uploads/2016/12/Master-Abstract-Booklet.pdf
27. Hadjistavropoulos T, Janzen Claude A, Hadjistavropoulos H, Marchildon GP, Kaasalainen S, Gallagher R, et al. Stakeholder opinions on a transformational model of pain management in long-term care. J Gerontol Nurs 2011 Jul;37(7):40-51. [doi: 10.3928/00989134-20100503-03] [Medline: 21634316]
28. Gray-Toft P, Anderson JG. Stress among hospital nursing staff: its causes and effects. Soc Sci Med A 1981 Sep;15(5):639-647. [doi: 10.1016/0271-7123(81)90087-0] [Medline: 6976625]
29. Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behav 1981 Apr;2(2):99-113. [doi: 10.1002/job.4030020205]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
LTC: long-term care
MBI: Maslach Burnout Inventory
mHealth: mobile health
NSS: Nursing Stress Scale
PACSLAC: Pain Assessment Checklist for Seniors With Limited Ability to Communicate

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by R Eckhoff, C Reis; comments to author 15.01.20; revised version received 05.02.20; accepted 10.02.20; published 22.04.20.

Please cite as:
Zahid M, Gallant NL, Hadjistavropoulos T, Stroulia E
Behavioral Pain Assessment Implementation in Long-Term Care Using a Tablet App: Case Series and Quasi-Experimental Design
JMIR Mhealth Uhealth 2020;8(4):e17108
URL: http://mhealth.jmir.org/2020/4/e17108/
doi:10.2196/17108
PMID:32319955

Original Paper

Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach

Jonas Busk[1], MSc, PhD; Maria Faurholt-Jepsen[2], DMSc, MD; Mads Frost[3], PhD; Jakob E Bardram[4], MSc, PhD; Lars Vedel Kessing[2,5], DMSc, MD; Ole Winther[1,6,7], PhD

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark
[2]Copenhagen Affective Disorder Research Center, Psychiatric Center Copenhagen, Rigshospitalet, Copenhagen, Denmark
[3]Monsenso ApS, Copenhagen, Denmark
[4]Department of Health Technology, Technical University of Denmark, Lyngby, Denmark
[5]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[6]Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[7]Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Jonas Busk, MSc, PhD
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Richard Petersens Plads, 324
Lyngby, DK-2800
Denmark
Phone: 45 31692766
Email: jbusk@dtu.dk

Abstract

Background: Bipolar disorder is a prevalent mental health condition that is imposing significant burden on society. Accurate forecasting of symptom scores can be used to improve disease monitoring, enable early intervention, and eventually help prevent costly hospitalizations. Although several studies have examined the use of smartphone data to detect mood, only few studies deal with forecasting mood for one or more days.

Objective: This study aimed to examine the feasibility of forecasting daily subjective mood scores based on daily self-assessments collected from patients with bipolar disorder via a smartphone-based system in a randomized clinical trial.

Methods: We applied hierarchical Bayesian regression models, a multi-task learning method, to account for individual differences and forecast mood for up to seven days based on 15,975 smartphone self-assessments from 84 patients with bipolar disorder participating in a randomized clinical trial. We reported the results of two time-series cross-validation 1-day forecast experiments corresponding to two different real-world scenarios and compared the outcomes with commonly used baseline methods. We then applied the best model to evaluate a 7-day forecast.

Results: The best performing model used a history of 4 days of self-assessment to predict future mood scores with historical mood being the most important predictor variable. The proposed hierarchical Bayesian regression model outperformed pooled and separate models in a 1-day forecast time-series cross-validation experiment and achieved the predicted metrics, R^2=0.51 and root mean squared error of 0.32, for mood scores on a scale of −3 to 3. When increasing the forecast horizon, forecast errors also increased and the forecast regressed toward the mean of data distribution.

Conclusions: Our proposed method can forecast mood for several days with low error compared with common baseline methods. The applicability of a mood forecast in the clinical treatment of bipolar disorder has also been discussed.

(JMIR Mhealth Uhealth 2020;8(4):e15028) doi:10.2196/15028

KEYWORDS

bipolar disorder; mood; early medical intervention; digital phenotyping; machine learning; forecasting; Bayesian analysis

Introduction

Background

Bipolar disorder is estimated as one of the most important causes of disability worldwide [1,2]. Bipolar disorder is characterized by recurrent episodes of depression, (hypo)mania, and mixed episodes intervened by periods of euthymia [3] and with a high degree of comorbidity, functional impairment, and increased risk of suicide [4]. The World Health Organization estimates that about 60 million people are affected by bipolar disorder worldwide and that the burden of depression and other mental health conditions is on the rise globally [5]. The cornerstone of treatment of patients with bipolar disorder is continuous and long-term maintenance treatment to reduce or prevent relapses, applying a variety of methods including psychopharmacological treatment and group-based psychoeducation. Long-term treatment also involves symptom monitoring, early identification of subsyndromal symptoms of depression and mania, and intervention to prevent or reduce the severity of affective episodes.

In this paper, we analyzed daily self-assessments, including mood scores, collected from patients with bipolar disorder through a smartphone-based system. Ecological momentary assessment (EMA) reflects the method used to collect assessments of individuals' real-time states repeatedly, over time, during naturalistic settings and may reduce recall bias [6]. At present, a median of 76% of adults across 18 advanced economies reported having a smartphone [7], and many people use a smartphone daily [8]. The rapid evolution and ubiquity of mobile networks have resulted in the increasing growth of electronic mental health technologies, including electronic platforms offering tolls for remote self-monitoring [9]. By using daily smartphone-based self-monitoring, potential recall bias in self-reported patient data is minimized. Thus, smartphones extend the use of EMA beyond its classical use for self-reports and offer the opportunity to collect fine-grained data unobtrusively and outside clinical settings [10]. By replacing paper-based self-assessments of traditional treatment methods with a smartphone-based system, users can ubiquitously enter and review their own data and share the data with clinicians, who can intervene if something appears alarming. Thus, smartphones provide a unique platform for monitoring and managing depression and mania [11-13]. Furthermore, modern smartphones provide the means for collecting rich sensor data, which are believed to capture valuable behavioral information that can be related to disease outcomes [14]. Digital self-reporting and data collection have the additional benefit of making data available for automatic analysis immediately, which can help support continuous disease monitoring.

We found it useful to distinguish between *mood detection*, ie, predicting the mood based on data from the same day, and *mood forecasting*, ie, predicting the mood one or more days ahead based on historical data. Smartphone-based mood detection is well studied but remains a difficult problem. Several papers have examined the use of passive smartphone data, such as location, communication logs, and device usage, to detect or classify daily self-reported mood labels [15-21]. A few recent studies [22,23] have addressed mood forecasting, which is a more challenging task than mood detection, as the causal chain between cause and outcome is longer and because of the uncertainty inherently associated with future events. DeepMood by Suhara et al [22] is a solution for forecasting severely depressed mood from self-reported histories using a recurrent neural network. Suhara et al [22] found that long-term historical information up to 14 days improves the accuracy of forecasting depressed mood classes and that the mood on the previous day is the most important predictor when forecasting severe depression for one day. A limitation of this method is that it needs labeled observations every day in a 14-day history to make predictions. A study by Taylor et al [23] employed a selection of multi-task learning (MTL) techniques to train personalized models for predicting future mood, stress, and health one day ahead. Taylor et al [23] found that using MTL techniques to account for individual differences provides substantial performance improvements over traditional machine learning methods. By utilizing a cluster of users based on age, gender, and personality, a new user needs only to be assigned to a cluster to enable prediction based on new data, when labeled data from a population of similar users are available to fit the initial model.

A major challenge when reviewing work on mood prediction and behavior tracking is that researchers often have different data collection strategies and apply custom modelling and labeling approaches, consequently making results difficult to compare and sometimes contradicting [14]. Another limitation is that most studies involve healthy subjects (ie, without a diagnosis), and it is therefore hard to generalize to patients suffering from affective disorders. Nonetheless, some common observations stand out. Several studies found that personalized models generally outperform generic models when sufficient data are available [16,19,23-25], demonstrating the importance of accounting for individual differences in the data. This can be accommodated by applying MTL techniques, which provides a way of improving generalization by learning several related tasks simultaneously [26]. By considering individuals as separate tasks in a combined model, MTL techniques can produce personalized predictions in a straightforward manner.

Our study differs from prior work in a number of ways. Where many studies collect data from nonclinical subjects (such as students and volunteers recruited on the Web), our data are collected in a randomized clinical trial from patients who received a diagnosis of bipolar disorder and were treated for it. Moreover, to the best of our knowledge, the size of our patient population (N=84) is significantly larger than any prior clinical studies. We also found that even though most studies record subjective mood on a continuous or ordinal scale, the prediction task is often reduced to a classification problem by binning the values into two or more classes, such as *neutral*, *depressed*, and *manic*. In this study, we treated mood prediction as a regression problem, which is more direct given the way data are collected and interpreted by users. Finally, rather than mood detection, we addressed the more challenging task of mood forecasting and applied a hierarchical Bayesian modelling approach, which is a popular method of MTL that is able to account for individual differences in the data.

Objectives

The main objective of this study was to examine the feasibility and technical foundation of forecasting daily mood scores in bipolar disorder based on daily smartphone self-assessments. We hypothesized that utilizing these data to establish an accurate, real-time mood forecast solution can help improve disease monitoring by providing additional insights that enable early intervention and thus eventually prevent the relapse of affective episodes and burdensome and costly hospitalizations.

Methods

Data Description

Data used in this study were collected between September 2014 and January 2018 during the MONARCA II randomized clinical trial [27] that was investigating the effect of smartphone-based monitoring. All patients with a diagnosis of bipolar disorder who had previously been treated at the Copenhagen Clinic for Affective Disorder, Copenhagen, Denmark, in the period from 2004 to 2016 and who, at the time of recruitment, were being treated at community psychiatric centers, by private psychiatrists, and by general practitioners were invited to participate in the trial. Patients were included in the study for a 9-month follow-up period if they received a diagnosis of bipolar disorder according to International Classification of Diseases, 10th Revision using the Schedules for Clinical Assessments in Neuropsychiatry [28] and were previously treated at the Copenhagen Clinic for Affective Disorder. Patients with schizophrenia, schizotypal, or delusional disorders, previous use of the MONARCA system, pregnancy, and a lack of Danish language skills were excluded. Patients with other comorbid psychiatric disorders and substance use were eligible for the trial. As a part of the MONARCA II trial, patients were randomized to either using a smartphone-based monitoring system (the Monsenso system) for daily self-monitoring (the intervention group) or treatment as usual (the control group). Patients included in the intervention group collected daily smartphone-based self-monitoring data and were included in the analyses in this paper. The inclusion and exclusion criteria were investigated and assessed by 1 clinical researcher (MJ).

Study participants were provided an Android smartphone app configured for the study and were instructed to evaluate subjective measures of illness activity on a daily basis by answering a daily self-assessment questionnaire including the items listed in Table 1. Specifically, mood was scored on a scale of −3, −2, −1, −0.5, 0, 0.5, 1, 2, and 3, where negative values indicate various degrees of depression, positive values indicate mania, and zero indicates neutral mood (euthymia).

Table 1. Items of the daily self-assessment questionnaire.

Attribute	Description	Value
Activity	Level of physical activity	−3 to 3
Alcohol	Alcoholic drinks consumed	0 to 10+
Anxiety	Level of anxiety	0 to 2
Irritable	Level of irritability	0 to 2
Cognitive difficulty	Level of cognitive discomfort	0 to 2
Medicine	Medicine adherence	0 to 2
Mixed mood	Experienced mixed mood	0 to 1
Mood	Experienced mood	−3 to 3
Sleep	Hours of sleep	0 to 24
Stress	Level of stress	0 to 2

Additionally, study participants were periodically evaluated by trained psychiatrists throughout the trial, up to five times (at baseline and after 4 weeks, 3 months, 6 months, and 9 months), on the following clinical rating scales for depression and mania: the Hamilton Depression Rating Scale (HDRS) [29] and Young Mania Rating Scale (YMRS) [30]. Each rating scale consists of a series of questions that are scored and totaled to summarize the current state of the patient with higher scores indicating more severe symptoms. Clinical researchers, who were blinded to all smartphone-based data, conducted all the clinical assessments. Thus, data on the severity of depressive and manic symptoms were collected in a rater-blinded manner. Both rating scales are clinically validated and generally accepted as accurate measures of illness severity in bipolar disorder.

Data Preprocessing

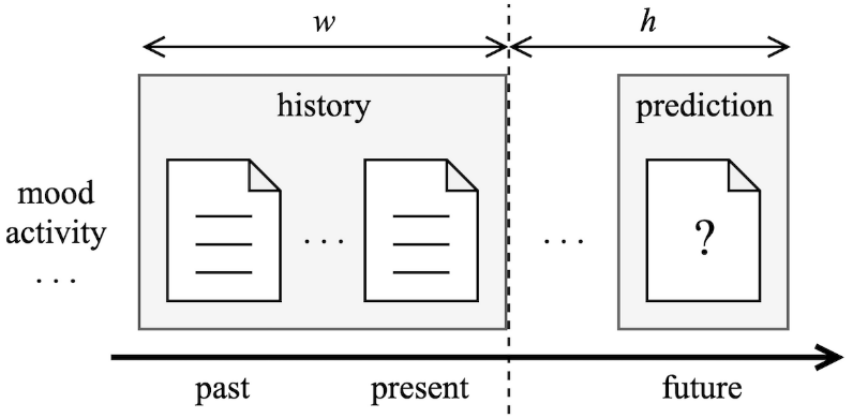
Two of the self-assessment items were preprocessed before the analysis. As the answer to the medicine item is categorical by design (medicine not taken, medicine taken, and medicine taken with changes), we encoded it as two exclusive binary variables indicating if medicine was not taken (medicine omitted) or if medicine was taken with changes (medicine changed). Additionally, we did not expect sleep duration to have a linear effect on mood, thus the sleep variable was replaced with two new features by subtracting the individual mean and splitting the result into a negative and positive component (sleep negative and sleep positive), indicating decreased or increased sleep relative to the mean. Finally, we normalized all self-assessment variables by their allowed minimum and maximum value. We also experimented with forward filling the missing data from the previous day but found that very little additional data were

gained; therefore, we left this step out of the final analysis presented in this paper.

Forecasting

Forecasting is the task of predicting the future, given all available information from the past and present [31,32]. For forecasting to be feasible, it should be reasonable to assume that the history of recorded information somehow relates to the predicted future events. A typical forecasting task is illustrated in Figure 1; *w* denotes the size of history used in the forecast and *h* denotes the horizon of how far into the future the target is predicted. In our case of using daily self-reports, both *w* and *h* are measured in days.

Figure 1. Forecasting is the task of predicting the future, given all relevant information from the past and the present. The window size, w, is the size of history defining the predictor variables and the horizon, h, is how far in the future the target variable is predicted.



Several methods for producing forecasts exist [32]. The simplest forecasting methods use only historic information about the target variable and do not consider any other information but are designed to utilize time-dependent patterns, such as seasonality and trend, to extrapolate observed data into the future. Another approach is to apply standard regression or classification models to predict the variable of interest based on relevant information, such as prior (lagged) observations of the target variable along with additional predictor variables. This approach has the benefit of allowing the use of a variety of different methods from the machine learning and statistical inference literature but may not be as good at capturing long-term time-dependent patterns. For short-term forecasts or data without long-term time dependence, however, this might not be a problem. For these reasons, we chose to apply the latter approach in this work.

Special care should be taken when evaluating the performance of a forecast. A genuine forecast only uses data available at the time of forecast, and thus no future data, to estimate its parameters [32]. Consequently, the size of in-sample residuals (training error) is not a reliable indicator of the true forecast error. The forecast performance can only be determined by fitting the model on training data observed before the test data. This needs to be considered when designing the experiment used to evaluate the forecast model, such as cross-validation. Time-series cross-validation addresses this by splitting the data into a sequence of consecutive test sets. The corresponding training sets consist only of data observed before each test set. Thus, no future information is included when constructing the forecasts. The cross-validation error is then computed across all the test sets. As we considered data from multiple individuals, we applied two different time-series cross-validation in our experiments:

1. Leave-all-out time-series cross-validation: Each individual's data are partitioned into a sequence of *T* consecutive similar-sized test sets. Then the test sets are pooled across all individuals. The corresponding training sets consist of all data observed before each test set, resulting in $T-1$ test and training set pairs (the first test set has no prior data).
2. Leave-one-out time-series cross-validation: Each individual's data are partitioned into a training set and subsequent test set. The training set is then pooled with all data from all other individuals, resulting in a number of test/training set pairs equal to the number of individuals, *J*.

These two experiments correspond to two different scenarios: the leave-all-out time-series cross-validation simulates a situation where a group of patients starts monitoring at the same time without any additional historical data, whereas the leave-one-out time-series cross-validation simulates a situation where each participant starts monitoring when data are already available from a population of similar individuals.

Hierarchical Bayesian Models

When analyzing data consisting of multiple related sets of measurements, such as individuals in a population, a basic approach is to completely pool all the data into a common model, assuming all sets have similar properties. A drawback of this method is there is a risk of losing important information at the individual level. To overcome this problem, an alternative approach is to model each set of data separately, assuming all sets are independent. However, information about how the individual sets relate to each other at the population level might be missed. Especially when each individual dataset is too small to construct a meaningful separate model, it is useful to include information from the population to make analysis feasible. A hierarchical (multi-level) Bayesian model is an intermediate solution allowing partial pooling of the data, thus providing a compromise between the completely pooled and separate models [33,34]. The hierarchical approach captures the overall characteristics of the population while allowing individual

differences and enables modeling of small related datasets, each getting a gradually more personalized model as more data are collected and included in the training set. Additionally, it allows for reasoning about previously unobserved individuals, assuming they come from the same population, which helps to overcome the cold start problem. Applying a Bayesian approach has the additional benefit of providing uncertainty in all model parameters and predictions, allowing for improved interpretability. Owing to these desirable properties, we applied hierarchical models in our analysis. In particular, we explored the use of hierarchical implementations of both linear and ordinal regression models.

Ordinary linear regression is a method of predicting the outcome of a continuous variable, modeled as the linear combination of the model parameters and predictor variables. Hierarchical Bayesian linear regression can be expressed by assuming that each set of parameters is drawn from a common population distribution (Figure 2). For individual j=1:J, observation i=1:N, target variable y_{ji}, and predictor variables $\boldsymbol{x_{ji}}$:

$$y_{ji} = \text{Normal}(\alpha_j + \beta_j^T \boldsymbol{x_{ji}}, \sigma)$$

where α_j and β_j are sampled from population distributions:

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \tau_\alpha)$$

$$\beta_j \sim \text{Normal}(\mu_\beta, \tau_\beta)$$

and the population means μ_α, μ_β and variances τ_α,τ_β as well as the standard error σ have independent normal priors.

Ordinal regression (sometimes referred to as ordinal classification) is a method of predicting a discrete variable that has a relative ordering of the possible outcomes. Thus, it can be thought of as an intermediate between regression and classification. An example of ordinal regression is ordered logistic regression. For an outcome belonging to one of K categories, $y_{ji} \in 1{:}K$, ordered logistic regression is determined by a latent continuous variable, $z_{ji} = \beta_j^T \boldsymbol{x_{ji}}$, along with a sequence of K+1 ordered cutpoints, $\boldsymbol{c_j}$, such that $c_{k-1} < c_k$ and $c_0 = -\infty$, $c_K = \infty$ by definition. If z_{ji} falls between two cutpoints, c_{k-1} and c_k, the outcome is predicted to belong to the corresponding category, $y_{ji} = k$, with high probability. This type of model can be justified by assuming the category, y_{ji}, is an incomplete measurement of the latent variable, z_{ji}:

$$y_{ji} \sim \text{OrderedLogistic}(z_{ji}, \boldsymbol{c_j})$$

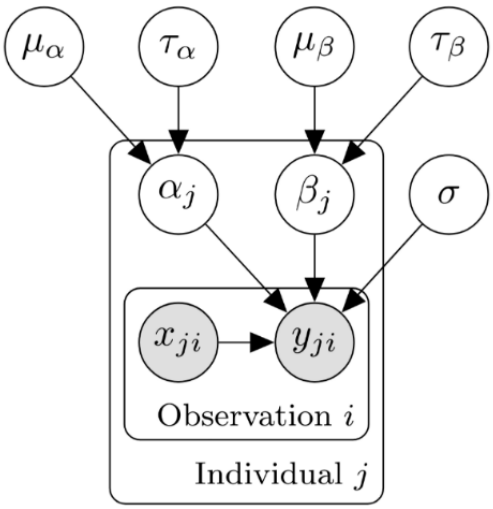
Hierarchical Bayesian ordinal regression can be expressed by assuming that each set of model parameters is drawn from a common population distribution:

$$\beta_j \sim \text{Normal}(\mu_\beta, \tau_\beta)$$

$$\boldsymbol{c_j} \sim \text{Normal}(\mu_c, \tau_c)$$

with independent normal priors on the population parameters, μ_β, τ_β,μ_c, τ_c, along with ordering constraints on μ_c and $\boldsymbol{c_j}$. In practice, we re-parameterized the hierarchical models to achieve more efficient sampling [35]. A practical difference of using ordinal regression over linear regression is that ordinal regression can never produce predictions (or uncertainties) outside the range of the training data. This can be an advantage when the target variable represents a bounded scale where values outside the scale do not have any meaning. Ordinary linear regression does not provide this guarantee; thus, the ordinal model can lead to more interpretable outcomes.

Figure 2. A Bayesian network of a hierarchical linear regression model. Individual regression intercept α_j and weights β_j are drawn from population distributions parameterized by μ_α, τ_α and μ_β, τ_β. This allows the model to account for individual differences while constraining individual parameters to be similar across the population.



We used the open-source statistical modeling platform, Stan [36], to specify and perform inference in the hierarchical models. Generally, the models were fitted using four sampling chains and 5,000 iterations, where the first half was warm-up and parameter tuning, resulting in 10,000 posterior samples. Our prior belief was that self-reported mood would be the strongest predictor of future mood, hence the population parameters corresponding to mood were assigned less restrictive priors than the other population parameters, which were assigned more restrictive priors to introduce regularization. The Stan code of the hierarchical models and more details on the priors is included in Multimedia Appendix 1. To provide appropriate baseline

results for comparison, a suite of naïve and standard machine learning regression models from the Scikit-learn machine learning library [37] and the popular XGBoost Python package [38] were also evaluated. These models were fitted both with pooled and separate data, where applicable.

Ethical Considerations

The Regional Ethics Committee in the Capital Region of Denmark (H-2-2014-059) and the Danish Data protection agency (2013-41-1710) approved the trial. The law on handling of personal data was respected. Before commencement, the trial was registered at ClinicalTrials.gov (NCT02221336). Electronic data collected from the smartphones were stored at a secure server at Concern IT, Capital Region, Denmark (I-suite number RHP-292 2011-03). The trial complied with the Helsinki Declaration of 1975, as revised in 2008.

Results

Descriptive Statistics

The dataset consists of 15,975 daily self-assessments and 280 clinical evaluations from 84 participants. This corresponds to an average of 190.2 self-assessments per individual and an average self-assessment adherence of 82.8% between the first and last submitted self-assessment. The population ranged from the ages of 21 to 71 years (mean 43.1, SD 12.4) and consisted of 62% (52/84) women. Figure 3 presents the distribution of self-reported mood scores across all individuals in the dataset (mean −0.14, SD 0.48). The majority of observed mood scores, y, are centered around zero, indicating euthymia ($-0.75<y<0.75=89.64\%$) with only few values indicating depression ($y<-0.75=8.68\%$) and even fewer values indicating mania ($y>0.75=1.68\%$). As expected, the self-reported mood scores and HDRS scores were negatively correlated ($r=-0.40$; $P<.001$) and self-reported mood scores and YMRS scores were positively correlated ($r=0.22$; $P<.001$).

Figure 3. Distribution of all self-reported mood scores (left) and individual mean mood scores (right). The mood scores are generally close to zero indicating neutral mood with only a few exceptions indicating depressed or elevated mood.

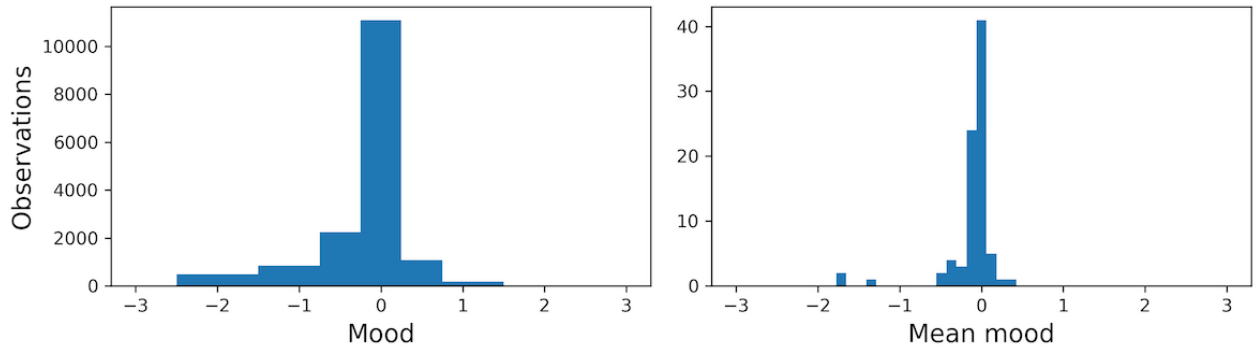


Figure 4. The mean of individual correlations of self-assessment items and mood lagged up to 7 days. Nonzero correlations indicate that items have some relation to mood on subsequent days that can be utilized for mood forecasting.

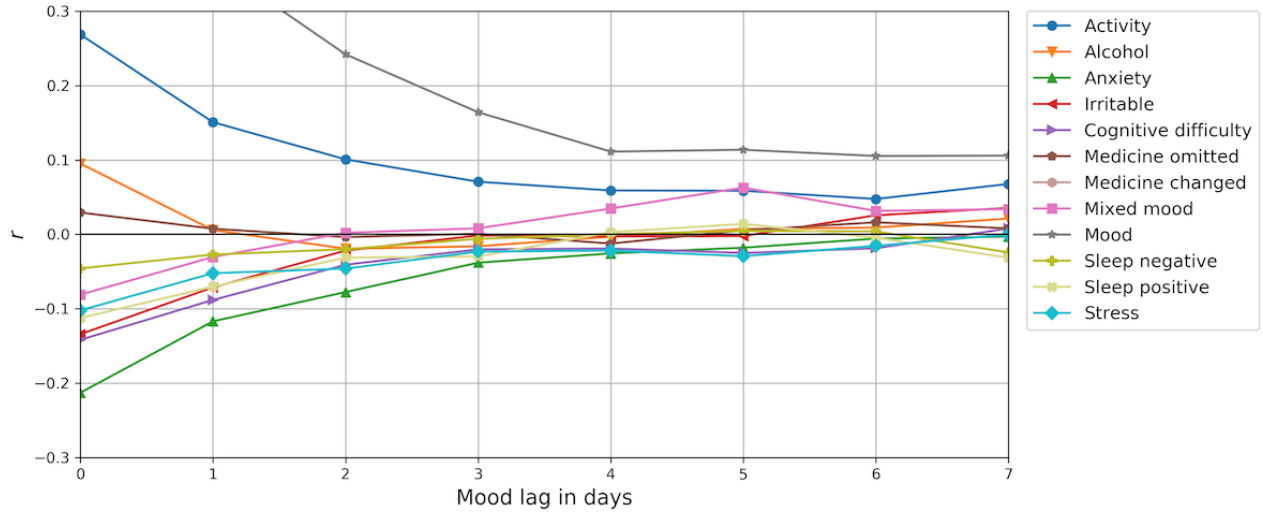


Figure 4 shows correlations of self-assessment items with self-reported mood lagged for up to seven days. Self-reported mood has a positive autocorrelation for the entire duration of 1 to 7 days. Additionally, activity has a positive correlation with mood for a few days, indicating that high activity levels coincides with elevated mood, and anxiety has a small negative correlation with mood, indicating that anxiety often coincides with negative mood scores. The remaining self-assessment items

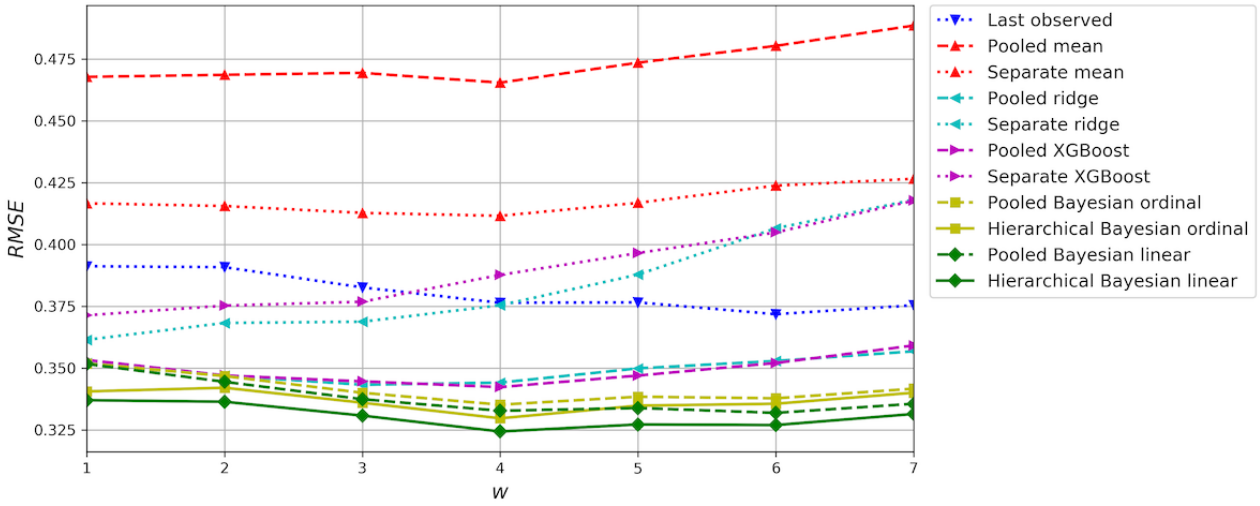
show small, diminishing correlations with lagged mood. A seasonality analysis of self-reported mood revealed no significant monthly or daily seasonality in the data and was left out for brevity.

Window Size Selection

To find the optimal window size, *w*, for forecasting mood, we evaluated a 1-day forecast with window sizes from 1 to 7 days. Each window size was evaluated in a *T*=24 leave-all-out time-series cross-validation experiment with data partitions with a size of one week. The predicted coefficient of determination (R^2), indicating the proportion of the data variance explained by each model (higher is better), and the root mean squared error (RMSE), measuring the square root of the mean of squared errors (lower is better), were computed across all the test sets.

Figure 5 shows the RMSE of the cross-validation for *w*=1 through 7 and *h*=1. The errors of the naïve mean models are almost constant, varying only because of the difference in datasets available for different values of *w*. The hierarchical Bayesian linear regression model achieved the lowest RMSE of all models for every window size with the best result at *w*=4 days, which we then used in the following analysis.

Figure 5. Window size selection results. The root mean squared error (RMSE) was evaluated in time-series cross-validation experiments for w=1 through 7 and h=1. The lowest RMSE was achieved by the hierarchical linear model at w=4.



Model Checks and Feature Importance

To evaluate how well the proposed hierarchical linear and ordinal models fit the data distribution, we trained them on the entire dataset of participants with at least two data points for *w*=4 and *h*=1 (*N*=5881). The hierarchical models achieved a similar fit with in-sample R^2=0.56 and in-sample RMSE=0.29. We then performed posterior predictive checks by testing the ability of the models to replicate (predict) the observed distribution of future mood from the observed history of predictor variables. In particular, we computed the ratio of observed mood values and replicated mood values less than −0.75 and greater than 0.75. The hierarchical linear model replicated 93% of the small values while the ordinal model replicated 65% of the small values. The hierarchical linear model replicated 73% of the large values while the ordinal model only replicated 24% of the large values. Thus, the hierarchical linear model is better at capturing the tails of the distribution whereas the ordinal model underestimates extreme values.

The importance of a predictor variable in a linear regression model can be measured as the absolute value of the *t*-statistic of its regression weight, β, computed as the mean weight scaled by its standard error: $t_\beta=\beta/SE(\beta)$ [39]. Table 2 presents the mean absolute *t*-statistic of the individual-level regression weights in the hierarchical Bayesian linear regression model for each of the predictor variables in a 4-day history. This shows that self-reported mood is the most important variable for predicting mood the next day, which is not surprising considering mood has a strong autocorrelation (Figure 4).

Table 2. Predictor variables sorted by overall feature importance measured by the mean absolute t-statistic of the individual-level regression parameters in the hierarchical Bayesian linear regression model for w=4 and h=1. Self-reported mood is the most important variable for predicting mood on the following day.

Predictor	|*t*|, mean (SD)			
	x_{t}	x_{t-1}	x_{t-2}	x_{t-3}
Mood	4.53 (3.35)	2.34 (0.55)	0.47 (0.28)	2.78 (0.18)
Anxiety	2.78 (0.05)	0.71 (0.02)	1.29 (0.01)	0.76 (0.00)
Irritable	2.74 (0.11)	1.22 (0.01)	0.95 (0.01)	1.30 (0.00)
Mixed mood	2.09 (0.06)	2.51 (0.02)	1.96 (0.01)	0.52 (0.01)
Medicine changed	0.36 (0.10)	0.08 (0.01)	2.15 (0.01)	0.64 (0.00)
Sleep positive	1.65 (0.01)	0.72 (0.00)	0.37 (0.00)	0.16 (0.00)
Cognitive difficulty	1.48 (0.09)	0.58 (0.02)	0.19 (0.00)	1.57 (0.00)
Alcohol	0.67 (0.02)	0.77 (0.01)	1.56 (0.01)	0.87 (0.00)
Medicine omitted	0.13 (0.01)	1.31 (0.00)	0.60 (0.00)	0.14 (0.00)
Stress	1.22 (0.12)	0.91 (0.02)	0.71 (0.01)	0.28 (0.01)
Activity	1.04 (0.03)	1.14 (0.02)	0.49 (0.01)	1.14 (0.01)
Sleep negative	0.41 (0.01)	0.52 (0.00)	0.48 (0.00)	0.52 (0.00)

Time-Series Cross-Validation Results

The results of the leave-all-out and leave-one-out time-series cross-validation experiments for w=4 and h=1 are presented in Table 3. In both experiments the naïve pooled mean model scored a predicted R^2 close to zero because it does not explain any variance in the data. A predicted R^2 score greater than zero indicates that some variance is explained while a negative R^2 score is worse than the pooled mean model. The *last observed* model simply repeats the last observed mood value, which performs considerably better than the mean model and represents a solid baseline.

Table 3. Results of the leave-all-out time-series cross-validation (left) and leave-one-out time-series cross-validation (right) experiments. The hierarchical Bayesian linear regression model achieves the best results. The pooled models are better than the separate models, overall.

Model	Leave-all-out		Leave-one-out	
	R^2 [a]	RMSE[b]	R^2 [a]	RMSE[b]
Last observed	0.342	0.376	0.151	0.385
Pooled mean	−0.007	0.465	−0.009	0.419
Pooled ridge	0.450	0.344	0.340	0.339
Pooled XGBoost	0.455	0.342	0.343	0.338
Separate mean	0.213	0.412	−0.443	0.502
Separate ridge	0.345	0.375	−0.471	0.506
Separate XGBoost	0.302	0.388	−0.682	0.541
Hierarchical Bayesian linear	0.511	0.324	0.347	0.337
Hierarchical Bayesian ordinal	0.495	0.330	0.343	0.339

[a]Coefficient of determination (R^2): higher is better.

[b]Root mean squared error (RMSE): lower is better.

The leave-all-out time-series cross-validation experiment was evaluated with T=24 and data partitions a size of one week, resulting in T−1=23 iterations of cross-validation. The hierarchical Bayesian linear model achieved the best result with the predicted R^2=0.511 and predicted RMSE=0.324, beating the naïve baseline and pooled and separate regression models. The hierarchical Bayesian ordinal model is a close second best.

The leave-one-out time-series cross-validation experiment was evaluated for each individual with the first 2 weeks of data pooled with data from the rest of the population in the training set and evaluated on the next 22 weeks of data from that individual, resulting in J=58 iterations of cross-validation. The hierarchical Bayesian linear model achieved the best predicted R^2=0.347 and predicted RMSE=0.337, but is similar to the best pooled regression models, indicating that the hierarchical model does a lot of pooling as well. The separate models fail to

generalize to the held-out test data in this experiment, achieving negative R^2 scores, because the training sets contain only 2 weeks of data. Overall, the hierarchical and pooled models performed better than the separate models, and all regression models generally outperformed the naïve baseline models when sufficient data were available.

Seven-Day Forecast

Thus far we have focused on evaluating a 1-day forecast, but it is also interesting to forecast mood on a more distant horizon. Figure 6 shows the mean RMSE of cross-validation for w=4 and h=1 through 7. The hierarchical Bayesian linear regression model achieves the lowest RMSE of all models for every value of h. As might be expected, the error generally grows with the size of the horizon. The errors of the naïve mean models are almost constant, varying only because of the difference in datasets available for different values of h. However, even at h=7, the best regression models are able to outperform the mean models, meaning they are able to capture useful information from prior self-assessments. Two examples of 7-day mood forecasts produced by the hierarchical linear regression model are presented in Figure 7.

Figure 6. Results of forecasting mood for up to seven days. The root mean squared error (RMSE) was evaluated in time-series cross-validation experiments for w=4 and h=1 through 7. As expected, the RMSE increases when forecasting further ahead. The proposed hierarchical models achieved consistently lower RMSEs than the baseline models.

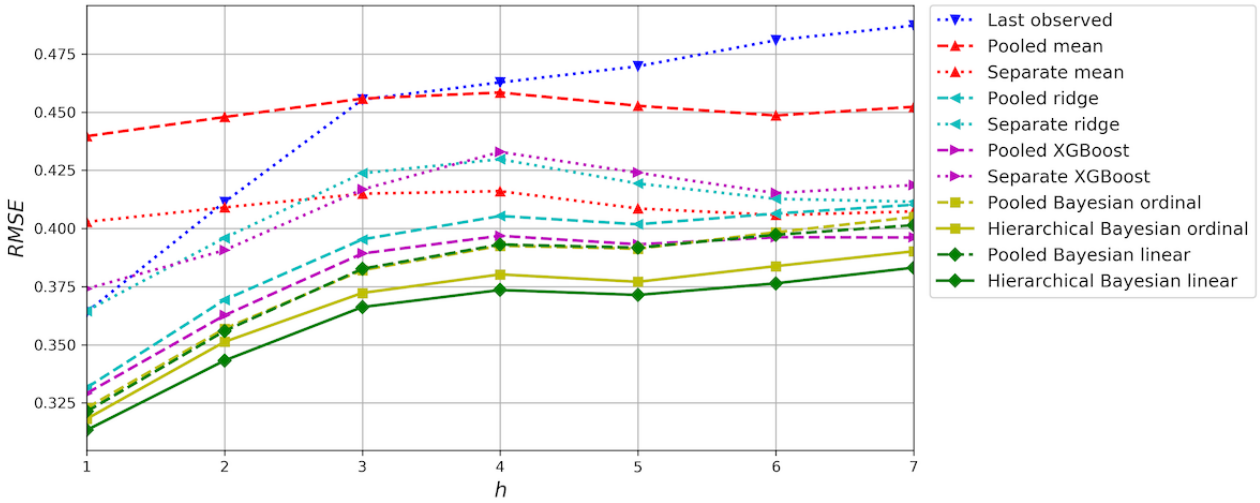
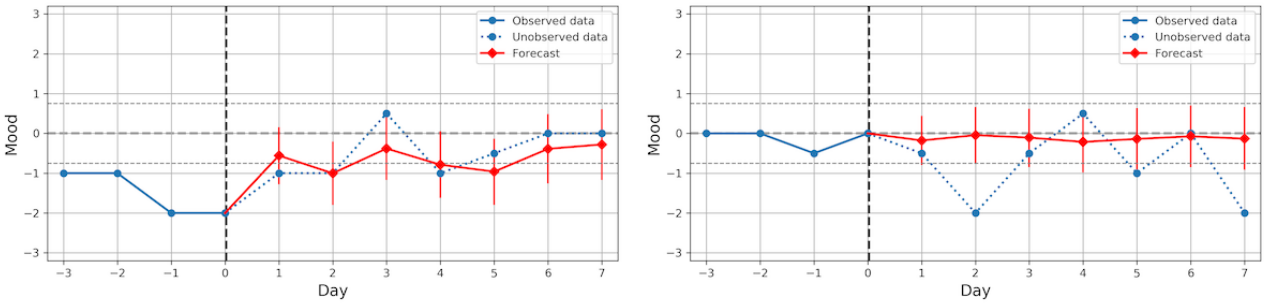


Figure 7. Examples of 7-day mood forecasts produced by the hierarchical linear regression model. The forecasted mood values are shown with 95% CI uncertainties and compared with observed data. The forecast to the left is rather accurate despite variation in the data, whereas the forecast to the right fails to anticipate future mood changes.



Discussion

Principal Findings

In this study, we have analyzed smartphone-based self-assessment data from a population of 84 patients with bipolar disorder with the purpose of forecasting subjective mood. The initial data analysis showed that the majority of observed mood scores are close to zero, indicating weak or no symptoms among the population for most of the study period. Yet, we found a significant negative correlation between self-reported mood scores and HDRS scores (r=−0.40; P<.001) and a significant positive correlation between self-reported mood scores and YMRS scores (r=0.22; P<.001). This confirms prior findings [40-42], suggesting that subjective mood is a valid indicator of the mental state in bipolar disorder and thereby also a clinically relevant feature for daily monitoring and forecasting. We did not observe any substantial seasonality or long-term trend of subjective mood, indicating that time-series models designed to utilize such time-dependent patterns [32] are not appropriate for forecasting mood. However, the recorded mood scores do show an autocorrelation several days ahead. Thus, we employed a multiple regression approach based on a history of predictor variables to forecast future mood scores. In particular, we proposed using a hierarchical Bayesian model to perform MTL, enabling personalized predictions while

considering common characteristics of the population. The hierarchical approach additionally makes it possible to reason about individuals for whom we have observed little data, thus overcoming the cold start problem.

Employing a regression model approach to produce a forecast required us to find an appropriate window size defining the predictor variables included in the model. With perfect data and a model robust to overfitting, increasing the window size should never result in a worse model, as any added noninformative variables could simply be ignored. In a real-world application, however, increasing the window size often results in fewer training examples because of missing data and similarly requires more data to enable prediction on new instances. Thus, finding the optimal window size is a trade-off that depends on data quality and model robustness. In our experiment, we found that including a history of up to four days improved the prediction error, but with more complete data, there is no reason the window size could not be increased even further. For instance, Suhara et al [22] found that their model for classifying depression benefited from long data histories up to 14 days, although it is our experience that collecting complete self-assessment histories over an extended period is very difficult.

By inspecting the inferred regression parameters of the hierarchical Bayesian model, we found historical mood to be the most important predictor of future mood. This result is not surprising as substantial changes in mood often occur over several days, and thus, future mood is likely to be similar to the mood in the immediate past. Consequently, the forecast is inclined to extrapolate the mood from previous days and gradually regress toward the mean of the data as uncertainty grows when forecasting further ahead. Although this forecast behavior succeeds at achieving a low error, its utility in a practical monitoring setting must be studied further. We see this as an interesting topic for future research. However, the results presented in this paper show that regression models based on self-assessment histories are able to consistently outperform naïve forecast baselines of either repeating the last observed value or predicting the mean of the pooled or separate data distributions up to seven days into the future (see Figure 7).

The proposed hierarchical linear and ordinal models achieved the best predictive performance in the time-series cross-validation experiments. In the leave-all-out cross-validation, the hierarchical Bayesian linear regression model achieved the best result (R^2=0.511; RMSE=0.324) with the hierarchical Bayesian ordinal model being a close second. In the leave-one-out cross-validation, the hierarchical Bayesian linear regression model also achieved the best result (R^2=0.547; RMSE=0.337) but was much closer to the performance of the best pooled models. These results show how the hierarchical approach solves the cold start problem by including information from the population when little individual data are observed and by gradually becoming more personalized as more data become available. In contrast to previous work, we found that pooled models outperformed separate models, indicating that the individual datasets did not contain sufficient information to produce accurate forecasts. Thus, the separate models were biased and consequently it proved more useful to disregard individual differences and include data from the population in a general model. The hierarchical models succeeded in finding a compromise between the pooled and separate approach by regularizing the personalized models with data from the population.

In forecasting mood for several days, the hierarchical models similarly achieved the best results. As expected, the forecast error increased when forecasting further ahead; however, we observed that the best regression models performed better than the naïve mean models for up to seven days. It is a remarkable result that a short self-assessment history of just a few days can forecast mood for several days, the most important reason being that substantial mood changes often happen gradually over a horizon longer than 7 days.

The data analyzed in this study were collected from a population of well-characterized patients with bipolar disorder during the MONARCA II randomized clinical trial [27] conducted by researchers with specific knowledge of bipolar disorder. Overall, the findings from this study are found to be generalizable to patients with bipolar disorder not presenting with an acute affective episode and who are willing to use a smartphone-based monitoring tool.

Limitations

We observed a low prevalence of severe symptoms in our data sample leading to some limitations. As the mood values have low variance, regression models will tend to regress toward the mean of the data, and naïve mean models are able to achieve low errors relative to the full range of the mood scale. It prevented us from assessing how well the proposed method performs in a population with more severe symptoms and how well the forecast is at anticipating severe cases.

A major motivation for our research and the MONARCA II study was to establish a real-time mood forecasting solution to improve monitoring and enable early intervention in patients with bipolar disorder [27]. However, it is still not clear how a real-time forecast system is affected by interventions, as the intervention can change the outcome and thus future training data, which could lead to a biased model that underestimates future mood scores. Thus, it would be crucial to monitor the performance of a real-time system continuously using held out, unbiased data for validation.

Perspectives

The mood forecast presented in this paper has used a history of self-reported features as input. However, several research projects have been investigating the use of sensor-based and automatically collected data as input for mood prediction. Sensor technology in modern smartphones enables tracking of a variety of behavioral features such as physical activity, location, and sleep along with communication and device usage logs. Additionally, sensor data can be captured with wearables such as wristbands and fitness trackers with high accuracy. Such sensor-based features could be used to augment or even reduce self-assessment in mood prediction tasks and thus reduce the need to prompt users for daily self-assessments. There is great

potential in utilizing objectively collected sensor data in semiautomatic mood detection and forecasting.

Mood prediction and forecasting can be used as early warning signs in clinical treatment. Furthermore, accurate symptom forecasting could be extended to detect risk of relapse of major affective episodes specifically, eg, by detecting if values exceed predefined thresholds over consecutive days. This could be useful in, eg, a telemedicine setup in which trained nurses or other clinical personnel supervise patients in outpatient treatment. This could help catch early onset of major depressive or manic phases that can be addressed and handled early, which again could reduce the severity of symptoms and the degree of treatment. Hence, the need for readmission could be reduced. We are currently working on implementing a Web-based forecasting system evaluated as part of the RADMIS (reducing the rate and duration of readmissions among patients with unipolar disorder and bipolar disorder using smartphone-based monitoring and treatment) trials [43] to study its practical application, including investigating if such a system could potentially reduce readmission and hospitalization.

In this paper, we have examined the technical foundation of mood forecasting aimed at improving continuous disease monitoring. However, for a patient, the prospect of experiencing depressed or elevated mood in the future might lead to changes in behavior and state of mind and, in the worst case, become a self-fulfilling prophecy. Therefore, real-time mood forecasting should be used with care and applied exclusively as a monitoring and early intervention tool for professionals rather than being presented directly to users.

Conclusions

Continuous symptom monitoring and early detection are important components in the treatment of patients with bipolar disorder. Smartphones provide a unique platform for self-assessment and management of depression and mania and have the additional benefit of making data available for immediate analysis. In this work, we have examined the feasibility of establishing a mood forecast system based on self-assessments to provide additional insights and enable early intervention. We found that our proposed method of applying hierarchical Bayesian regression models was able to consistently outperform commonly used machine learning methods and forecast subjective mood for up to seven days.

Acknowledgments

This study was funded by the Innovation Fund Denmark through the RADMIS project and the Copenhagen Center for Health Technology. The authors would like to thank everyone who participated in the MONARCA II trial and the clinical staff at the Psychiatric Center Copenhagen who helped facilitate the trial and collect the dataset used in this work.

Conflicts of Interest

JB, MJ, and OW have no conflicts of interest. MF and JEB are founders and shareholders of Monsenso. LK has been a consultant for Sunovion and Lundbeck in the last 3 years.

Multimedia Appendix 1
Stan code of the hierarchical linear regression and ordinal regression models and details on choice of model priors.
[PDF File (Adobe PDF File), 94 KB - mhealth_v8i4e15028_app1.pdf]

References

1. Pini S, de Queiroz V, Pagnin D, Pezawas L, Angst J, Cassano GB, et al. Prevalence and burden of bipolar disorders in European countries. Eur Neuropsychopharmacol 2005 Aug;15(4):425-434. [doi: 10.1016/j.euroneuro.2005.04.011] [Medline: 15935623]
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
3. Goodwin FK, Jamison KR. Manic-Depressive Illness. New York: Oxford University Press; 1990.
4. Sanchez-Moreno J, Martinez-Aran A, Colom F, Scott J, Tabares-Seisdedos R, Sugranyes G, et al. Neurocognitive dysfunctions in euthymic bipolar patients with and without prior history of alcohol use. J Clin Psychiatry 2009 Aug;70(8):1120-1127. [doi: 10.4088/JCP.08m04302] [Medline: 19758523]
5. World Health Organization. 2019 Nov 28. Mental Disorders URL: https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders [accessed 2019-03-15] [WebCite Cache ID 76tGO34od]
6. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
7. Silver L. Pew Research Center. 2019 Feb 5. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-05-29]
8. Ericsson. Ericsson Mobility Report URL: https://www.ericsson.com/en/mobility-report [accessed 2020-01-28]

9. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
10. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
11. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: ACM; 2013 Presented at: CHI'13; April 27 - May 2, 2013; Paris, France p. 2627-2636. [doi: 10.1145/2470654.2481364]
12. Frost M, Doryab A, Faurholt-Jepsen M, Kessing LV, Bardram JE. Supporting Disease Insight Through Data Analysis: Refinements of the Monarca Self-Assessment System. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. USA: ACM; 2013 Presented at: UbiComp'13; September 8 - 12, 2013; Zurich, Switzerland p. 133-142. [doi: 10.1145/2493432.2493507]
13. Bardram JE, Frost M. The personal health technology design space. IEEE Pervasive Comput 2016;15(2):70-78. [doi: 10.1109/MPRV.2016.37]
14. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
15. Bogomolov A, Lepri B, Pianesi F. Happiness Recognition from Mobile Phone Data. In: Proceedings of the 2013 International Conference on Social Computing. 2013 Presented at: SocialCom'13; September 8-14, 2013; Alexandria, VA, USA. [doi: 10.1109/socialcom.2013.118]
16. Canzian L, Musolesi M. Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces Analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. USA: ACM; 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
17. Farhan AA, Yue C, Morillo R, Ware S, Lu J, Bi J, et al. Behavior vs Introspection: Refining Prediction of Clinical Depression via Smartphone Sensing Data. In: Proceedings of the 2016 IEEE Wireless Health. 2016 Presented at: WH'16; October 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764553]
18. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]
19. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services. USA: ACM; 2013 Presented at: MobiSys'13; June 25 - 28, 2013; Taipei, Taiwan p. 465-466. [doi: 10.1145/2462456.2483967]
20. Ma Y, Xu B, Bai Y, Sun G, Zhu R. Daily Mood Assessment Based on Mobile Phone Sensing. In: Proceedings of the 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: BSN'12; May 9-12, 2012; London, UK. [doi: 10.1109/bsn.2012.3]
21. 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 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
22. Suhara Y, Xu Y, Pentland A. DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 715-724. [doi: 10.1145/3038912.3052676]
23. Taylor SA, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affective Comput 2017 Dec 19:1-1. [doi: 10.1109/TAFFC.2017.2784832]
24. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
25. Constantinides M, Busk J, Matic A, Faurholt-Jepsen M, Kessing LV, Bardram JE. Personalized versus Generic Mood Prediction Models in Bipolar Disorder. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. 2018 Presented at: UbiComp'18; October 8 - 12, 2018; Singapore, Singapore p. 1700-1707. [doi: 10.1145/3267305.3267536]
26. Caruana R. Multitask learning. Mach Learn 1997;28(1):41-75. [doi: 10.1023/A:1007379606734]
27. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram J, Kessing LV. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones--the MONARCA II trial protocol: a randomized controlled single-blind parallel-group trial. BMC Psychiatry 2014 Nov 25;14:309 [FREE Full text] [doi: 10.1186/s12888-014-0309-5] [Medline: 25420431]
28. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990 Jun;47(6):589-593. [doi: 10.1001/archpsyc.1990.01810180089012] [Medline: 2190539]
29. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967 Dec;6(4):278-296. [doi: 10.1111/j.2044-8260.1967.tb00530.x] [Medline: 6080235]

30. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
31. Bontempi G, Taieb SB, Le Borgne YA. Machine learning strategies for time series forecasting. In: Business Intelligence. Berlin, Heidelberg: Springer; 2013:62-77.
32. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. Second Edition. OTexts: Melbourne, Australia; 2018.
33. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Third Edition. Boca Raton, Florida: Chapman and Hall/CRC; 2013.
34. Murphy KP, Bach F. Machine Learning: A Probabilistic Perspective. Cambridge: MIT Press; 2012.
35. Betancourt MJ, Girolami M. arXiv e-Print archive. 2013. Hamiltonian Monte Carlo for Hierarchical Models URL: https://arxiv.org/abs/1312.0906 [accessed 2020-01-28]
36. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Soft 2017;76(1):1-32. [doi: 10.18637/jss.v076.i01]
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
38. 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. USA: ACM; 2016 Presented at: KDD'16; August 13 - 17, 2016; California, San Francisco, USA p. 785-794. [doi: 10.1145/2939672.2939785]
39. Molnar C. Interpretable machine learning: A Guide for Making Black Box Models Explainable. North Carolina, United States: Lulu Press; 2019.
40. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Christensen EM, Bardram JE, et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. Int J Methods Psychiatr Res 2016 Dec;25(4):309-323. [doi: 10.1002/mpr.1502] [Medline: 27038019]
41. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. Bipolar Disord 2015 Nov;17(7):715-728. [doi: 10.1111/bdi.12332] [Medline: 26395972]
42. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]
43. Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, Busk J, et al. Reducing the rate and duration of Re-ADMISsions among patients with unipolar disorder and bipolar disorder using smartphone-based monitoring and treatment - the RADMIS trials: study protocol for two randomized controlled trials. Trials 2017 Jun 15;18(1):277 [FREE Full text] [doi: 10.1186/s13063-017-2015-3] [Medline: 28619114]

Abbreviations

EMA: ecological momentary assessment
HDRS: Hamilton Depression Rating Scale
MTL: multi-task learning
RADMIS: reducing the rate and duration of readmissions among patients with unipolar disorder and bipolar disorder using smartphone-based monitoring and treatment
RMSE: root mean squared error
YMRS: Young Mania Rating Scale

Edited by G Eysenbach; submitted 13.06.19; peer-reviewed by D Hidalgo-Mazzei, H Hassani, M Ostacher, A Graham; comments to author 17.09.19; revised version received 25.10.19; accepted 17.12.19; published 01.04.20.

Please cite as:
Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Vedel Kessing L, Winther O
Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach
JMIR Mhealth Uhealth 2020;8(4):e15028
URL: https://mhealth.jmir.org/2020/4/e15028
doi:10.2196/15028
PMID:32234702

Original Paper

Efficacy of a Mobile Texting App (HepTalk) in Encouraging Patient Participation in Viral Hepatitis B Care: Development and Cohort Study

Chul Hyun[1], MD, PhD; Joseph McMenamin[2], MD, JD; Okhyun Ko[3], BS; Soonsik Kim[3], MPH

[1]The Center for Viral Hepatitis, Englewood, NJ, United States

[2]W Medical Strategy Group, Englewood Cliffs, NJ, United States

[3]Korean Community Services Public Health and Research Center, New York, NY, United States

Corresponding Author:
Chul Hyun, MD, PhD
The Center for Viral Hepatitis
35 Van Nostrand Avenue
Englewood, NJ, 07631
United States
Phone: 1 845 661 1477
Email: chulhyunmd@gmail.com

Abstract

Background: Chronic hepatitis B virus (HBV) infection is a major cause of liver-related morbidity and mortality among Asian Americans in the United States. Despite the available resources, a majority of HBV-infected individuals are not able to access adequate health care owing to numerous barriers.

Objective: This study aimed to assess the efficacy of a newly developed mobile texting app (*HepTalk*) in overcoming these barriers and improving patient engagement and health care access among HBV-infected and nonimmune individuals.

Methods: *HepTalk* was employed for two-way communication between participants and patient navigators. A total of 82 Korean American participants who were either HBV infected or nonimmune to HBV, identified from a community hepatitis B campaign in New York, were enrolled in the study. After informed consent was obtained, both the frequency and themes of the text messages were evaluated. The effects of this communication on linkage to care at the end of the 6-month intervention period were analyzed and discussed.

Results: On average, patient navigators sent and received 14 and 8 messages per participant, respectively, during the 6-month period. The themes of the messages were similar to the following 4 categories: finding providers, scheduling appointments with providers, health education, and financial issues. Of the 82 participants, 78 were linked to care within 6 months (a 95% linkage rate).

Conclusions: *HepTalk* may be employed as an effective and strategic tool to facilitate communicative interaction between patients and patient navigators or health care providers, thereby improving patient engagement and health care access.

(JMIR Mhealth Uhealth 2020;8(4):e15098) doi:10.2196/15098

KEYWORDS

chronic hepatitis B; hepatitis B virus infection; linkage to care; mobile texting app; remote consultation

Introduction

Background

Patient communication with health care providers (HCPs) is crucial for positive health outcomes. Numerous barriers, however, can hinder the process of accurate and timely communication between patients and HCPs. For example, often, well-meaning HCPs provide patients medically accurate information, but unintentionally phrase it in a way that those with limited English proficiency may have trouble understanding, thus failing to adequately foster patients' engagement with their care [1-3]. The high cost of visiting a physician may deter some patients from seeking or keeping appointments [4-7]. Outside of scheduled appointments, HCPs have only limited methods of providing patients with information and educational materials related to their health problems. Furthermore, patients may be located too far from

HCPs and thus are unable to access care when the need arises [8,9]. Accordingly, there is a need for efficient communication systems or methods that can facilitate two-way communication between patients and HCPs.

Given the need for linguistically and culturally competent care in today's diverse community, communication between patients and HCPs can be even more challenging when it involves patients and HCPs of different races and ethnicity [3,10]. Chronic hepatitis B (CHB) is a major cause of liver-related mortality and morbidity worldwide, causing numerous clinical challenges in adequate treatment and management. CHB is also a good example of health disparities whose management is complicated by a lack of adequate communication between patients and HCPs speaking different languages and from different cultures [10-12].

There is a marked ethnic and racial disparity in the prevalence of CHB in the United States. Of 2 million Americans with CHB, a majority consists of immigrants from Asia, Africa, and other parts of the world [12,13]. Specifically, 3% to 10% of Asian Americans, of whom a large majority are immigrants from China, Korea, and Southeast Asia, have chronic hepatitis B virus (HBV) infection compared with less than 0.2% of non-Hispanic white Americans. CHB management is especially critical in Asian Americans for a number of reasons. A majority of Asian Americans with CHB have been infected since their birth and have a high probability of remaining chronically infected. Among all population groups in the United States, Asian Americans are most likely to develop liver cancer [10]. Despite these risks, many people with CHB are not accessing care largely because of the absence of symptoms and the barriers that prohibit their access [12,13]. As CHB is prevalent among immigrant populations, many of whom are not proficient in the English language, serious language barriers impede communication with providers and accessing health care in the United States [10-12]. Two major issues burden CHB management in the United States: (1) a significant percentage of the population at risk is not vaccinated; and (2) a majority of chronically HBV-infected individuals, many of whom may require antiviral treatment, are not currently accessing care.

We have previously investigated the use of mobile text messaging in facilitating the connection between HCPs and individuals with CHB or at risk for it [14]. In this pilot study involving 32 participants, the frequency and contents of the messages between participants and patient navigators were evaluated. The results of the study, which showed a strong positive outcome in conduits to care after only a 3-month intervention period, suggest that mobile text messaging intervention could provide a platform for patients to engage with HCPs, thereby potentially improving their access to health care and their understanding of their own needs.

Objective

To further examine the efficacy of mobile texting in engaging patients in collaborating in their own viral hepatitis B care, we developed a text messaging app, *HepTalk*. We evaluated the effects of its use in two groups of individuals similar to those in our previous study: people with CHB, who are not currently accessing care and thus need to see HCPs for further evaluation; and nonimmune individuals who need vaccination at a health care facility.

Methods

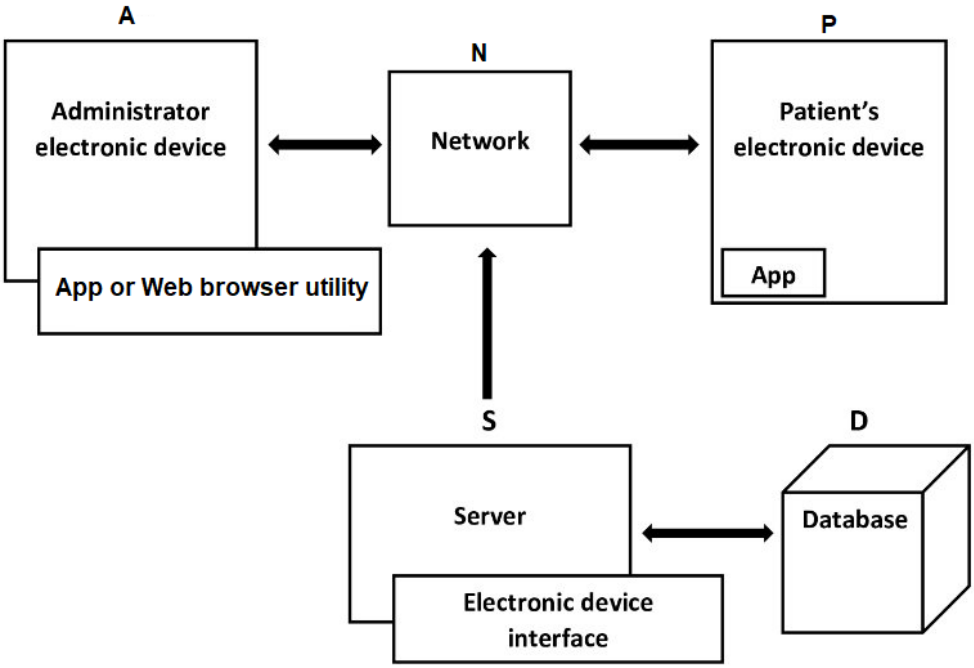
App Development

A mobile texting app called *HepTalk* was developed for a communication channel resistant to hacking that facilitates two-way communication between patients and HCPs. *HepTalk* permits encrypted texting between providers and consenting patients. Data so generated are retained throughout the period when patients are connected to providers and for 1 additional year, after which they are destroyed. Typically, the communication channel *HepTalk* is implemented as an app on an electronic device such as a smartphone or other computing device and may be used to facilitate consultation, patient navigation, and education. It may also be used for instant message or texting communication between patients and HCPs. In this manner, the secure communication channel may educate, engage, and empower patients and improve their access to HCPs.

Hacking-Resistant Communication Channel

As shown in the functional diagram in Figure 1, the system includes the following components: a server (S) having an electronic interface, an administrator electronic device (A) with an application on a Web browser utility, and a patient electronic device (P) such as a smartphone with an app. All these electronic devices (S, A, and P) are communicatively coupled via a network (N). S is configured to generate a hacking-resistant communication channel between the app on the administrator's electronic device and the app on the patient's smartphone. The communication channel is set up in accordance with settings stored on a database (D), and S stores a transcript of communication on the secure communication channel on D.

Figure 1. Functional diagram of a system providing a hacking-resistant communication channel through the app *HepTalk*. A: administrator electronic device; D: database; N: network; P: patient electronic device; S: server.



Downloading the App and Logging In

HepTalk is provided a link on the Apple and Android app stores. Patients may thus download the app onto their own smartphones. Once the patients download *HepTalk* app and provide log-in information to the app, the server (S) notifies the administrator (A) for approval. The administrator may approve or disapprove the patient. The approval process may include screening the patient to determine if he has the indicated medical condition. Once the patient is approved, the administrator may match the patient with an HCP (or patient navigator) and create a hacking-resistant communication channel.

Participant Recruitment

Participants in a community-based hepatitis B awareness campaign run in Queens, New York, between January 2017 and June 2017, were Korean American adults who were found to be either nonimmune (nonimmune group) to HBV or hepatitis B surface antigen seropositive (CHB group). Korean Americans refer to those with Korean ancestry or those born in Korea but living in the United States. The campaign consisted of community-based hepatitis B screening and education organized by the Center for Viral Hepatitis (CVH) and Korean Community Services. We identified a total of 75 nonimmune participants and 28 participants with CHB not linked to HCPs.

A total of 82 (61 from the nonimmune group and 21 from the CHB group) participants agreed to participate in this study. Before the study, participants signed an informed consent form available in both Korean and English. In addition, a self-administered survey was used to determine demographic characteristics. The items in the survey included gender, date of birth, country of birth, contact information, years of residence in the United States, and preferred language. All participants preferred to communicate in Korean. Once the participants logged into the app and were linked to patient navigators, text communication began, and the messages were counted and analyzed.

Data Collection and Analysis

All participants owned smartphones with text messaging capability. Once the participants were logged in, texting communication was initiated by the patient navigators. Members of the nonimmune group were sent an initial message advising them to receive hepatitis B vaccination at a local health care facility and asking at the same time if the participant needed help in locating such a facility. Members of the CHB group were also sent initial messages advising them to see physicians for further evaluation of HBV infection. The participants in the latter group were also given a list of local health care facilities and doctors' offices where they could potentially receive further care.

The frequency of text messaging to and from the patient navigators was recorded. The contents of the messages were then analyzed by classifying them under 1 of the 4 following thematic categories: medical access (finding physicians or health facilities), reminders and schedules, financial costs and insurance, and health information and education. Medical access referred to finding physicians or health facilities the participants felt comfortable visiting for evaluation or vaccination. Reminders and schedules include notifications to alert the participants to make and keep appointments. Financial costs and insurance referred to medical expense and having or lacking a health insurance plan. Finally, health information and education included all messages providing health information on CHB and prevention of HBV infection by vaccination. These included updated guidelines and recommendations on CHB available from community centers, and their current activities, and printed and Web resources (eg, US Centers for Disease Control and Prevention).

Linkage to Care

Outcome in linkage to care (LTC) was evaluated at the end of 6 months. In the nonimmune group, the participants who received at least two hepatitis B vaccinations were considered *linked*. In the CHB group, the participants who saw a physician at least once for further evaluation of their CHB status were also considered linked to care. All the CHB participants who were linked to physicians had a hepatitis B DNA test, and their attendance at scheduled office visits was confirmed by the patient navigator.

Patient Navigator

The patient navigator program was used to provide participants with specific information on the prevention and management of CHB and to link them to HCPs who have expertise in hepatitis B care within their community. Patient navigators are health care professionals who can assist the participants in finding their way through the health care system and who work to overcome obstacles by identifying and providing resources for Korean Americans with socioeconomic and communication barriers.

This study involved 2 patient navigators. One held a Bachelor of Science degree in public health, and the other held a Master of Public Health degree. They were employees of community service organizations and were familiar with clinicians and other health care resources within the community. The participants in the nonimmune group were provided with a list of health care facilities where they could be vaccinated. All participants in the CHB group were given a list of community HCPs with expertise in hepatitis B. Throughout all the participant office visits and appointments, patient navigators used *HepTalk* to guide the participants through the process of evaluation. Communication between subjects and patient navigators relied on text messaging (ie, *HepTalk*), and the patient navigators kept detailed records of all communication with the subjects.

Ethics Approval and Consent to Participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee. The Investigative Committee on Clinical Research, institutional review board of Holy Name Medical Center, Teaneck, New Jersey, approved this study.

Availability of Data and Materials

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Results

Demographic Characteristics

Between July and December 2017, we registered a total of 82 individuals and followed them up with *HepTalk* texts. Table 1 shows the demographic characteristics of both groups. The nonimmune group consisted of 61 individuals requiring HB vaccination. Of these 61 participants, 28 were men and 33 women; the average age was 47 years. The CHB group consisted of 21 individuals who had been found to be hepatitis B surface antigen seropositive. There were 10 men and 11 women; the average age was 54 years. All the participants were Korea-born immigrants who preferred to communicate in Korean. Once registered on *HepTalk*, they were assigned to patient navigators, with whom all the participants subsequently communicated through *HepTalk*. The nonimmune group was advised to go to recommended medical facilities to receive vaccinations, whereas the CHB group was advised to see providers for further assessment and management of their CHB. All the communication between the participants and the patient navigator took place through *HepTalk* during the 6-month period between July 2017 and December 2017.

Table 1. Demographic characteristics and linkage to care rate after a 6-month intervention period.

Groups	Vaccine group (n=61)	Chronic hepatitis B group (n=21)
Sex, n (%)		
Male	28 (46)	10 (48)
Female	33 (54)	11 (52)
Age (years)		
Mean	47	54
Range, n (%)		
21-29	8 (13)	0 (0)
30-39	6 (10)	1 (5)
40-49	18 (30)	4 (19)
50-50	19 (31)	11 (52)
60-69	8 (13)	5 (24)
70-79	2 (3)	0 (0)

Communication Between the Participants and Patient Navigators

Tables 2 and 3 demonstrate the number and frequency of messages sent and received by patient navigators. All participants received at least one message sent by the assigned patient navigator at the beginning of the 6-month intervention period. The patient navigators sent a total of 852 messages to members of the nonimmune group. All 61 nonimmune participants responded with a total of 463 messages, averaging 7.6 messages per respondent (Table 2). Patient navigators sent 299 messages to the CHB group, consisting of 21 participants. All, except 5 participants, responded with a total of 157 messages, averaging 9.8 messages per texting respondent (Table 3).

Table 2. Number of messages sent and received for group 1.

Number of messages	Number of participants receiving messages	Number of participants sending messages
1-5	15	38
6-10	19	9
11-15	9	2
16-20	6	4
21+	12	8

Table 3. Number of messages sent and received for group 2.

Number of messages	Number of participants receiving messages	Number of participants sending messages
0	0	5
1-5	7	7
6-10	3	3
11-15	2	1
16-20	3	4
21+	6	1

Frequency and Common Themes of the Communication

The contents of the messages received by both groups were categorized into 1 of the following 4 themes: finding medical access, reminders and scheduling, education and health information, and financial and insurance.

The greatest proportion of the messages in both groups was related to medical access, followed by the other 3 categories, whose proportions were similar in frequency. The proportion of the messages pertaining to HCPs (medical access) and scheduling was greater than half of all the messages received (282/461, 61.2% and 85/147, 57.8% in the nonimmune and CHB groups, respectively).

However, the frequency of communication was markedly different between the nonimmune and CHB groups. Although three-fourths of the total communication in the nonimmune group was carried out within 3 months of intervention, the same proportion of the total communication in the CHB group was carried out within the first month. This difference may be related to the fact that to be considered *linked* in the nonimmune group, a participant had to have 2 vaccinations 1 month apart, whereas to be considered *linked* in the CHB group, a participant required only 1 visit to an HCP. The linkage in the CHB group was rapid, with a mean of 14 days. The most common themes of the messages involved medical access and reminders in both groups.

High Level of Achievement in Linkage to Care

At the end of the 6-month intervention period, 78 of 82 individuals were linked to care, demonstrating an overall linkage rate of 95%. Specifically, 58 of 61 individuals in the nonimmune group received their first 2 vaccinations within an average period of 65 days. In the CHB group, 20 of 21 participants made 2 or 3 visits to their providers' offices for the evaluation of their CHB within the 6-month period (Table 4). Those participants who accessed care in the CHB group made their first visit to HCPs within an average period of 14 days of *HepTalk* intervention.

Table 4. History of hepatitis B virus infection in the hepatitis B surface antigen-seropositive participants linked to care.

Years since hepatitis B virus was diagnosed	Number of participants linked to care (n=20), n (%)
<5	2 (10)
5-10	3 (15)
11-20	2 (10)
>20	13 (65)

History of Hepatitis B Virus Infection in Chronic Hepatitis B Group

We evaluated how many of the 20 linked CHB participants had known their infection status before the current screening and *HepTalk* intervention. As shown in Table 4, 18 of the 20 participants had been aware of their infection for more than 5 years before the texting intervention began. Only 5 of these patients had seen physicians at least once in the past, but they did not return for continued care.

It is noteworthy that the remaining 15 of these linked patients had known their HBV status for over a decade and yet had not seen an HCP for further evaluation before their current linkage.

Discussion

Principal Findings

Secure communication via texting between patients and HCPs may be used for consultation, patient engagement and education, and direct instant messaging [15-17]. Much evidence supports the use of texting between patients and HCPs and positive health outcomes. Text messaging can provide laboratory results, reminders of appointments, medication administration and flu vaccination, and other services [18-20]. Specifically, text messaging has been shown to improve adherence to medication and attendance at medical appointments among HIV and other chronic disease patients [21,22]. Text4Health projects have also helped to engage the underserved population and improved their health [23]. SmokeFreeText, for instance, more than doubled the smoking cessation rate among teens by texting smoking cessation messages to them [24]. Furthermore, the Text4Baby Campaign has helped expectant mothers to receive crucial prenatal care resources, thereby fostering the safe delivery of their babies [25].

As texting has become a progressively effective tool in patient engagement, the number of health apps has dramatically increased during the past decade. As of 2013, there were more than 1700 diabetes mellitus apps in all the app stores combined (Apple app store, Google Play store, and Windows) [26]. These mobile health apps in smartphones can collect and deliver health care data and monitor patients' vital signs in real time [27,28].

The results of this study demonstrate that *HepTalk* can be employed to boost patient engagement and improve outcomes in hepatitis B care. This study further supports the finding of our previous study, which suggested that a form of mobile texting combined with the patient navigator program facilitated communication between the patients and HCPs and enhanced LTC [14]. *HepTalk* provided an effective communication mechanism through which patient navigators were able to guide participants to appropriate health care resources. The benefits of *HepTalk* are enormous, including communication speed, accessibility, and reduced patient expense. *HepTalk* is not limited by geographic boundaries and is able to help people lacking transportation. Last but not least, as all the communication was in the patients' native language, *HepTalk* could also overcome linguistic and cultural barriers.

Rapid and Effective Patient Engagement on Patients' Terms

Our results also demonstrated a high level of texting communication between the participants and patient navigators. First, patient navigators sent a total of 1151 messages to all 82 participants, averaging 14.0 messages per participant. Second, patient navigators received a total of 620 messages from 77 participants, averaging 8.1 messages per participant (Tables 2 and 3). All but 5 participants in the CHB group replied to patient navigators through *HepTalk*, demonstrating a 94% response rate from the participants. The 5 participants who did not reply to the patient navigators were sent 2 repeat messages advising them to see a provider, providing a list of community HCPs. Of these 5 participants, 4 were finally linked to providers in the community. Thus, texting communication took place on the participants' terms and allowed them the flexibility to interact with the patient navigators in the manner they felt most comfortable with. Most of the messages were read and replied to within a few hours, further demonstrating the likelihood and speed of interaction.

The mean number of messages received and sent by each participant in the CHB group during the 6-month period was 14 and 10, respectively. Very few messages were sent or received once the LTC was confirmed. As the mean length of time needed to establish LTCs was only 14 days, linkage in the CHB group was rapid, suggesting patients' motivation to engage with care. The observed high rate of LTC is remarkable, especially given that the majority of the patients in the CHB group had been diagnosed years before this campaign and had not previously accessed HCPs (Table 4).

The efficacy of *HepTalk* in participant engagement observed in this study is attributable to successful communicative interaction between the patient navigators and participants. The themes noted in the texts were diverse and covered important topics that typically would be covered between patients and HCPs in their offices. Although finding HCPs was a slightly dominant theme, the frequency of the other 3 themes, related to making and keeping appointments with HCPs, health education, and financial factors, was equal; the number of text messages during the 6-month period was the same in all, that is, 3 messages.

Positive Health Behavior Change

It is noteworthy that the majority of the CHB patients in this study had been aware of their infections for many years before this study. Of 20 CHB patients linked to care, 18 had known their infection status for 5 years or longer, yet they had not seen HCPs for continuing care (Table 4). These results are congruent with the previous findings that only a minority of HBV-infected people could access care. According to these previous investigations, only about 40% of subjects with CHB screened in a community setting were successfully linked to care [10-13].

How did the *HepTalk* app, then, help the HBV infected participants to change their minds and access care they had failed to do so in the past? *HepTalk* combined with the patient navigator program in this study helped to overcome barriers caused by language and culture, as all texting took place in the

Korean language. Furthermore, as the texting cost nothing, patients faced no financial burden. Finally, patients did not have to spend much time in finding HCPs and health care facilities and arranging medical appointments, thereby diminishing the burden of finding providers. Once patients were able to understand through effective communication (eg, same language and culture) the need for vaccination and further evaluation, they were motivated to access care. Thus, *HepTalk* helped to drive positive behavioral change in the HBV-infected participants who might otherwise have failed to address their problems in the past.

It has been well established that health behaviors can have an impact on individuals' physical and mental health and quality of life [29,30]. Behaviors that relate to accessing care and behaviors that involve care seeking and adherence to treatment, for instance, can be crucial in determining the effective management of CHB and numerous other chronic diseases. By overcoming various barriers, *HepTalk* provided an environment designed to empower and direct the individuals with CHB to change their behaviors.

Beneficial Features of HepTalk

The benefits of *HepTalk* are substantial, including accessibility, flexibility, ease and speed of communication, and reduced patient expense. *HepTalk* is designed specifically as a hepatitis B app to facilitate hepatitis B care. Patients can easily identify the app on the smartphone screen, allowing quick access. *HepTalk* does not rely on cellphone data plans; it can be used with Wi-Fi alone. This *Wi-Fi–compatible* feature is not only convenient as Wi-Fi is readily available in cities but also can help to reduce cost, as smartphone plans charge for texting. *HepTalk* is thus more affordable than SMS. Unlike most other apps, *HepTalk* allows communication between a patient navigator (or HCP) and only 1 patient at a time. No second party can be invited, protecting the privacy of the communication. Administration, however, can reach out to any number of patients at once to make announcements. *HepTalk* can also provide patients with unlimited access to educational material and information outside of their scheduled appointments. In addition, *HepTalk* has many features that regular SMS or texting apps lack, including an ability to file patients' demographics and categorize the patients (eg, nonimmune and CHB groups). These capabilities allow easy and convenient access and communication between the participants and the patient navigators. Furthermore, although not available worldwide, *HepTalk* is not limited by geographic boundaries and can serve people lacking transportation. Last but not least, *HepTalk* can also overcome many issues related to cultural competence. Patients with a specified culture and language can connect with patient navigators or HCPs who share the same culture and language, thus allowing optimal communication between the parties.

One difficult aspect of *HepTalk* use from patients' perspectives, however, is the downloading and log-in process. Unlike most other apps, *HepTalk* log-in requires approval by administration. Approval requires a process of screening the patient to determine if he or she has an appropriate medical condition. Although a few patients had difficulty with log-in, this screening process may be considered more of an advantage than a disadvantage because it allowed distinguishing individuals who need the *HepTalk* service from those who do not.

Limitations of This Study

We have to consider important shortcomings of this study. First, the sample size (n=82) may not be large enough to reflect results to be expected among Korean Americans, in general, in the United States, whose proportion is much less than that of other ethnic groups. Although it is well known that age, education, and socioeconomic status influence patients' health care access behavior and their uptake and adaptation of mobile technology, we have not assessed these factors among the participants enrolled in the study. Second, this study did not directly compare the effects of *HepTalk* texting on patients who refrained from *HepTalk* texting. Although we have not conducted a study where we employed a control group denied access to *HepTalk* texting, this study and previous studies have demonstrated a significant lack of LTC in people with CHB in the absence of any mobile texting interventions [10-13]. Third, our study did not compare *HepTalk* with regular SMS texting, thus limiting specific comments on direct comparison between the two. However, it should be noted that there are unique features that *HepTalk* offers, and SMS or other apps do not, which may make *HepTalk* more desirable for patients' use (see the Discussion above). Fourth, we considered CHB participants who saw HCPs at least once during the 6-month period for evaluation with DNA testing to be linked to care. However, it may be premature to consider these participants linked because some of these *linked* participants may not return to HCPs for follow-up in the future. Thus, although the texting intervention was significantly longer in this study compared with the intervention period employed in our previous study [14], studies with an intervention period greater than 6 months may be preferable to assess the sustainability of the efficacy of the *HepTalk* communication. This is particularly important because we have noted that a small but significant portion of CHB patients with long histories of infection saw HCPs once but did not sustain their LTC. Finally, it should be noted that *HepTalk* alone would not have led to the observed successful LTC. The exact role and contribution of the patient navigator program to the overall success of *HepTalk*-mediated LTC need to be better defined. For instance, the benefit arising from the use of *HepTalk* alone cannot be distinguished from the benefit arising from the linguistic and cultural competence of patient navigator program implementation.

In conclusion, the results of this study suggest that *HepTalk* can serve as an effective communication tool that may empower patients to access health care. The patient navigator program and mobilization of local HCPs who have expertise in CHB were also crucial to successful linkage of the patients from the testing site to providers. Mobile texting combined with community-based patient navigation programs such as those employed in this study may be implemented in other minority ethnic populations to enhance hepatitis B care or, perhaps, the care of patients with other chronic illnesses. Future studies with *HepTalk* or other instant messenger apps evaluating a larger population at different economic and sociocultural levels for longer periods would be needed to better assess the efficacy

and sustainability of text messaging intervention in the enhancement of LTC in hepatitis B care.

Acknowledgments

The authors acknowledge the work by members of and volunteers at the CVH and the Asian Liver Center of Holy Name Medical Center. The study was supported by Gilead Foundation. The funder partially financed the process of data collection and analysis. The funder had no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

CH conceived the study, participated in the design of the study, and drafted the manuscript. SK and OK participated in the design of the study and performed the statistical analysis. JM participated in the analysis and interpretation of the data and manuscript development. All authors read and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Chu C. Cross-cultural health issues in contemporary Australia. Ethn Health 1998;3(1-2):125-134. [doi: 10.1080/13557858.1998.9961854] [Medline: 9673469]
2. Jacobs E, Chen AH, Karliner LS, Agger-Gupta N, Mutha S. The need for more research on language barriers in health care: a proposed research agenda. Milbank Q 2006;84(1):111-133 [FREE Full text] [doi: 10.1111/j.1468-0009.2006.00440.x] [Medline: 16529570]
3. Meuter RF, Gallois C, Segalowitz NS, Ryder AG, Hocking J. Overcoming language barriers in healthcare: A protocol for investigating safe and effective communication when patients or clinicians use a second language. BMC Health Serv Res 2015 Sep 10;15:371 [FREE Full text] [doi: 10.1186/s12913-015-1024-8] [Medline: 26357948]
4. Parikh PB, Yang J, Leigh S, Dorjee K, Parikh R, Sakellarios N, et al. The impact of financial barriers on access to care, quality of care and vascular morbidity among patients with diabetes and coronary heart disease. J Gen Intern Med 2014 Jan;29(1):76-81 [FREE Full text] [doi: 10.1007/s11606-013-2635-6] [Medline: 24078406]
5. DeVoe JE, Fryer GE, Phillips R, Green L. Receipt of preventive care among adults: insurance status and usual source of care. Am J Public Health 2003 May;93(5):786-791. [doi: 10.2105/ajph.93.5.786] [Medline: 12721145]
6. Wilper AP, Woolhandler S, Lasser KE, McCormick D, Bor DH, Himmelstein DU. Hypertension, diabetes, and elevated cholesterol among insured and uninsured US adults. Health Aff (Millwood) 2009;28(6):w1151-w1159. [doi: 10.1377/hlthaff.28.6.w1151] [Medline: 19843553]
7. Garfield R, Orgera K, Damico A. Kaiser Family Foundation. 2019 Jan 25. The Uninsured and the ACA: A Primer - Key Facts about Health Insurance and the Uninsured Amidst Changes to the Affordable Care Act URL: https://www.kff.org/uninsured/report/the-uninsured-and-the-aca-a-primer-key-facts-about-health-insurance-and-the-uninsured-amidst-changes-to-the-affordable-care-act/ [accessed 2019-10-16]
8. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
9. Kullgren JT, McLaughlin CG, Mitra N, Armstrong K. Nonfinancial barriers and access to care for US adults. Health Serv Res 2012 Feb;47(1 Pt 2):462-485 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01308.x] [Medline: 22092449]
10. Philbin MM, Erby LA, Lee S, Juon H. Hepatitis B and liver cancer among three Asian American sub-groups: a focus group inquiry. J Immigr Minor Health 2012 Oct;14(5):858-868 [FREE Full text] [doi: 10.1007/s10903-011-9523-0] [Medline: 21901445]
11. Hutton DW, Tan D, So SK, Brandeau ML. Cost-effectiveness of screening and vaccinating Asian and Pacific Islander adults for hepatitis B. Ann Intern Med 2007 Oct 2;147(7):460-469. [doi: 10.7326/0003-4819-147-7-200710020-00004] [Medline: 17909207]
12. Hu KQ, Pan CQ, Goodwin D. Barriers to screening for hepatitis B virus infection in Asian Americans. Dig Dis Sci 2011 Nov;56(11):3163-3171. [doi: 10.1007/s10620-011-1840-6] [Medline: 21861105]
13. Cohen C, Holmberg S, McMahon B, Block JM, Brosgart CL, Gish RG, et al. Is chronic hepatitis B being undertreated in the United States? J Viral Hepat 2011 Jun;18(6):377-383. [doi: 10.1111/j.1365-2893.2010.01401.x] [Medline: 21143343]
14. Hyun CS, Yoon S, Kim S, Lee S, McMenamin J. Patient navigator facilitated text messaging intervention improves linkage to care in viral hepatitis B: a pilot study. JMTM 2017 Aug;6(2):11-18. [doi: 10.7309/jmtm.6.2.2]
15. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]

16. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
17. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
18. Hammonds T, Rickert K, Goldstein C, Gathright E, Gilmore S, Derflinger B, et al. Adherence to antidepressant medications: a randomized controlled trial of medication reminding in college students. J Am Coll Health 2015;63(3):204-208. [doi: 10.1080/07448481.2014.975716] [Medline: 25338175]
19. McKenzie RB, Berquist WE, Foley MA, Park KT, Windsheimer JE, Litt IF. Text messaging improves participation in laboratory testing in adolescent liver transplant patients. J Particip Med 2015;7:pii: e7 [FREE Full text] [Medline: 26213633]
20. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
21. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012 Mar 14(3):CD009756 [FREE Full text] [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
22. Arya M, Kallen MA, Street RL, Viswanath K, Giordano TP. African-American patients' preferences for a health center campaign promoting HIV testing: an exploratory study and future directions. J Int Assoc Provid AIDS Care 2014;13(6):488-491 [FREE Full text] [doi: 10.1177/2325957414529823] [Medline: 24739209]
23. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. J Am Med Assoc 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]
24. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
25. Evans WD, Abroms LC, Poropatich R, Nielsen PE, Wallace JL. Mobile health evaluation methods: the Text4baby case study. J Health Commun 2012;17(Suppl 1):22-29. [doi: 10.1080/10810730.2011.649157] [Medline: 22548595]
26. Isaković M, Sedlar U, Volk M, Bešter J. Usability pitfalls of diabetes mHealth apps for the elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
27. Dias D, Cunha JP. Wearable health devices-vital sign monitoring, systems and technologies. Sensors (Basel) 2018 Jul 25;18(8):pii: E2414 [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]
28. Di Rienzo M, Rizzo F, Parati G, Brambilla G, Ferratini M, Castiglioni P. MagIC System: A New Textile-Based Wearable Device for Biological Signal Monitoring. Applicability in Daily Life and Clinical Setting. In: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. 2006 Presented at: IEMBS'06; January 17-18, 2005; Shanghai, China. [doi: 10.1109/iembs.2005.1616161]
29. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
30. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]

Abbreviations

CHB: chronic hepatitis B
CVH: Center for Viral Hepatitis
HBV: hepatitis B virus
HCP: health care provider
LTC: linkage to care

Edited by G Eysenbach; submitted 19.06.19; peer-reviewed by T Muto, Y Qiu, J Wallace; comments to author 16.07.19; revised version received 10.02.20; accepted 29.02.20; published 01.04.20.

Please cite as:
Hyun C, McMenamin J, Ko O, Kim S
Efficacy of a Mobile Texting App (HepTalk) in Encouraging Patient Participation in Viral Hepatitis B Care: Development and Cohort Study
JMIR Mhealth Uhealth 2020;8(4):e15098
URL: https://mhealth.jmir.org/2020/4/e15098
doi:10.2196/15098
PMID:32234704

Original Paper

Evaluating the Feasibility of Frequent Cognitive Assessment Using the Mezurio Smartphone App: Observational and Interview Study in Adults With Elevated Dementia Risk

Claire Lancaster[1], BSc, MSc, PhD; Ivan Koychev[2], PhD, MRCPsych; Jasmine Blane[2], BSc, MSc; Amy Chinner[2], BSc; Leona Wolters[2], BSc, MSc; Chris Hinds[1], BA, DPhil, MSc

[1]Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

[2]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Claire Lancaster, BSc, MSc, PhD
Nuffield Department of Population Health
University of Oxford
Big Data Institute
Old Road Campus
Oxford, OX3 7LF
United Kingdom
Phone: 44 1865 743893
Email: claire.lancaster@bdi.ox.ac.uk

Abstract

Background: By enabling frequent, sensitive, and economic remote assessment, smartphones will facilitate the detection of early cognitive decline at scale. Previous studies have sustained participant engagement with remote cognitive assessment over a week; extending this to a period of 1 month clearly provides a greater opportunity for measurement. However, as study durations are increased, the need to understand how participant burden and scientific value might be optimally balanced also increases.

Objective: This study explored the *little but often* approach to assessment employed by the Mezurio app when prompting participants to interact every day for over a month. Specifically, this study aimed to understand whether this extended duration of remote study is feasible, and which factors promote sustained participant engagement over such periods.

Methods: A total of 35 adults (aged 40-59 years) with no diagnosis of cognitive impairment were prompted to interact with the Mezurio smartphone app platform for up to 36 days, completing short, daily episodic memory tasks in addition to optional executive function and language tests. A subset (n=20) of participants completed semistructured interviews focused on their experience of using the app.

Results: Participants complied with 80% of the daily learning tasks scheduled for subsequent tests of episodic memory, with 88% of participants still actively engaged by the final task. A thematic analysis of the participants' experiences highlighted schedule flexibility, a clear user interface, and performance feedback as important considerations for engagement with remote digital assessment.

Conclusions: Despite the extended study duration, participants demonstrated high compliance with the schedule of daily learning tasks and were extremely positive about their experiences. Long durations of remote digital interaction are therefore definitely feasible but only when careful attention is paid to the design of the users' experience.

(JMIR Mhealth Uhealth 2020;8(4):e16142) doi:10.2196/16142

KEYWORDS

technology assessment; cognition; smartphone; mhealth; mobile phone; Alzheimer disease; early diagnosis; feasibility study; ecological momentary assessment

Introduction

Smartphones will benefit the development of clinical interventions for Alzheimer disease (AD) by enabling detailed cognitive phenotyping at scale [1]. Rich, descriptive, and high-frequency data from every cognitive response, taken from precise on-device sensors, will allow us to move beyond the standard summary metrics of reaction time and accuracy provided by conventional neuropsychological assessment [2], facilitating the detection of subtle performance differences [3]. In comparison with repeat in-clinic tests, remote digital data collection using mobile technologies uniquely enables frequent, longitudinal assessment, with a significantly lower participant burden, lower administration cost, and higher potential for deployment at scale [4,5]. By facilitating a *little but often* approach to assessment, smartphones increase the reliability of cognitive profiling, reducing the impact of within-participant variability, eg, in association with daily stressors, mood, and sleep [6-8]. Together, this maximizes the value of smartphones for the detection of preclinical or prodromal disease, encouraging greater and more targeted recruitment to clinical trials for early AD [3,9]. Frequent measurement allows acute longitudinal change in cognitive function to be accurately assessed [10]; therefore, smartphones can contribute to more exact monitoring of disease progression and therapeutic response [1]. Although digital technology promises clear benefits for dementia research, usability is the greatest challenge for these tools to be widely adopted [1]. This research will consequently investigate participants' views following the longitudinal use of a smartphone measurement platform.

The feasibility of remote, digital cognitive assessment in older adults is commonly assessed according to study compliance [10-13]. Emphasis has been placed on deploying short, frequent assessments, with adherence to a schedule of self-reported function and active cognitive tasks in the range of 72% to 82% when older adults with no objective cognitive impairment were prompted to interact 5 times a day for a week [14,15]. Further support for the acceptability of high-frequency *microinteractions* (<1 min) is evidenced by 77% compliance in older adults prompted via a smartphone notification to complete four assessments at random intervals across each day for a week [16]. However, to maximize the opportunity of remote, digital phenotyping, it is important to establish whether compliance can be sustained over substantially longer study durations (eg, from several weeks up to several years). Reports of 8.1% participant attrition to a 6-month schedule of remote cognitive tests in individuals at increased familial risk of dementia support the promise of digital tools for long-term assessment [17]; however, further research is needed to understand how to promote sustained participant engagement in digital health research.

Despite the significance of participant compliance for the value of research outcomes, there has been limited evaluation of which factors foster continued engagement with repeat, digital cognitive assessment. A focus group discussion between younger and older adults highlighted participant autonomy in scheduling, positive feedback, and specific instructions on how to approach each cognitive task as important factors for engagement [18]. More broadly, for sustaining long-term engagement (3-17 months) with a Web-based health platform, older adults highlighted personalized reminders from the platform, incorporating the tool into their daily routine, and observed variation or progress within the platform as important [19].

This study tested the feasibility of the Mezurio smartphone app, specifically the utility of this tool for significantly longer-term, high-frequency assessment (36 days) than the 7-day assessment periods explored previously [14-16], along with which factors contribute to successful participant engagement. Mezurio contains a collection of novel cognitive tasks, designed and built by the Mezurio research team to facilitate the detection of preclinical AD. These tasks measure long-term episodic memory, language, and executive function through a range of input modalities including voice, movement, and touch. In addition, the tasks follow a comparable *little but often* approach to remote assessment, as tested previously [14,20], with emphasis placed on providing participants autonomy to schedule tasks according to their daily routine. The developers have worked closely with lay older adults (including adults with mild cognitive impairment) to create a research experience anticipated to be both clear and engaging.

Feasibility was objectively tested according to study compliance and participant attrition across the baseline period of assessment, with the longitudinal follow-up at 6 and 12 months still ongoing. Semistructured interviews, focused on themes of approachability, acceptability, and engagement, were used to evaluate which factors contribute to participants' willingness to engage in daily digital cognitive assessment for approximately a month. This proof of concept provides the first in-depth evaluation of the feasibility of frequent, remote, and digital cognitive assessment in middle-aged adults, with the high proportion of individuals with a familial risk of AD making this a highly relevant sample for Mezurio's eventual use case as a tool for detecting preclinical disease [21,22]. Research outcomes will be essential for establishing app utility ahead of the inclusion of Mezurio in larger clinical trials for preclinical AD, as well as improving the use of digital tools in health research more broadly.

Methods

Participants

Table 1 reports the demographic characteristics of 35 volunteers (34/35, 97% non-Hispanic whites, aged 40-59 years) recruited via post or an email from the Oxford Health National Health Service Trust site of the larger PREVENT dementia program (n=68) [18]. Although a family history of dementia was not an inclusion criterion for this digital substudy, a high proportion (23/35, 66%) of participants reported a first-degree relative with dementia (15/35, 43% AD). The characteristics of the subset of participants (n=20) who completed a semistructured interview following their period of smartphone assessment are also shown in Table 1, with no significant group differences between participants providing feedback and those not completing the interview—independent groups *t* tests (2-tailed): age (*P*=.85)

and years of education (P=.68); gender (χ^2_1=.84, P=.36) and immediate family history of dementia (χ^2_2=3.30, P=.19).

Table 1. Demographic characteristics of participants.

Demographics	Total sample (n=35)	Interview subgroup (n=20)
Age (years), mean (SD)	52.57 (5.10)	52.20 (5.15)
Gender (female), n (%)	26 (74)	16 (80)
Years of education, mean (SD)	15.49 (2.74)	15.30 (3.16)
Family history, n (%)	23 (66)	15 (75)

The PREVENT Dementia Program

PREVENT dementia [23] is an ongoing prospective study that aims to investigate interactions between risk factors for dementia and traditional biomarkers in midlife. Participants were invited to join PREVENT dementia via a number of routes, including the ConCERT-D and *Join Dementia Research* databases, as well as via the study website and social media. Volunteers were screened at the time of initial consent into the PREVENT dementia program to ensure they met the following criteria: (1) aged 40 to 60 years, (2) no diagnosis of dementia according to the Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems criteria, and (3) no known contradiction to a magnetic resonance imaging scan.

Informed Consent

This study was ethically approved (University of Oxford Medical Sciences Inter-Divisional Research Ethics Committee: R48717/RE001) and is compliant with the Helsinki Declaration of 1975. Written consent was required upon entry to the study.

The Mezurio Smartphone App

Participants were asked to complete a selection of freely available cognitive tasks within the Mezurio smartphone app, each intended to take 5 min or less to complete. The Mezurio research team designed and built these tasks to measure three core cognitive domains: episodic memory (Gallery Game and Story Time), connected language (Story Time), and executive function (Tilt Task). Figure 1 provides a visual representation of all three tasks. Story Time and Tilt Task were only introduced once a subset of participants (23/35, 66%) had completed their baseline month of assessment, with subsequent recruits (12/35, 34%) asked to complete all three tasks. The first 23 participants were offered the opportunity to switch to this extended version of Mezurio for ongoing follow-up data collection at 6 and 12 months, with 19 participants opting to make this change. No data were collected on why participants did not choose to take part in the extended version of the app.

Figure 1. Cognitive tasks deployed within Mezurio: (A) Gallery Game, including the gold star animation presented following a successful learning iteration, (B) Story Time, and (C) Tilt Task, with the participant's remaining lives (or hearts) and number of levels completed (trophies) shown on the right.

Gallery Game

All participants (n=35) completed Gallery Game during the baseline period of assessment [24], which comprised regular cross-modal paired-associate learning tasks with subsequent tests of recognition and recall memory following ecologically relevant delays (for the current middle-aged population specified as 1, 2, 4, 6, 8, 10, or 13 days). Within each learning activity, participants were asked to encode distinct pairings of object photo–stimuli and touch screen *swipe* directions (left, right, or up), with the number of object-direction associations progressively increasing alongside iterative checks of immediate recall for pairings until the learning criterion was achieved—for more details, see the study by Lancaster et al [24]. Positive feedback (a gold star animation) was presented following each

learning iteration if immediate recall was 100% correct (see Figure 1). Incorrect responses were indicated by the object-direction association being cued so the participant could repeat the learning trial, followed by an immediate second test of recall. Participants did not receive explicit feedback on recognition or recall test performance. A number of practice sessions, analogous to the main Gallery Game learning as well as recognition and recall tests, were included at the start of the study schedule. Excluding practice days, participants were prompted to complete up to 22 learning tasks, in addition to the associated recognition and recall tests for encoded stimuli. Participants interacted with Gallery Game once a day for up to 29 days.

Story Time

Story Time tests *connected* language; participants were asked to narrate a short comic strip aloud while the device microphone was recording and then repeat the story from memory immediately and following an approximate 24-hour delay. No feedback on performance is provided by Mezurio. Participants were prompted to narrate six comic strips, in addition to the immediate and delayed verbal recall test associated with each. A total of 12 participants completed Story Time during their baseline period of Mezurio assessment.

Tilt Task

Tilt Task, a measure of executive function, relies on in-built device movement sensors. Participants were asked to tilt their phone to move the central cursor toward the next target in sequence, with the executive challenge of each presented sequence increasing with successive levels. Inaccurate responses were registered by the central cursor rebounding off the nontarget lure and participants losing one of their three lives. Positive feedback on performance was symbolized by the collection of trophies as participants progressed through the levels. If a participant does not reach the end of the level, the participant cannot move onto the next level.

Note, a number of participants were unable to complete this initial prototype of Tilt Task because of the required sensors (an accelerometer, a gyroscope, and a magnetometer) being absent in a range of Android devices. In addition, a number of participants also reported technical issues with the movement recognition software included in Tilt Task. Consequently, 9 participants were presented with Tilt Task at baseline, with performance data available for only 7 individuals. Subsequent versions of Mezurio screen device sensors when launching the app, with a bug-fix implemented to improve movement recognition during this task.

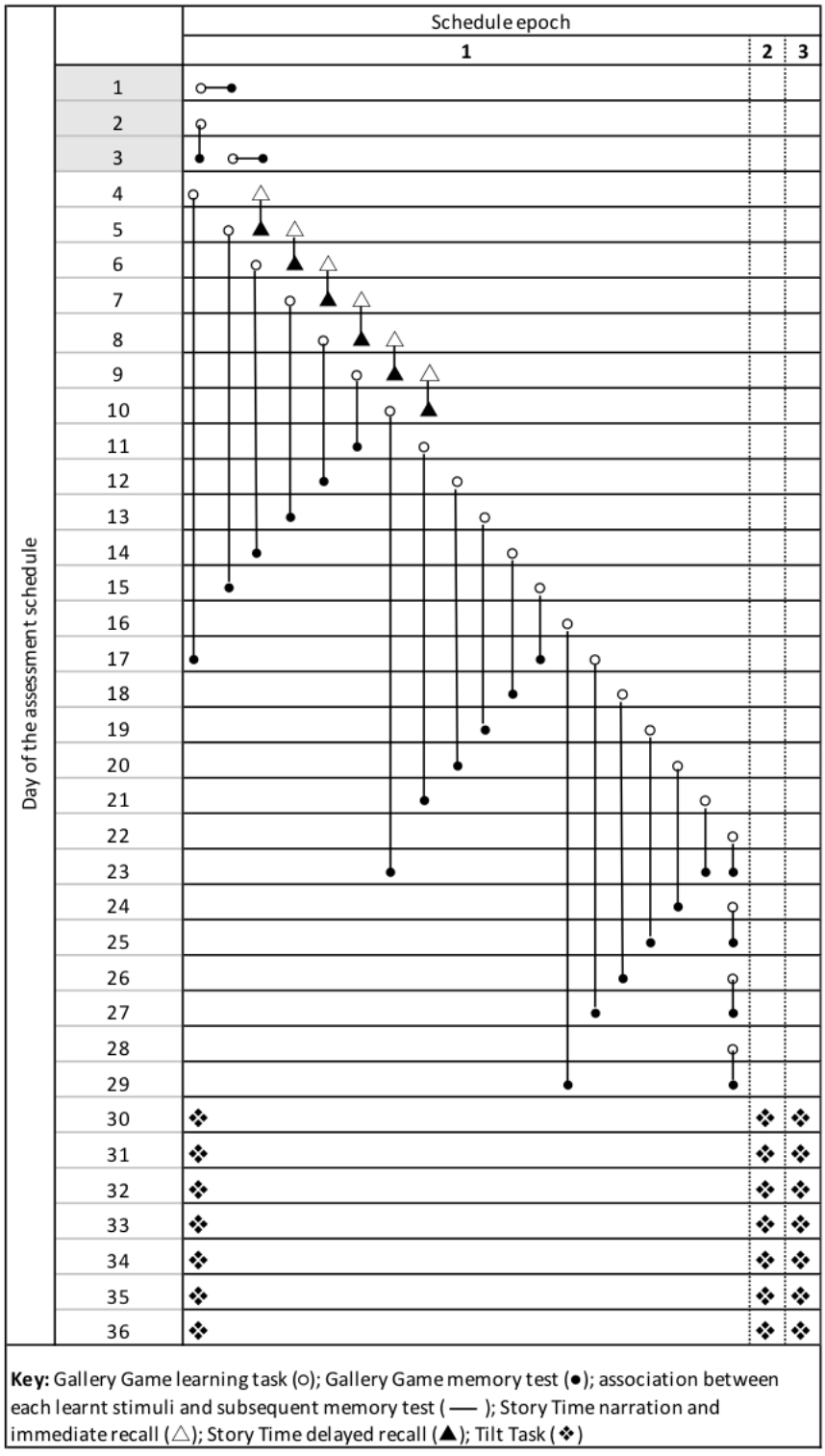
Procedure

Participants were invited to download Mezurio onto their personal smartphone (Apple or Android; 28/35, 80%); in cases where the participant's smartphone was running an incompatible operating system or the participant did not wish to use his or her own device, a number of Android devices were available for loan (n=7). The research team provided written instructions for download and a unique authentication ID.

Once installed, Mezurio prompted participants to complete regular cognitive tasks, following a schedule of interactions specified by the app. This schedule is study dependent and can be flexibly tailored according to the research aims and population, with cognitive tasks scheduled for a maximum of 36 days at baseline in this research, dependent on whether participants opted to complete the extended (Gallery Game, Story Time, and Tilt Task) or Gallery Game–only version of Mezurio (see Figure 2). The ongoing follow-up assessment at 6 and 12 months follows a comparable study schedule. The length of study schedule, along with the distribution of tasks within this active test period, was designed to limit daily participant burden while ensuring each individual was presented with multiple opportunities to complete each *microassessment*, anticipated to provide sufficient statistical power for subsequent analysis of cognitive outcomes.

Mezurio typically asked participants to interact once a day, with the exception of individuals opting to complete the extended version of the app, who were additionally prompted to complete Tilt Task thrice daily (see Figure 2). Mezurio encouraged participants to complete the cognitive tasks at the same time each day using local notifications scheduled by the app and displayed on the smartphone's home screen. A second notification was sent 15 min later if the task was not initiated. Participants chose the time of their scheduled tasks when first opening Mezurio; these notification times could be changed at any time from within the home screen of the app. Following the first task notification, participants had 16 hours to complete each activity before the task *expired*, with the exception of the thrice-daily Tilt Task, which must be completed within 2 hours. Restrictions on task availability were included to both prevent cognitive assessments from accumulating, thereby creating unequal daily participant burden which may adversely impact acceptability, and limit the temporal proximity of task completions (at least eight hours between daily tasks and at least one hour between thrice-daily tasks). The average duration of each task was as follows: Gallery Game learning task (2 min, 16 seconds), Story Time narration (2 min, 17 seconds), and Tilt Task (5 min, 33 seconds).

Figure 2. A schematic of the assessment schedule presented to participants completing the extended version of Mezurio, including Gallery Game practice tasks (days 1-3). Schedule epochs represent the three distinct times of the day during which a participant may receive a task notification (eg, after breakfast, lunch, and dinner).



Semistructured Interviews

Participants were invited to provide feedback on their experiences using the Mezurio smartphone app, with the first 20 volunteers selected to complete a semistructured interview (duration: approximately 30 min). Participants who had withdrawn from the study were not invited to complete these semistructured interviews. At the time of interview, all participants had completed their baseline period of assessments, with 13 participants having completed or in the process of completing their 6-month follow-up. As a result of some participants completing the interview after the start of their longitudinal follow-up, only 4 participants had experience using a Gallery Game–only version of Mezurio; however, 3 participants were further unable to interact with Tilt Task as

their smartphone did not have adequate sensors (a study smartphone was loaned to participants with an incompatible personal device before the next wave of longitudinal data collection). The schedule of interview questions (Multimedia Appendix 1) included a mixture of closed-answer questions requiring a score out of 10 and open-ended questions, with prompts tailored to each participant's responses. The presentation of questions broadly mapped to the order in the schedule, with revisions according to individual responses. Interviews were audio recorded and transcribed verbatim.

Data Analysis

Compliance

The proportion of Gallery Game learning tasks attempted by each participant at baseline is considered as the primary outcome of compliance, as this aspect of the study schedule was uniformly presented to all participants. The presentation of Gallery Game memory tests (recognition and recall) was conditional on successful completion of the associated learning task; therefore, compliance with this aspect of Gallery Game was not considered as a separate, independent outcome. The association between participant characteristics and compliance with Gallery Game learning tasks (reverse-score square root transformation to account for negative skew) was screened, using between-group 2-tailed *t* tests for categorical (immediate family history and gender) and simple linear regression for continuous (age and years of education) variables (alpha=.05). Attrition to study participation was proxied as how far through the schedule of Gallery Game learning tasks were participants last active. To further examine whether participation declined as a function of time in the study, the relationship between the assessment index (1-22) and the proportion of participants attempting each scheduled learning task was subject to a nonparametric correlational analysis.

The proportion of Story Time narration (n=12) and Tilt Task (n=7) attempted by each participant is also reported; however, the small number of individuals exposed to these assessments at baseline precludes meaningful interpretation. The impact of participant characteristics and time in the study on compliance is not subject to statistical tests because of the limited sample size, but descriptive statistics are provided for the proportion of participants interacting with these tasks each day.

Interviews

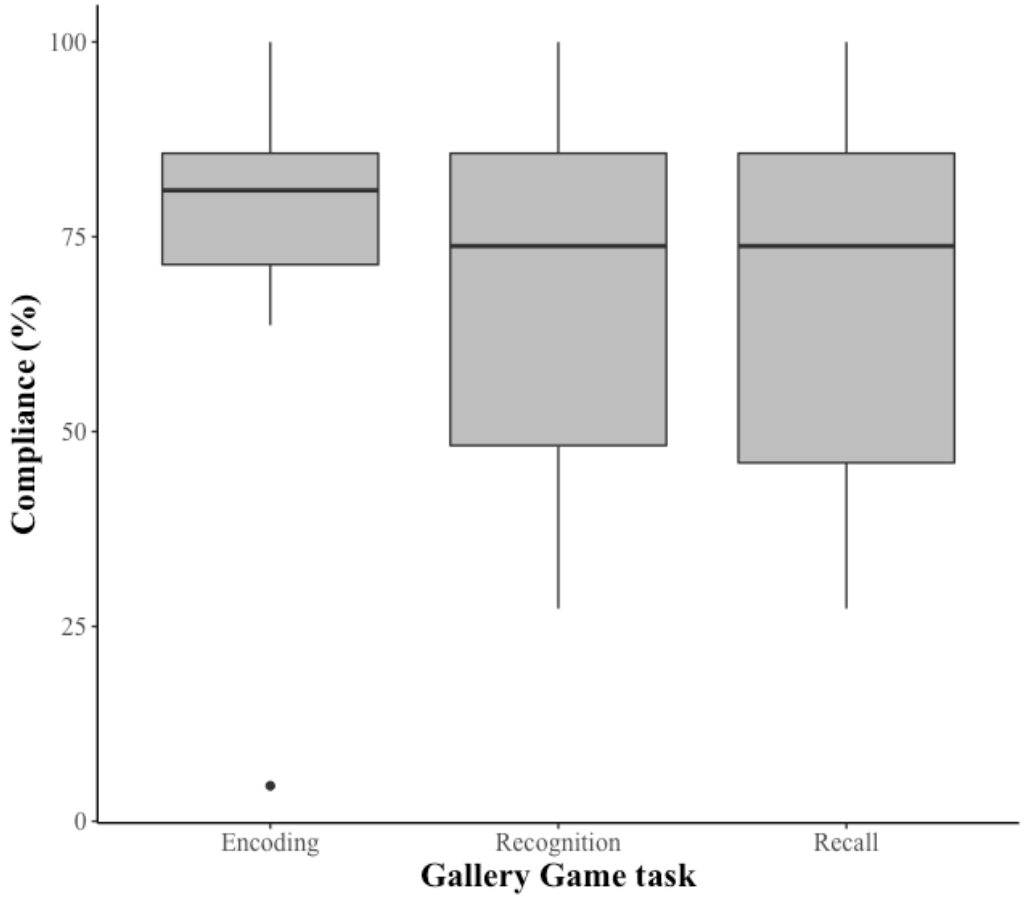
Interviews were subject to a thematic analysis [25], with the lead author (CL) taking a deductive approach to analyze the transcripts, focused on the following research themes: (1) approachability of the Mezurio app, including a more general consideration of smartphone technology, (2) acceptability of the research ask, and (3) engagement with the cognitive tasks. Additional themes emerging from the data were considered, but these are not considered here as factors contributing to the feasibility of a smartphone-based assessment. The lead author (CL) read each transcript iteratively to extract words and phrases representing the key themes of participant experience.

Results

Compliance

Average compliance to the schedule of Gallery Game learning tasks was 77.85% (SD 15.88). One outlier (4.55%) only completed the first learning task, later withdrawing from the study because of the time commitment. Excluding this participant, average compliance was 80.00% (SD 9.60). Volunteer characteristics are shown in Table 1. An immediate family history of dementia (mean 79.13, SD 9.33) was associated with poorer compliance to the study schedule in comparison with those with no family history (mean 81.79, SD 10.38;$tt_{43.33}$=17.25; *P*<.001), as was being female (mean 79.30, SD 9.38) as opposed to being male (mean 82.30, SD 10.62;$t_{41.67}$=12.41; *P*<.001). However, these associations must be considered with caution because of limited group sizes. There was no significant association between age (*P*=.10) and years of education (*P*=.74) with compliance to the schedule of Gallery Game learning tasks. Participants completed an average of 67.26% (SD 20.61) of recognition tests and 66.63% (SD 20.59) of recall tests (see Figure 3).

Figure 3. Proportion compliance for Gallery Game tasks (learning, recognition, and recall) across the baseline month of assessments.



Attrition across the baseline period of Gallery Game learning tasks was limited, with 88% of the participants completing the final scheduled learning task. The mean proportion of the schedule completed at the point of final learning was as follows: 98.15% (SD 5.70); range 75% to 100%. In addition, the percentage of participants compliant with each of the 22 scheduled learning tasks presented across the baseline assessment period is shown in Table 2. There was a nonsignificant, limited relationship between the learning task index and compliance (=0.09; *P*=.60).

Average compliance to the schedule of story narration tasks was 75.00% (SD 26.11), with the proportion of participants attempting each of the six scheduled daily tasks as follows: (1) 83%, (2) 92%, (3) 75%, (4) 75%, (5) 67%, and (6) 58%. Participants completed 38% (SD 31.83) of thrice-daily Tilt Tasks; however, the range of tasks (1-17) attempted by participants (out of a possible 21) may reflect the technical problems some participants reported for this task. The proportion of participants active each day of Tilt Task was as follows: (1) 100%, (2) 71%, (3) 57%, (4) 43%, (5) 29%, (6) 57%, and (7) 29%, but note the very small sample size (n=7).

Table 2. The proportion of participants (n=35) completing each of the 22 Gallery Game learning tasks distributed across the baseline Mezurio assessment period.

Task index	Compliance (%)
1	97
2	89
3	91
4	94
5	86
6	80
7	83
8	83
9	83
10	89
11	89
12	89
13	94
14	89
15	80
16	86
17	83
18	77
19	67
20	83
21	67
22	83

A Thematic Analysis of App Experience

Quotations evidencing the extracted themes are present in Multimedia Appendix 2.

Approachability of the Mezurio App

The Approachability of Smartphones

The approachability of smartphones was confirmed in this study's middle-aged sample, with all interviewed participants owning a personal device (12 Apple iPhone users and 8 Android users) and 90% of the participants reporting daily or more frequent smartphone usage (the remaining participants stated they used their phone on most days). Indeed, 35% (7/20) of participants reported active attempts to decrease smartphone usage in their daily lives, supporting the prevalence of these technologies in the target population, with one participant stating:

> *It's a double thing where by the convenience is amazing but it's also very intrusive, and I have to consciously switch it off or put it away or put the ringer off or something for it to not be constant.* [OX017]

Comparable smartphone ownership is inferred in the wider study population (n=35); the cited reasons for borrowing a study device (n=7) were not wishing to use their personal phone for the purpose of the study (n=2) and smartphone incompatibility because of either an outdated operating system (n=2) or inadequate sensors for the completion of Tilt Task (n=3), rather than participants not owning a device. Smartphones were used by participants for a wide range of functions, including communication, work, finance, travel, and shopping; however, 35% (7/20) of the participants reported viewing their phone as a functional tool rather than a device used for enjoyment. The majority (15/20, 75%) of this middle-aged population did not use phones for gaming, with the two reasons reported for this being time commitment (3/20, 15%) and lack of interest (3/20, 15%). A small number of participants had previously completed the AD-focused SeaHero Quest (2/20, 10%) [26] or used *brain-training* apps (2/20, 10%).

Remote Setup

Remote setup of the Mezurio app was well accepted, with the clarity of the installation process scoring an average of 9.03 out of 10 (SD 1.52); participant OX066 recalled the Mezurio setup process as:

> *I clicked the click and in....*

A total of 80% (16/20) of the participants reported that the written instructions for installation were sufficient; indeed, 1 participant felt the quantity of written guidance provided by the

research team was a barrier to completing what in reality was a very simple task, stating the following:

> *it was sufficiently information intensive to make me feel like I had to put it down and find time to do it whereas successful calls to action just make it super easy for you to go click, click, done... I think, I think more information is not necessarily better communication.* [OX037]

Task Onboarding

Task onboarding for Gallery Game was appropriate for this sample, with instructions scoring an average of 9.30 out of 10 (SD .10). Participants highlighted a number of features that were effective for remote introduction to Gallery Game, including simple, short written instructions, the inclusion of *pop-up* responsive help within the task, and the provision of a task strategy at the start of the assessment period. Participants were allowed up to 7 days to practice the Gallery Game task; the consensus (13/20, 65%) was that an initial practice period of 1 or 2 days is sufficient for understanding task demands. For participants completing Tilt Task, the instructions were again judged to be clear, with participants highlighting being provided with an opportunity to practice, coupled with a gradual increase in task complexity, as important for task onboarding; OX011 stated the following:

> *and the way it gradually increases with complexity. I mean, that's a hard thing to describe because it's one of those things you just really need to see and try it and then the instructions kind of make sense if you see what I mean.*

There was less consensus around the clarity of task onboarding for Story Time; participants were unclear of the amount of descriptive detail to provide during narration:

> *they say describe in as much detail as possible erm, but I'm not sure you really do want that because somebody who takes things so literally, like me, and then I spend forever doing the first scene.* [OX066]

The User Interface

The user interface of Gallery Game and Tilt Task was acceptable, with the manual response mechanism of the Tilt Task evaluated positively by 40% (8/20) of participants. Qualitative feedback suggested that the spoken interface of Story Time increased the burden of remote, digital assessment (7/20, 35%) by limiting situations in which the task could be completed, with participant OX015 explaining:

> *it was just another layer of something I had to do to complete it, you know, I had to take myself of to a room or a quiet space to do it...*

Across Mezurio, visual clarity of the app was highlighted (6/20, 30%) as an important feature, with OX026 stating:

> *I liked the layout because it's clear, it's not, sort of, forced into a scrumpled little heap that you have to read carefully. It's well laid out and the writing is a good size to read...*

Acceptability of the Research Ask

The approach in this study to remote cognitive assessment was scored 8.37 out of 10 (SD 1.34) for participant acceptance. The limited number of participants (4/20, 20%) completing the Gallery Game–only version of Mezurio precludes traditional significance testing; however, the mean acceptability of this group (mean 9.00) is comparable with ratings by those completing the extended version of the app (mean 8.20).

Daily Time Commitment

Daily time commitment was an important feature in the acceptability of the current participation ask, with short task durations being important in allowing the app to work around participants' schedules:

> *I think having it short is good because if you make it too long then you're going to find that people like me, if we're out at work then trying to fit in makes it more of a problem.* [OX062]

Increased daily participation load in the extended version of Mezurio, characterized by occasional, additive scheduled activities (eg, a Gallery Game and Story Time on the same day) and longer daily task durations contributed to an increased sense of burden (3/20, 15%). This study asked participants to interact with the app every day for up to 36 days; this study duration was not a significant problem for the current research group, with limited (3/20, 15%) discussion of this as an issue.

Scheduling the Tasks

Scheduling the tasks via local phone-based notifications benefitted study compliance (n=20), with many participants noting that the regularity of app timings helped them complete their daily tasks, eg, OX033 stated:

> *the fact that it was very regular, it was once a day, such a regular thing, it just became part of my daily, you know, habit.*

Specifically, the inclusion of short, cognitive assessments within the participants' daily routine was highlighted as important, with poor compliance generally accounted for by changes to an individual's regular patterns of activity, eg, weekends and trips (3/30, 15%). Participants using the extended version of Mezurio were prompted to complete Tilt Task thrice daily, with the alignment of task notifications to the daily routine (eg, breakfast, lunch, and dinner) reported to be increasingly important for high-frequency assessment. A total of 3 participants reported finding the 3-times-a-day routine to be too much to fit into their daily routine.

Flexibility of the Study Schedule

Flexibility of the study schedule was emphasized as a requisite for high-frequency, remote assessment. Within the app, participants had the option of changing the time of their next prompt; 7 participants identified this feature as a strength, with participant OX066 stating:

> *...really nice that you could change the times. I got quite good at thinking that's not going to work tomorrow and changing them.*

Participants (5/20, 25%) suggested that Mezurio could be improved in the future through the inclusion of a *snooze* function for notifications, allowing participants to set a second reminder for later in the day. Actively maintaining the intention to complete Mezurio after noncompliance with the initial prompt was associated with increased subjective burden (2/20, 10%). Presently, participants are able to re-enter the Mezurio app and complete their daily task within a certain expiry window (once-daily tasks: 16 hours; thrice-daily tasks: 2 hours). This feature benefitted research acceptability; however, there was still a wish for greater flexibility. A further suggestion to improve Mezurio is the ability for participants to monitor the task expiry time (3/20, 15%).

Perceived Burden

Perceived burden of the research ask was largely dependent on the external stressors to a participant's time, independent from the app, which is explained as follows:

> *I mean, conceptually it wasn't burdensome it was absolutely fine, you know, I didn't feel I was being put upon, erm, practically, however, I just found that the, my ability to, err, err, to, to, to adhere to the schedule... just got compromised by the ins and outs of daily life.* [OX037]

Furthermore, daily-life stressors were associated with subjective reports of poorer cognitive performance on that day, supporting the need to sample cognition at multiple time points to maximize reliability. Asking participants to interact with the app thrice daily was considered burdensome by some participants (3/20, 15%).

Engagement

Task Enjoyment

Task enjoyment was scored out of 10 as following: Gallery Game mean 7.26 (SD 2.02), n=19; Tilt Task mean 6.44 (SD 2.62), n=12; and Story Time mean 7.18 (SD 1.49), n=14. The cognitive demand of included assessments was cited as a primary source of participant enjoyment (9/20, 45%), with participants reflecting on the satisfaction of challenging themselves in comparison with their previous performance and the development of personal strategies to aid performance (3/20, 15%):

> *I was surprised how...a very simple task, task, was actually quite hard to do and so I could feel it challenging my brain, and that, that felt good.* [OX066]

In contrast, barriers to research enjoyment included concern over their own performance (6/20, 30%), frustration at difficult tasks (4/20, 20%), and limited variation in the day-to-day task demand (2/20, 10%):

> *by the last one I just felt "how many more!"...I guess there will be reasons why you've done several days in a row but if you could do sev, a few days and then a break, and then come back and do a few days.* [OX042]

Enjoyment was not considered an important factor for compliance by a small number of participants describing the tasks as functional as opposed to fun. In relation to this, a main motivator for engagement was commitment to the research aims rather than personal pleasure.

The Inclusion of Feedback

The inclusion of feedback was identified by participants as the foremost way to promote participant engagement with remote cognitive assessment. Although limited explicit feedback on performance outcomes is provided in the current version of Mezurio, participants had an intuitive sense of their own performance (7/20, 35%). Participants highlighted a need for explicit feedback on how their performance changed across the course of the research, with feedback on performance in comparison with peers identified as a potential factor to increase study participation. In addition, greater dissemination of the research background, objective, and methods within the app was highlighted as a potential to promote further engagement:

> *...it would be really helpful to get some commentary on what it's contributing to, even if that needs to be left until the end...but within the game. It's not to get an email, you know a week later with a 2-page pdf, that's mainly because I just won't read it....* [OX037]

Discussion

Principal Findings

This study tested the feasibility of digital cognitive assessment, deployed through the Mezurio smartphone app, in a middle-aged group relevant to the detection of early preclinical AD [22,23]. Participants were prompted to complete a prolonged schedule of daily assessments (36 days); a qualitative evaluation of Mezurio's user interface and task design, alongside study compliance and attrition, was used to explore whether smartphone-centered tools can substantially extend the breadth of cognitive assessment in which participants will engage. Excluding 1 participant who subsequently withdrew because of the time commitment, compliance with the schedule of Gallery Game learning tasks averaged 80%, with 88% of the participants still active at the end of the assessment period (36 days), confirming the feasibility of frequent, long-term cognitive assessment in a digital environment. Critically, participant feedback supported the acceptability of Mezurio's approach to digital assessment, with an intuitive user interface, flexible scheduling around personalized prompts, and engagement within the tasks themselves identified as important factors contributing to a positive research experience. Preliminary reports of compliance with the schedule of story narration and tilt executive function measures were at 75% and 38%, respectively, but the small number of individuals exposed to these activities at baseline and the technical issues identified in this first deployment of Tilt Task limit interpretation of these data.

Importantly, this study demonstrated participant engagement with daily cognitive assessments across a significantly longer study duration than the previously reported 7-day window [14-16]. This provides evidence that smartphone technologies enable far greater sampling of cognition than plausible with in-person tests. Consistent with the high levels of participation reported in older adults who are asked to complete frequent

mobile cognitive assessments [14-16], high compliance in this middle-aged group confirmed the feasibility of a *little but often* approach to smartphone-based cognitive testing. Furthermore, limited attrition supports the potential utility of smartphone-based tools for long-term monitoring. Participation *dropout* is reported in digital research spanning multiple months [17,27], with cited reasons for withdrawing from the research including technical problems, time commitment, and loss of interest in repetitive tasks [27]. Although study reuptake at months 6 and 12 for this initial pilot of Mezurio remains to be established, the progression of strategies to sustain participant engagement is important for the quality of collected data. In addition, the generalizability of research outcomes must be considered, with the limited published data in this field suggesting compliance to a schedule of remote cognitive assessment differs between study groups; 27% compliance has been reported in young adults [28] compared with 84% to 91% compliance in adults with reported substance abuse [29,30].

A major benefit of smartphone assessment is the ability for participants to independently contribute to research from a home environment; therefore, it is critical to establish the remote usability of such measures before implementing them at scale. Middle-aged participants in this study positively evaluated the clarity of setup and task instructions for Mezurio, with concise written instructions, an opportunity to practice, and responsive inbuilt help identified as design strengths—developed through iterative patient and public involvement. All participants were frequent smartphone users; however, although the adoption of smartphone technology is increasing in older age groups (49% of adults in the United Kingdom aged 55-64 years and 17% of adults aged over 65 years reported owning a device in 2015) [31], generalizability of app usability remains to be established. Subjective feedback from participants, including those with no previous experience with smartphones, on a smartphone app for monitoring the symptoms of chronic pain identified inexperience with technology and the need for technical support as potential issues [32]. Future work will further establish the feasibility of Mezurio in a wider general population.

Scheduling, both the length of individual assessments and participant autonomy in the timing of smartphone-based notifications, was critical for the acceptability of reseach ask [10,18,19]. Each task within Mezurio is designed to take 5 min or less to complete; although acceptability of the current participation ask was high (8.37/10), time commitment was reported as a primary reason for noncompliance, emphasizing the value of a *little but often* approach for repeat mobile assessment [14-16]. The allowance to personalize phone-based notifications to suit participants' daily routine and the flexibility given to delay responding to scheduled assessments were identified as important factors in limiting research burden by participants providing discursive feedback. Although the current version of Mezurio does not objectively record the scheduling behavior of participants, including notification times and the lag between participants receiving such a prompt and initating their next task, this may be explored in future research to improve the utility of Mezurio for time-sensitive research protocols, eg, monitoring treatment response.

The mental challenge of daily activities was identified as an important factor for participant engagement with repeat Mezurio assessments, with the suite of tasks within this app intended for use by adults with no clinical diagnosis of cognitive impairment in contrast to a more traditional approach to neuropsychological assessment. However, in a limited number of participants, subjective performance was associated with anxiety, which is relevant for the future modification of these tasks for use in individuals with mild cognitive impairment. Presently, Mezurio provides limited explicit feedback on task performance, with participants suggesting that future personalization of this aspect of the app would benefit sustained participant engagement, consistent with previous focus groups [18,19]. In accordance with Mezurio's intended screening utility for preclinical dementia, careful consideration of how to ethically communicate performance outcomes to participants is needed. Importantly, feedback in this study suggested that short communications within the app, intended to inform and promote interest in the research area, may similarly contribute to long-term engagement.

Limitations

A limitation of this study is that the Tilt Task and Story Time were included as an optional, later amendment, meaning that participants did not have a uniform experience of using the app. In addition, the reasons were not evaluated for participants declining to take part in this extended version of Mezurio at 6-month follow-up and for declining or withdrawing from (n=1) this ancillary technology substudy of PREVENT more broadly; this would have provided additional insight into the approachability of smartphone cognitive assessment. In addition, a number of participants experienced technical issues, which prevented or disrupted their completion of Tilt Task assessments, likely to impact reported Tilt Task compliance and subjective feedback. Building on this early, pilot study, a number of improvements have been made to the software underpinning this task.

When drawing on these results, it is worth noting that this study's sample was small and recruited from an existing, more intensive prospective cohort. Specifically, as these participants were *research aware* and had already demonstrated a high commitment to dementia research, mirrored in their qualitative feedback, the promise of using digital biomarkers to screen cognition remotely may not generalize to a wider population. The GameChanger study [33] has been launched to directly investigate whether the feasibility of Mezurio generalizes across a wide UK demographic, with over 16,000 participants completing remote high-frequency cognitive assessments within the app to date. It is also notable that the recruited sample comprised cognitively healthy individuals; thus, the acceptability and feasibility of using the app in impaired populations remains to be studied. Conclusions from this study, along with further input from patient and public involvement sessions, have been used to strengthen the research design implemented in the Mezurio smartphone app.

Conclusions

This research supports the feasibility of the Mezurio smartphone app for extensive cognitive profiling in middle-aged adults, evidencing high compliance to a substantially longer schedule

of daily interactions than previously explored. The scheduling of smartphone interactions, clarity of user experience, and task design were critical for reported engagement with Mezurio, with the qualitative feedback presented here providing an important direction for implementing digital tools in future health research. Following this initial demonstration of the viability of remote Mezurio assessment as a complementary method to in-clinic assessment, ongoing work seeks to replicate this feasibility in a wider population, as well as in adults with a diagnosis of dementia. In addition, establishing the scientific utility of these novel cognitive tasks in comparison with traditional markers of preclinical dementia (eg, brain-based biomarkers, neuropsychological outcomes, and prospective decline) is still in progress. However, the conclusions in this study are an important first step in justifying a participant-orientated mobile design for the progression of efficient, early screening of cognitive decline.

Acknowledgments

The development of the Mezurio app was supported by funding from the Robertson Foundation, National Institute for Health Research Oxford Health Biomedical Research Centre, and Eli-Lilly and F Hoffmann-La Roche. Additional support for this study was provided by the Oxford Clinical Academic Graduate School Clinical Lecturer Support Scheme, the Academy of Medical Sciences Clinical Lecturer Starter Grant (SGL016\1079), and the Medical Research Council Deep and Frequent Phenotyping study grant (MR/N029941/1). The PREVENT research program was developed with a grant from Alzheimer's Society.

Conflicts of Interest

IK declares paid membership of the Advisory Board of Mantrah Ltd, a company developing digital solutions to support the care of individuals with dementia.

Multimedia Appendix 1
The schedule for the semistructured interviews.
[DOCX File , 16 KB - mhealth_v8i4e16142_app1.docx]

Multimedia Appendix 2
Thematic outcomes of self-reported research experience with the Mezurio app, evidenced by participant quotations.
[DOCX File , 21 KB - mhealth_v8i4e16142_app2.docx]

References

1. Gold M, Amatniek J, Carrillo MC, Cedarbaum JM, Hendrix JA, Miller BB, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2018;4:234-242 [FREE Full text] [doi: 10.1016/j.trci.2018.04.003] [Medline: 29955666]
2. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2:pii: 9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
3. Hird N, Ghosh S, Kitano H. Digital health revolution: perfect storm or perfect opportunity for pharmaceutical R&D? Drug Discov Today 2016 Jun;21(6):900-911. [doi: 10.1016/j.drudis.2016.01.010] [Medline: 26821131]
4. Areàn PA, Ly KH, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [Medline: 27489456]
5. Resnick HE, Lathan CE. From battlefield to home: a mobile platform for assessing brain health. Mhealth 2016;2:30 [FREE Full text] [doi: 10.21037/mhealth.2016.07.02] [Medline: 28293603]
6. Brose A, Schmiedek F, Lövdén M, Lindenberger U. Daily variability in working memory is coupled with negative affect: the role of attention and motivation. Emotion 2012 Jul;12(3):605-617. [doi: 10.1037/a0024436] [Medline: 21787075]
7. Hess TM, Popham LE, Emery L, Elliott T. Mood, motivation, and misinformation: aging and affective state influences on memory. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2012;19(1-2):13-34 [FREE Full text] [doi: 10.1080/13825585.2011.622740] [Medline: 22059441]
8. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. Cogn Neuropsychol 2007 Oct;24(7):755-789. [doi: 10.1080/02643290701754158] [Medline: 18066734]
9. Clay I. Impact of digital technologies on novel endpoint capture in clinical trials. Clin Pharmacol Ther 2017 Dec;102(6):912-913. [doi: 10.1002/cpt.866] [Medline: 29027665]
10. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038 [FREE Full text] [doi: 10.1093/geroni/igy038] [Medline: 30619948]
11. Klimova B. Mobile phone apps in the management and assessment of mild cognitive impairment and/or mild-to-moderate dementia: an opinion article on recent findings. Front Hum Neurosci 2017;11:461 [FREE Full text] [doi: 10.3389/fnhum.2017.00461] [Medline: 28970789]

12. Onoda K, Yamaguchi S. Revision of the cognitive assessment for dementia, iPad version (CADi2). PLoS One 2014;9(10):e109931 [FREE Full text] [doi: 10.1371/journal.pone.0109931] [Medline: 25310860]
13. Ruggeri K, Maguire Á, Andrews JL, Martin E, Menon S. Are we there yet? Exploring the impact of translating cognitive tests for dementia using mobile technology in an aging population. Front Aging Neurosci 2016;8:21 [FREE Full text] [doi: 10.3389/fnagi.2016.00021] [Medline: 27014053]
14. Allard M, Husky M, Catheline G, Pelletier A, Dilharreguy B, Amieva H, et al. Mobile technologies in the early detection of cognitive decline. PLoS One 2014;9(12):e112197 [FREE Full text] [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
15. Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep;26(3). [doi: 10.1002/mpr.1521] [Medline: 27539327]
16. Lange S, Süß HM. Measuring slips and lapses when they occur - ambulatory assessment in application to cognitive failures. Conscious Cogn 2014 Mar;24:1-11. [doi: 10.1016/j.concog.2013.12.008] [Medline: 24384496]
17. Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E. Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality Proof-of-Principle study. JMIR Mhealth Uhealth 2017 May 25;5(5):e68 [FREE Full text] [doi: 10.2196/mhealth.6939] [Medline: 28546139]
18. Jenkins A, Lindsay S, Eslambolchilar P, Thornton IM, Tales A. Administering cognitive tests through touch screen tablet devices: potential issues. J Alzheimers Dis 2016 Oct 4;54(3):1169-1182. [doi: 10.3233/JAD-160545] [Medline: 27567875]
19. van Middelaar T, Beishuizen CR, Guillemont J, Barbera M, Richard E, Moll van Charante EP, HATICE consortium. Engaging older people in an internet platform for cardiovascular risk self-management: a qualitative study among Dutch HATICE participants. BMJ Open 2018 Jan 21;8(1):e019683 [FREE Full text] [doi: 10.1136/bmjopen-2017-019683] [Medline: 29358447]
20. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
21. Finch CE. The neurobiology of middle-age has arrived. Neurobiol Aging 2009 May;30(4):515-20; discussion 530. [doi: 10.1016/j.neurobiolaging.2008.11.011] [Medline: 19231030]
22. Irwin K, Sexton C, Daniel T, Lawlor B, Naci L. Healthy aging and dementia: two roads diverging in midlife? Front Aging Neurosci 2018;10:275 [FREE Full text] [doi: 10.3389/fnagi.2018.00275] [Medline: 30283329]
23. Ritchie CW, Ritchie K. The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onset Alzheimer's disease. BMJ Open 2012;2(6):pii: e001893 [FREE Full text] [doi: 10.1136/bmjopen-2012-001893] [Medline: 23166135]
24. Lancaster C, Koychev I, Blane J, Chinner A, Chatham C, Taylor K, et al. Gallery Game: Smartphone-based assessment of long-term memory in adults at risk of Alzheimer's disease. J Clin Exp Neuropsychol 2020 Jan 24:1-15. [doi: 10.1080/13803395.2020.1714551] [Medline: 31973659]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Coutrot A, Schmidt S, Coutrot L, Pittman J, Hong L, Wiener JM, et al. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. PLoS One 2019;14(3):e0213272 [FREE Full text] [doi: 10.1371/journal.pone.0213272] [Medline: 30883560]
27. Valdes EG, Sadeq NA, Bush AL, Morgan D, Andel R. Regular cognitive self-monitoring in community-dwelling older adults using an internet-based tool. J Clin Exp Neuropsychol 2016 Nov;38(9):1026-1037. [doi: 10.1080/13803395.2016.1186155] [Medline: 27266359]
28. Price E, Moore G, Galway L, Linden M. Validation of a smartphone-based approach to in situ cognitive fatigue assessment. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e125 [FREE Full text] [doi: 10.2196/mhealth.6333] [Medline: 28818818]
29. Pal R, Mendelson J, Clavier O, Baggott MJ, Coyle J, Galloway GP. Development and testing of a smartphone-based cognitive/neuropsychological evaluation system for substance abusers. J Psychoactive Drugs 2016;48(4):288-294. [doi: 10.1080/02791072.2016.1191093] [Medline: 27260123]
30. Jones A, Tiplady B, Houben K, Nederkoorn C, Field M. Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. Psychopharmacology (Berl) 2018 May;235(5):1487-1496 [FREE Full text] [doi: 10.1007/s00213-018-4860-5] [Medline: 29497782]
31. Ofcom. 2015. The Communications Market Report 2015 URL: https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr15/uk [accessed 2020-01-31]
32. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. JMIR Mhealth Uhealth 2017 Mar 24;5(3):e37 [FREE Full text] [doi: 10.2196/mhealth.6496] [Medline: 28341616]
33. GameChanger. URL: https://joingamechanger.org [accessed 2020-01-30]

Abbreviations

AD: Alzheimer disease

Edited by G Eysenbach; submitted 05.09.19; peer-reviewed by F Lanfranchi, D Leightley, M Whelan; comments to author 30.10.19; revised version received 18.12.19; accepted 19.12.19; published 02.04.20.

Please cite as:
Lancaster C, Koychev I, Blane J, Chinner A, Wolters L, Hinds C
Evaluating the Feasibility of Frequent Cognitive Assessment Using the Mezurio Smartphone App: Observational and Interview Study in Adults With Elevated Dementia Risk
JMIR Mhealth Uhealth 2020;8(4):e16142
URL: https://mhealth.jmir.org/2020/4/e16142
doi:10.2196/16142
PMID:32238339

Original Paper

Integrated Care Intervention Supported by a Mobile Health Tool for Patients Using Noninvasive Ventilation at Home: Randomized Controlled Trial

Erik Baltaxe[1,2], MD; Cristina Embid[1,2], MD, PhD; Eva Aumatell[1], MSc; María Martínez[1], MSc; Anael Barberan-Garcia[1,2], PhD; John Kelly[3], MSc; John Eaglesham[3], MSc; Carmen Herranz[1,2], MSc; Eloisa Vargiu[4], PhD; Josep Maria Montserrat[1,2], MD, PhD; Josep Roca[1,2], MD, PhD; Isaac Cano[1,2], PhD

[1]Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Universitat de Barcelona, Barcelona, Spain

[2]Center for Biomedical Research Network in Respiratory Diseases, Madrid, Spain

[3]Advanced Digital Innovation (UK) Ltd, Salts Mill, United Kingdom

[4]Eurecat Technological Center of Catalonia, Barcelona, Spain

Corresponding Author:
Isaac Cano, PhD
Hospital Clínic de Barcelona
Institut d'Investigacions Biomèdiques August Pi i Sunyer
Universitat de Barcelona
Villarroel 170
Barcelona, 08036
Spain
Phone: 34 932275540
Email: ISCANO@clinic.cat

Abstract

Background: Home-based noninvasive ventilation has proven cost-effective. But, adherence to therapy still constitutes a common clinical problem. We hypothesized that a behavioral intervention supported by a mobile health (mHealth) app could enhance patient self-efficacy. It is widely accepted that mHealth-supported services can enhance productive interactions among the stakeholders involved in home-based respiratory therapies.

Objective: This study aimed to measure changes in self-efficacy in patients with chronic respiratory failure due to diverse etiologies during a 3-month follow-up period after the intervention. Ancillary objectives were assessment of usability and acceptability of the mobile app as well as its potential contribution to collaborative work among stakeholders.

Methods: A single-blind, single-center, randomized controlled trial was conducted between February 2019 and June 2019 with 67 adult patients with chronic respiratory failure undergoing home-based noninvasive ventilation. In the intervention group, a psychologist delivered a face-to-face motivational intervention. Follow-up was supported by a mobile app that allowed patients to report the number of hours of daily noninvasive ventilation use and problems with the therapy. Advice was automatically delivered by the mobile app in case of a reported problem. The control group received usual care. The primary outcome was the change in the Self Efficacy in Sleep Apnea questionnaire score. Secondary outcomes included app usability, app acceptability, continuity of care, person-centered care, and ventilatory parameters.

Results: Self-efficacy was not significantly different in the intervention group after the intervention (before: mean 3.4, SD 0.6; after: mean 3.4, SD 0.5, P=.51). No changes were observed in adherence to therapy nor quality of life. Overall, the mHealth tool had a good usability score (mean 78 points) and high acceptance rate (mean score of 7.5/10 on a Likert scale). It was considered user-friendly (mean score of 8.2/10 on a Likert scale) and easy to use without assistance (mean score of 8.5/10 on a Likert scale). Patients also scored the perception of continuity of care and person-centered care as high.

Conclusions: The integrated care intervention supported by the mobile app did not improve patient self-management. However, the high acceptance of the mobile app might indicate potential for enhanced communication among stakeholders. The study identified key elements required for mHealth tools to provide effective support to collaborative work and personalized care.

Trial Registration: ClinicalTrials.gov NCT03932175; https://clinicaltrials.gov/ct2/show/NCT03932175

(JMIR Mhealth Uhealth 2020;8(4):e16395) doi:10.2196/16395

KEYWORDS

behavioral change; eHealth; noninvasive ventilation; mobile health; chronic diseases

Introduction

In the 1950s, the polio epidemic demonstrated the safety and efficacy of noninvasive ventilation (NIV) to decrease mortality [1]. Since then, the use of this therapeutic approach at home has reduced hospital admissions, has favorably impacted health-related quality of life, improved sleep quality, and reduced mortality in patients with chronic respiratory failure due to diverse etiologies [2-8]. These results have driven a steady increase in the prevalence of patients using home-based NIV in Europe, ranging from 4.5 to 20 per 100,000 adults [9-11].

Despite its proven cost-effectiveness [12], patient adherence to home-based NIV could still improve, which should further enhance health care–related efficiencies of the intervention [13]. Monitoring and optimization of physiological settings have enhanced adherence by improving the timely detection of problems such as mask leaks and patient-ventilator asynchronies [14]. Nevertheless, improvement in the behavioral aspects such as patient motivation and empowerment for self-management are important factors to consider when addressing adherence to respiratory therapies.

The current study sought to explore the transfer of previous positive experiences with behavioral interventions in other fields (ie, physical activity) [15-20] to home-based NIV. Specifically, we addressed the concept of self-efficacy, defined as the individual's perceived capability to perform a particular behavior [21]. Self-efficacy expectations can be affected by enablers or barriers such as the perception of physical function or the capacity for self-management. Therefore, a person who does not believe in her or his capacity to perform the desired action will fail to adopt, initiate, and maintain it. Self-efficacy is therefore seen as the most influential motivational factor and the strongest predictor of behavioral intentions [21].

We propose the use of a behavioral mobile health (mHealth) intervention, which can be framed by Bandura's model [22], to support changes in self-efficacy. This model is based on the concepts of health risk perceptions, health outcome expectancies, and the patients' confidence to engage in certain behaviors. The model has been widely applied in studies of the adoption, initiation, and maintenance of health-promoting behaviors [23].

In addition to self-efficacy as a way to influence behavioral change, previous reports by Hernandez et al [24] and Cano et al [25] identified two common hinderances for the effective implementation of complex respiratory therapies (ie, long-term oxygen therapy, continuous positive airway pressure therapy, home NIV, and home-based nebulizer therapy). First, interaction and communication, which could greatly benefit from digital tools supporting collaborative work, are needed among several stakeholders, namely health professionals at different health care tiers (eg, primary care, specialized care), patients and carers, companies undertaking equipment maintenance, and others. Second, improvement in therapeutic adherence is needed, which could be achieved by empowering patients to perform self-management.

Within this context, information and communication technologies (ICT) have been identified as promising tools to enhance the coordination between stakeholders and contribute to improved health outcomes [26,27]. Nonetheless, the implementation remains immature [28] due to a lack of evidence in a real-world context for the capacity of ICT to sustain behavioral changes, including self-efficacy, in patients with chronic, complex conditions. It is widely accepted that, despite current limitations, patients with chronic, complex conditions are an ideal population for which care coordination, patient and medical staff satisfaction, and patient empowerment are of the utmost importance to produce health benefits.

The principal objective of this study was to explore the capacity of a behavioral mHealth intervention to increase patient empowerment for self-management and adherence to therapy. The secondary aim was to learn, based on the experience of professionals and patients, how the mHealth tool should evolve to support collaborative work.

Methods

Study Design and Participants

A single-blind, single-center, randomized controlled trial with two parallel arms (1:1 ratio) was conducted. Patients were randomized to a control group or an intervention arm, which consisted of the behavioral mHealth intervention in addition to usual care. Inclusion criteria were as follows: all adult patients with hypercapnic ventilatory failure due to chest wall, neuromuscular, lung parenchyma, or airway disease already receiving treatment with NIV irrespective of treatment duration and in possession of a mobile phone or tablet that could support the use of the mHealth app (MyPathway). MyPathway [29] is a secure, digital communication channel connecting patients to clinicians and services. It is an app-based tool for both patients and clinicians to use on phones or tablets. See Multimedia Appendix 1 for more details. Patients with severe psychiatric or neurological diseases were excluded, as well as patients hospitalized at the time of assessment.

Intervention

In addition to usual care, the behavioral mHealth intervention included a face-to-face motivational interview by a psychologist (EA) to assess the patient's adherence profile and lifestyle habits, with a follow-up through the MyPathway app. In contrast, the control group received only usual care, which consisted of manual discharge and review of the NIV machine data by the treating pulmonologist and respiratory nurse. Respiratory parameters were changed, if needed, according to clinical data (anamnesis and physical examination) in addition to NIV data.

At the time of enrollment, semi-structured motivational interviews were conducted individually. Participants were asked about their treatment adaptation experience, lifestyle (physical activity and food habits), and use of ICT. In each session, field notes were taken anonymously, and no recordings were made. The intervention consisted of a 10-50–minute face-to-face session at the hospital or participants' home that followed the principles of a collaborative and evocative motivational interview, favoring the participant's autonomy. The techniques used were open questions, active listening, empathy, returning reflected thoughts, exploring a change in goals, summarizing, and giving feedback. Also, during the enrollment visit, patients were given verbal and written explanation on how to use the app. Free access was granted after receiving an invitation via the hospital health information system (SAP), which prompted the participant to register using an email address as the username. The app could also be downloaded to the carers' phone in case the patient did not have a smartphone.

During the follow-up, the MyPathway app was used by study participants for bidirectional interaction with the research team. It consisted of positive feedback or reinforcement messages in response to the number of hours of NIV use reported by the patient daily. Also, general advice on specific NIV clinical problems was automatically provided by the app according to the patients' weekly input. Additional educational material on physical activity, diet, and sleep hygiene could be accessed at any time via a dedicated link. A web-based clinical portal enabled the research team to monitor the patient-reported NIV hours of use and clinical problems. As indicated, a dedicated nurse with clinical and technical knowledge (one of the authors, MM) took the role of case manager to support collaborative work. She used the web-based portal to identify adherence problems and contacted participants via telephone or at home (for those with severe mobility problems) to enquire about and solve potential clinical or technical problems.

Procedures and Study Outcomes

The primary outcome was a change in self-efficacy, as measured using the Self Efficacy in Sleep Apnea (SEMSA) questionnaire. The SEMSA is a US-designed self-report questionnaire comprised of 26 items that are rated from 1 to 4 on a 4-point Likert scale [30]. The arithmetic mean of the Likert rating for each participant is computed for the overall SEMSA score and each of the 3 factors. The total score ranges from 1 to 4. Higher scores indicate greater risk perception, higher benefit expectancy with treatment, and greater perceived self-efficacy [30].

Secondary outcomes included usability of the ICT tool, as measured using the System Usability Scale [31]; patient satisfaction, as measured using the Net Promoter Score [32] in addition to 3 custom general satisfaction questions measured on a Likert scale; continuity of care, as measured using the Nijmegen Continuity Questionnaire [33]; and the Person-Centred Coordinated Care Experience Questionnaire as described by Leijten et al [34]. Moreover, ventilator-specific data such as the mean hours of daily use, unintentional leaks (L/s), minute ventilation (L/min), tidal volume (mL), and backup rate (breaths/min) were downloaded directly from the NIV machine.

Tertiary outcomes included mortality; health-related quality of life, as measured using the EuroQol 5D questionnaire [35,36]; and sleepiness, as measured using the Epworth Sleepiness Score.

The impact of the motivational mHealth tool recommendations on diet and exercise was indirectly measured by body weight changes.

All assessments were completed at baseline and the final visit scheduled 3 months later. The follow-up was conducted in the outpatient clinic for the control group and remotely by the nurse case manager (MM) using the MyPathway app and its clinical portal for the intervention group. When deemed necessary, the nurse case manager visited the patient at home, or a visit was scheduled in the outpatient clinics. There was no active follow-up for the control group.

Randomization and Masking

All eligible patients were contacted by telephone to briefly explain the study and invite them to participate. Those showing interest were invited to the hospital outpatient clinics. Study investigators (EB, EA, and MM) explained the study face-to-face, and, in case of acceptance, signed consent was obtained. Afterward, the patient was randomized. Before patient enrollment, the randomization scheme was generated using the website randomization.com by one of the researchers (EB). Blocks of 4 were used. Only after the participant provided consent, the investigator opened the envelope with the allocated study group.

Due to the nature of the intervention, neither the participants nor the investigators in direct contact with the participants were blinded. Only the investigator in charge of data analysis was blinded.

Sample Size Calculation, Data Management, and Statistical Analysis

Accepting an α risk of 0.05 and a β risk of 0.2 in a two-sided test, 31 subjects in the intervention group and 31 subjects in the control group were required to achieve a statistically significant difference ≥0.35 units in the SEMSA overall score [37]. The common SD was assumed to be 0.46 [38]. A 10% drop-rate was anticipated.

Baseline and end-of-study data (questionnaires) were collected face-to-face at the outpatient clinic by the investigators (EB, EA, and MM). Study data were collected and managed using the REDCap electronic case report form [39,40] hosted at the Hospital Clínic de Barcelona. Data on patient-reported NIV use and clinical problems with NIV were collected online using MyPathway.

Results are presented as mean (SD) or n (%). Comparisons were conducted using Chi-square or Fisher exact tests for categorical variables and Student *t* or Wilcoxon tests, depending on the distribution of the variables, for numerical variables.

Ethics

Study approval was obtained from the Ethics Committee for Clinical Research of Hospital Clínic de Barcelona (HCB/2019/0510). Patients read, understood, and accepted

informed consent, which was signed before enrolment to the study.

Results

Study Population

Between February and March 2019, all patients already being treated with NIV at the noninvasive ventilation clinic at the Hospital Clínic de Barcelona were assessed for eligibility. From an initial sample of 169 eligible patients, 50 (30%) did not meet the inclusion criteria, including 32 who did not have a smartphone or tablet, and 23 (14%) declined participation. Therefore, 67 patients were randomized between February and May 2019 (see the CONSORT flow diagram in Multimedia Appendix 2). One patient from the intervention group withdrew consent during the trial due to the worsening of his clinical condition. Baseline demographic and clinical characteristics are shown in Table 1 and Multimedia Appendix 3.

Patient-Reported Outcomes

For the primary outcome, the mean SEMSA score for self-efficacy was not significantly different in the intervention group after the intervention (before: 3.4, SD 0.6; after: 3.4, SD 0.5, *P*=.51).

The perceived risks, outcome expectancies, Epworth Sleepiness Score, and EuroQol 5Q-5D questionnaire score were also not significantly different in the intervention group after the intervention (see Multimedia Appendix 3). As for the patient experience questionnaires, neither the Nijmegen Continuity Questionnaire nor the Person-Centred Coordinated Care Experience Questionnaire were statistically significantly different between the groups (see Multimedia Appendix 3).

Table 1. Baseline characteristics of the study groups

	Intervention (n=33)	Control (n=34)	*P* value
Age (years), mean (SD)	68 (15.8)	65 (14.7)	.31
Male gender, n (%)	19 (58)	19 (58)	>.99
Weight, mean (SD)	86 (31.6)	78 (22.4)	.15
Educational level (n, %)			.73
No schooling	3 (9)	1 (3)	
School education	12 (36)	13 (38)	
Professional formation	17 (52)	19 (56)	
Doctorate or equivalent	1 (3)	1 (3)	
BMI (kg/m^2), mean (SD)	30.5 (7.1)	28.9 (7.4)	.35
Smoking status, n (%)			<.001
Never	12 (36)	16 (49)	
Former	18 (55)	16 (48)	
Current	2 (6)	1 (3)	
Smoking (packs/year), mean (SD)	55.5 (35.7)	52.5 (33)	.003
Diagnostic group, n (%)			
Neuromuscular	4 (12)	8 (24)	.25
Chest wall	11 (33)	10 (30)	.81
Obesity-hypoventilation	5 (15)	5 15)	>.99
Airway obstructive disease	3 (9)	2 (6)	.66
OSA[a] to CSA[b]	10 (30)	8 (24)	.60
Number of comorbidities per patient, mean (SD)	2 (1.5)	1.8 (1.6)	.68
Comorbidities, %			
Cancer	3	3	>.99
Congestive heart disease	33	27	.60
Ischemic heart disease	24	15	.37
Diabetes	27	36	.47
Stroke	9	9	>.99
Hypertension	67	52	.20
Dementia	3	0	.32
Neurological disorders other than stroke	3	0	.32
Depression/anxiety	18	18	>.99
Dyslipidemia	15	27	.54
Time on noninvasive ventilation (years), mean (SD)	6.75 (6.5)	4.5 (3.5)	.08
AHI[c], mean (SD)	46 (28.8)	35 (31.6)	.37
CT90[d] (%), mean (SD)	47 (37.3)	44 (40.4)	.91
Mean ventilatory parameters, mean (SD)			
IPAP[e] (cm H_2O)	16 (4.7)	14 (4.7)	.06
EPAP[f] (cm H_2O)	7 (2.8)	6 (2.1)	.31
Leak (L/s)	0.05 (0.2)	0.5 (0.09)	.03
Number of hours used per day	7.4 (2)	6.8 (3)	.28

[a]OSA: obstructive sleep apnea.

[b]CSA: central sleep apnea.

[c]AHI: global apnea-hypopnea index for all diagnostic groups.

[d]CT90: cumulative sleep time percentage with oxyhemoglobin saturation <90%.

[e]IPAP: inspiratory positive airway pressure.

[f]EPAP: expiratory positive airway pressure.

Clinical Outcomes

Adherence was measured as the number of hours the NIV was used per day, as recorded by the ventilator. The mean adherence value was not significantly different in the intervention group after the intervention (before: 7.4 hours, SD 2 hours; after: 7.7 hours, SD 2 hours). Mean minute ventilation was the only significantly different ventilatory parameter after the 3-month intervention in the intervention group (before: 7.0 L/min, SD 2 L/min; after: 6.4 L/min, SD 2.1 L/min, *P*=.03). The remaining ventilatory parameters and weight are shown in Multimedia Appendix 3. None of the patients died during the trial.

mHealth Tool Use, Usability, and Acceptability

The Net Promoter Score was –3 (10/33, 31% promoters; 11/33, 34% passives; 11/33, 34% detractors). The 3 Likert-scale questions about the general satisfaction with the app that were rated from 1 (very bad) to 10 (very good) resulted in a mean score of 7.5/10 for the general impression of the app, mean score of 8.2/10 for the user friendliness, and mean score of 8.5/10 for usability of the app without assistance. The mean System Usability Scale score was 78, a reasonably good grading. Up to 42% of the participants used the link to the educational material, and only 18% (6/33) consulted the terms of use. The mean number of hours of NIV use per day, reported using the mHealth tool, was 7.23 hours (SD 2.48 hours). Use of NIV for more than 4 hours per day during two-thirds of the study period was reported by 45% (15/33) of the patients. Likewise, the reported mean number of days during which NIV was used more than 4 hours in the entire intervention group was 35.6 days (SD 23.6 days). At the end of the study period, 3 participants stopped reporting due to app problems, 1 participant stopped using the app due to health problems, another participant stopped using the app for unknown reasons, and 3 participants decided to use the app on an alternative day basis.

Also, 30% (10/33) of the participants used the app through a family member or carer. It is of note that the nurse case manager was able to solve two-thirds of the technical problems that arose during the first 3 weeks of the study.

The qualitative analysis of the motivational interview as well as the detailed description of the requirements for mHealth to support collaborative work among stakeholders will be reported elsewhere. However, Table 2 summarizes a list of features that the research team agreed were key functional requirements of mHealth tools to effectively support collaborative work among stakeholders involved in home-based respiratory therapies.

Table 2. Requirements to support collaborative work within the noninvasive ventilation service.

Feature	Description of the requirement(s)
Adaptive case management	Capacity to enable the case manager to combine predesigned tasks and approach new cases by reusing structured experiences with previous cases. Over time, the case manager, or other authorized health professionals, should be able to adapt the work plan in a timely fashion to specific patient's requirements without any direct technological support
Team collaboration	Cloud-based, General Data Protection Regulation-compliant, enterprise-proven team collaboration tools to allow patients and health care professionals to break down silos and collaborate seamlessly from any device (mobile phone, tablet, or desktop) towards the health continuum care pathway
Multimedia communication	Enterprise-grade, scalable, high-quality, real-time communication among concurrent participants for file sharing, voice, video, and screen-share sessions with industry-standard encryption
Intelligent bots	Capacity to develop and integrate intelligent bots to guide professionals through continuum care pathways and to improve health risk assessment and service selection
Integration with hospital information systems	Use of HL7 Fast Healthcare Interoperability Resource interoperable middleware to integrate with provider-specific hospital information systems

Discussion

Principal Findings for Patient-Reported Outcomes

We report the results of a behavioral mHealth intervention based on a face-to-face interview and the use of an mHealth tool (MyPathway app) during a 3-month follow-up period with patients with hypercapnic chronic respiratory failure under home-based long-term NIV. To the best of our knowledge, this is the first randomized controlled trial using digital tools to support behavioral changes in this population [41-44].

In this study, the mean self-efficacy score was already high at baseline (Table 1), and we did not find a significant effect of the intervention on behavioral changes. Several explanations can be proposed for these results. First, the intervention may need to be more intensive (ie, more than one face-to-face session) [45]. Second, all the participating patients were long-term users without significant sleep symptoms at the time

of enrollment (average use >6 years with an average Epworth Sleepiness Score <10). Therefore, we could hypothesize that behavioral changes had occurred previously, as evidenced by the good average use of NIV (7.4 h/day) and high scores for self-efficacy at baseline. The inclusion of patients who have been newly prescribed NIV in future studies may show a positive impact of the intervention. Third, we may argue that, although NIV use was good among this sample of long-term users, adherence was more a function of necessity or imposition (by family or physicians) than a real feeling of self-management and that most of these chronic patients had not considered initiating behavioral changes [46,47]. Along this line of thought, the population we studied had mobility problems or poor general health, creating barriers for behavioral change [20]. Therefore, any intervention at this stage is likely to be ineffective. This may also be reflected by the lack of interest in consulting the educational material in the app (<50% of the patients did so). Last, we should note that the control group consisted of more patients with neuromuscular pathophysiology. However, the pathophysiology should not affect or have a direct relationship with the measured behavioral outcomes or the capacity and readiness to use the app. Accordingly, educational level is a more important factor [48,49], and both study groups had similar educational levels.

Usability, Acceptability, and Requirements for Supporting Collaborative Work

Notwithstanding the clinical results, it is important to note that the mHealth tool was well received by the patients and their family/caregivers. Despite their complex conditions (2 comorbidities on average) with considerable needs and burdensome treatment, all patients used the app regularly, grading it as generally good, user-friendly, and easy to use without help. Moreover, the System Usability Scale score was good.

As stated in the methods section, we want to highlight the fact that one of the authors (MM) undertook a new professional role during the study period. She became the clinical case manager with additional technical knowledge on the mHealth tool. Patients appreciated this new role very much despite the use of telephone or Whatsapp for bilateral communication. We found that the app lacked this function, and based on our experience, this should become an integral part of any app that includes case management with technical skills. This type of communication functionality should be cloud-based and General Data Protection Regulation-compliant. Moreover, future developments should consider adaptive case management functionality. Also, this communication should be supported by artificial intelligence to help guide professionals though continuum care pathways and improve health risk assessment and service selection. Finally, integration with hospital information systems may facilitate the whole process. This is in line with a recent report on the digital transformation of health care in Europe, which draws upon the experiences of 17 integrated care programs where the importance of communication technologies, new professional roles, and the relevance of clinical workflow evaluation were highlighted [50].

In this respect, we measured 2 process outcomes [51] related to patient experience [52]: continuity of care and person-centered care. Our study population, which included patients as well as their family and carers for one-third of the cases in the intervention group, evaluated both parameters very well. The importance of well-designed clinical workflows with embedded digital health tools may have an impact on not only an NIV service but also other respiratory services. Commonalities include high-complexity patients with clinical and social needs from different stakeholders (eg, physicians, providers, technicians, social workers) and health care tiers (eg, primary care, specialized care). Hernandez et al [24] showed how this complexity can hamper the effectiveness of long-term oxygen therapy. As mentioned, Table 2 shows the proposed elements to overcome the barriers for the successful implementation of digital health tools within clinical workflows relating to respiratory therapies.

Finally, stakeholders play an important role in the design and evaluation of digital health tools [53,54] and, as such, their input should be taken into account when evaluating a service in which there is considerable interplay between patients, different health care tiers, and social and technical services [55]. For an mHealth tool to produce health care value, it should be embedded in the clinical pathways of a well-evaluated clinical service and not as a standalone tool [56].

Strengths and Limitations of the Study

Our study considered the whole population of patients attending the clinic, resulting in a realistic clinical scenario. Another important strength of our study is its potential to demonstrate the positive interaction and collaborative work among the nurse case manager, patients, and family members or caregivers of complex patients using digital health tools. Previous studies [57-59] reported the use of digital tools by family caregivers, emphasizing the importance of including this group of stakeholders, not only as users but also in the co-design process. This stakeholder involvement is also a further step in scaling up digital health tools within clinical workflows [59], which, in our case, were evaluated well. An interesting aspect of our study was the collateral use of qualitative data collected from the motivational interviews and by the nurse case manager during follow-up. The qualitative results presented in Table 2 can be used to support the implementation of mHealth tools in different contexts, keeping in mind the inherent limitations of qualitative research data. We do acknowledge that, by using an existing app, the co-design phase was skipped. Also, we did not measure the technological literacy of our older population (average age 69 years), but, according to Martinez-Alcala et al [60], adults older than 60 years, if highly motivated, are capable of learning and acquiring digital literacy skills. Nonetheless, for some of our older patients (24% were 70-79 years old), especially those with physical limitations (eg, visual impairment), the motivation to learn and exploit all the app functionality was low, although the perceived usefulness was high. This agrees with other reports on the use of technology by older adults [61,62]. Another potential limitation was the heterogeneity of the study population, which directly influenced the mean number of hours of use of the NIV machines and precludes any interpretation. Nonetheless, we observe a strength

in terms of the generalizability of the mHealth tools within the heterogeneous population. Finally, a clear limitation of our study was the exclusion of new NIV patients, where the behavioral intervention may have had more impact. This warrants further study.

Conclusions

The behavioral mHealth intervention explored in this study did not show any effect on self-efficacy, adherence with NIV, or quality of life in our population of experienced NIV users. Nonetheless, we showed the potential of the mHealth app to manage complex patients and foster collaborative work among stakeholders. Regarding a clinical service that was graded well in terms of continuity of care and person-centered care, in which the needs of relevant stakeholders are properly addressed, we see the potential to further study mHealth tools to induce behavioral change in home-based ventilated patients as well as in other respiratory therapies.

Acknowledgments

We gratefully acknowledge the administrative help from Ms Gemma Guerra. We also want to thank all the patients that participated in this study and to the European Regional Development Fund, FEDER (NEXTCARE project COMRDI15, NextHealth RIS3Cat Community), the FIS-PITES (PI18/00841 and PI17/01068), and the Generalitat de Catalunya (CERCA programme and 2017 SGR-617 specific grant). This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement n° GA-689802 (CONNECARE).

Conflicts of Interest

JK and JE are employed by Advanced Digital Innovation (UK) Ltd, the creator of the MyPathway app. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Mypathway adaptation for home-based non-invasive ventilation.
[DOCX File , 562 KB - mhealth_v8i4e16395_app1.docx]

Multimedia Appendix 2
Consort flow diagram.
[PDF File (Adobe PDF File), 53 KB - mhealth_v8i4e16395_app2.pdf]

Multimedia Appendix 3
Baseline characteristics and clinical outcomes tables.
[DOCX File , 36 KB - mhealth_v8i4e16395_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 101 KB - mhealth_v8i4e16395_app4.pdf]

References

1. Lassen H. A PRELIMINARY REPORT ON THE 1952 EPIDEMIC OF POLIOMYELITIS IN COPENHAGEN WITH SPECIAL REFERENCE TO THE TREATMENT OF ACUTE RESPIRATORY INSUFFICIENCY. The Lancet 1953 Jan 03;261(6749):37-41. [doi: 10.1016/s0140-6736(53)92530-6] [Medline: 13011944]
2. Clini E, Vitacca M, Foglio K, Simoni P, Ambrosino N. Long-term home care programmes may reduce hospital admissions in COPD with chronic hypercapnia. Eur Respir J 1996 Aug;9(8):1605-1610 [FREE Full text] [doi: 10.1183/09031936.96.09081605] [Medline: 8866580]
3. Köhnlein T, Windisch W, Köhler D, Drabik A, Geiseler J, Hartl S, et al. Non-invasive positive pressure ventilation for the treatment of severe stable chronic obstructive pulmonary disease: a prospective, multicentre, randomised, controlled clinical trial. Lancet Respir Med 2014 Sep;2(9):698-705. [doi: 10.1016/S2213-2600(14)70153-5] [Medline: 25066329]
4. Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. The Lancet Neurology 2006 Feb;5(2):140-147. [doi: 10.1016/s1474-4422(05)70326-4]
5. de Lucas-Ramos P, de Miguel-Díez J, Santacruz-Siminiani A, González-Moro JMR, Buendía-García MJ, Izquierdo-Alonso JL. Benefits at 1 year of nocturnal intermittent positive pressure ventilation in patients with obesity-hypoventi lation syndrome. Respir Med 2004 Oct;98(10):961-967 [FREE Full text] [doi: 10.1016/j.rmed.2004.03.014] [Medline: 15481272]

6. Budweiser S, Heinemann F, Fischer W, Dobroschke J, Wild PJ, Pfeifer M. Impact of ventilation parameters and duration of ventilator use on non-invasive home ventilation in restrictive thoracic disorders. Respiration 2006;73(4):488-494. [doi: 10.1159/000088712] [Medline: 16205051]
7. Ward S, Chatwin M, Heather S, Simonds AK. Randomised controlled trial of non-invasive ventilation (NIV) for nocturnal hypoventilation in neuromuscular and chest wall disease patients with daytime normocapnia. Thorax 2005 Dec;60(12):1019-1024 [FREE Full text] [doi: 10.1136/thx.2004.037424] [Medline: 16299118]
8. Toussaint M, Chatwin M, Soudon P. Mechanical ventilation in Duchenne patients with chronic respiratory insufficiency: clinical implications of 20 years published experience. Chron Respir Dis 2007;4(3):167-177. [doi: 10.1177/1479972307080697] [Medline: 17711917]
9. Lloyd-Owen SJ, Donaldson GC, Ambrosino N, Escarabill J, Farre R, Fauroux B, et al. Patterns of home mechanical ventilation use in Europe: results from the Eurovent survey. Eur Respir J 2005 Jun;25(6):1025-1031 [FREE Full text] [doi: 10.1183/09031936.05.00066704] [Medline: 15929957]
10. Laub M, Berg S, Midgren B, Swedish Society of Chest Medicine. Home mechanical ventilation in Sweden--inequalities within a homogenous health care system. Respir Med 2004 Jan;98(1):38-42 [FREE Full text] [doi: 10.1016/j.rmed.2003.08.005] [Medline: 14959812]
11. de Lucas Ramos P, Rodríguez González-Moro JM, Paz González L, Santa-Cruz Siminiani A, Cubillo Marcos JM. [Current status of home mechanical ventilation in Spain: results of a national survey]. Arch Bronconeumol 2000 Nov;36(10):545-550 [FREE Full text] [Medline: 11149196]
12. Nonoyama ML, McKim DA, Road J, Guerriere D, Coyte PC, Wasilewski M, et al. Healthcare utilisation and costs of home mechanical ventilation. Thorax 2018 Jan 26. [doi: 10.1136/thoraxjnl-2017-211138] [Medline: 29374088]
13. Mansell SK, Cutts S, Hackney I, Wood MJ, Hawksworth K, Creer DD, et al. Using domiciliary non-invasive ventilator data downloads to inform clinical decision-making to optimise ventilation delivery and patient compliance. BMJ Open Respir Res 2018;5(1):e000238 [FREE Full text] [doi: 10.1136/bmjresp-2017-000238] [Medline: 29531743]
14. Janssens J, Borel J, Pépin JL, SomnoNIV Group. Nocturnal monitoring of home non-invasive ventilation: the contribution of simple tools such as pulse oximetry, capnography, built-in ventilator software and autonomic markers of sleep fragmentation. Thorax 2011 May;66(5):438-445. [doi: 10.1136/thx.2010.139782] [Medline: 20971980]
15. Barberan-Garcia A, Ubré M, Roca J, Lacy AM, Burgos F, Risco R, et al. Personalised Prehabilitation in High-risk Patients Undergoing Elective Major Abdominal Surgery: A Randomized Blinded Controlled Trial. Ann Surg 2018 Jan;267(1):50-56. [doi: 10.1097/SLA.0000000000002293] [Medline: 28489682]
16. Troosters T, Maltais F, Leidy N, Lavoie KL, Sedeno M, Janssens W, et al. Effect of Bronchodilation, Exercise Training, and Behavior Modification on Symptoms and Physical Activity in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2018 Oct 15;198(8):1021-1032. [doi: 10.1164/rccm.201706-1288OC] [Medline: 29664681]
17. Slovinec D'Angelo ME, Pelletier LG, Reid RD, Huta V. The roles of self-efficacy and motivation in the prediction of short- and long-term adherence to exercise among patients with coronary heart disease. Health Psychol 2014 Nov;33(11):1344-1353. [doi: 10.1037/hea0000094] [Medline: 25133848]
18. Kenyon DB, Kubik MY, Davey C, Sirard J, Fulkerson JA. Alternative high school students' physical activity: role of self-efficacy. Am J Health Behav 2012 Mar;36(3):300-310 [FREE Full text] [doi: 10.5993/AJHB.36.3.2] [Medline: 22370432]
19. Teixeira PJ, Marques MM. Health Behavior Change for Obesity Management. Obes Facts 2017;10(6):666-673 [FREE Full text] [doi: 10.1159/000484933] [Medline: 29237167]
20. McAuley E, Szabo A, Gothe N, Olson EA. Self-efficacy: Implications for Physical Activity, Function, and Functional Limitations in Older Adults. Am J Lifestyle Med 2011 Jul;5(4) [FREE Full text] [doi: 10.1177/1559827610392704] [Medline: 24353482]
21. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122-147. [Medline: 15064770]
22. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
23. Schwarzer R, Fuchs R. Self-efficacy and health behaviours. In: Conner M, Norman P, editors. Predicting health behaviour: Research and practice with social cognition models. Philadelphia: Open University Press; 1996:163-196.
24. Hernandez C, Aibar J, de Batlle J, Gomez-Cabrero D, Soler N, Duran-Tauleria E, NEXES Consortium. Assessment of health status and program performance in patients on long-term oxygen therapy. Respir Med 2015 Apr;109(4):500-509 [FREE Full text] [doi: 10.1016/j.rmed.2015.01.005] [Medline: 25771036]
25. Cano I, Dueñas-Espín I, Hernandez C, de Batlle J, Benavent J, Contel JC, et al. Protocol for regional implementation of community-based collaborative management of complex chronic patients. NPJ Prim Care Respir Med 2017 Jul 14;27(1):44 [FREE Full text] [doi: 10.1038/s41533-017-0043-9] [Medline: 28710482]
26. Cano I, Alonso A, Hernandez C, Burgos F, Barberan-Garcia A, Roldan J, et al. An adaptive case management system to support integrated care services: Lessons learned from the NEXES project. J Biomed Inform 2015 Jun;55:11-22 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.011] [Medline: 25796455]

27. Ambrosino N, Vitacca M, Dreher M, Isetta V, Montserrat JM, Tonia T, ERS Tele-Monitoring of Ventilator-Dependent Patients Task Force. Tele-monitoring of ventilator-dependent patients: a European Respiratory Society Statement. Eur Respir J 2016 Sep;48(3):648-663 [FREE Full text] [doi: 10.1183/13993003.01721-2015] [Medline: 27390283]
28. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century /* COM/2012/0736 final */ Internet. OPOCE; URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52012DC0736&from=EN [accessed 2018-07-31]
29. MyPathway. URL: https://mypathway.healthcare/ [accessed 2019-03-01]
30. Montplaisir J, Bédard MA, Richer F, Rouleau I. Neurobehavioral manifestations in obstructive sleep apnea syndrome before and after treatment with continuous positive airway pressure. Sleep 1992 Dec;15(6 Suppl):S17-S19. [doi: 10.1093/sleep/15.suppl_6.s17] [Medline: 1470802]
31. Sauro J. Measuring Usability With The System Usability Scale (SUS). URL: https://measuringu.com/umux-lite/ [accessed 2019-02-01]
32. Reichheld F. The One Number You Need to Grow. Harv Bus Rev 2003;81(12):124 [FREE Full text]
33. Uijen AA, Schellevis FG, van den Bosch WJHM, Mokkink HGA, van Weel C, Schers HJ. Nijmegen Continuity Questionnaire: development and testing of a questionnaire that measures continuity of care. J Clin Epidemiol 2011 Dec;64(12):1391-1399. [doi: 10.1016/j.jclinepi.2011.03.006] [Medline: 21689904]
34. Leijten FRM, Hoedemakers M, Struckmann V, Kraus M, Cheraghi-Sohi S, Zemplényi A, SELFIE consortium. Defining good health and care from the perspective of persons with multimorbidity: results from a qualitative study of focus groups in eight European countries. BMJ Open 2018 Aug 30;8(8):e021072 [FREE Full text] [doi: 10.1136/bmjopen-2017-021072] [Medline: 30166294]
35. Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. Value Health 2016;19(6):704-719 [FREE Full text] [doi: 10.1016/j.jval.2016.06.001] [Medline: 27712695]
36. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee TA. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. Respir Med 2008 Apr;102(4):519-536 [FREE Full text] [doi: 10.1016/j.rmed.2007.11.016] [Medline: 18180151]
37. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat 2004 Feb;14(1):97-110. [doi: 10.1081/BIP-120028508] [Medline: 15027502]
38. Philip P, Bioulac S, Altena E, Morin CM, Ghorayeb I, Coste O, et al. Specific insomnia symptoms and self-efficacy explain CPAP compliance in a sample of OSAS patients. PLoS One 2018;13(4):e0195343 [FREE Full text] [doi: 10.1371/journal.pone.0195343] [Medline: 29617414]
39. 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]
40. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
41. Hamine S, Gerth-Guyette E, Faulx D, Green B, Ginsburg A. 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]
42. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
43. Lee J, Choi M, Lee S, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
44. Yi JY, Kim Y, Cho Y, Kim H. Self-management of Chronic Conditions Using mHealth Interventions in Korea: A Systematic Review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
45. Voils CI, King HA, Maciejewski ML, Allen KD, Yancy WS, Shaffer JA. Approaches for informing optimal dose of behavioral interventions. Ann Behav Med 2014 Dec;48(3):392-401 [FREE Full text] [doi: 10.1007/s12160-014-9618-7] [Medline: 24722964]
46. Mogler BK, Shu SB, Fox CR, Goldstein NJ, Victor RG, Escarce JJ, et al. Using insights from behavioral economics and social psychology to help patients manage chronic diseases. J Gen Intern Med 2013 May;28(5):711-718 [FREE Full text] [doi: 10.1007/s11606-012-2261-8] [Medline: 23229906]
47. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
48. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]

49. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
50. Baltaxe E, Czypionka T, Kraus M, Reiss M, Askildsen JE, Grenkovic R, et al. Digital Health Transformation of Integrated Care in Europe: Overarching Analysis of 17 Integrated Care Programs. J Med Internet Res 2019 Sep 26;21(9):e14956 [FREE Full text] [doi: 10.2196/14956] [Medline: 31573914]
51. Donabedian A. The Quality of Care. JAMA 1988 Sep 23;260(12):1743-1748 [FREE Full text] [doi: 10.1001/jama.1988.03410120089033]
52. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
53. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf 2014 Apr;23(4):290-298 [FREE Full text] [doi: 10.1136/bmjqs-2013-001862] [Medline: 24025320]
54. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-Focused, Web-Based Interventions: Systematic Review and Meta-Analysis (Part 2). J Med Internet Res 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
55. Baltaxe E, Cano I, Herranz C, Barberan-Garcia A, Hernandez C, Alonso A, et al. Evaluation of integrated care services in Catalonia: population-based and service-based real-life deployment protocols. BMC Health Serv Res 2019 Jun 11;19(1):370 [FREE Full text] [doi: 10.1186/s12913-019-4174-2] [Medline: 31185997]
56. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
57. Piette JD, Striplin D, Marinec N, Chen J, Trivedi RB, Aron DC, et al. A Mobile Health Intervention Supporting Heart Failure Patients and Their Informal Caregivers: A Randomized Comparative Effectiveness Trial. J Med Internet Res 2015 Jun 10;17(6):e142 [FREE Full text] [doi: 10.2196/jmir.4550] [Medline: 26063161]
58. Lapid MI, Atherton PJ, Clark MM, Kung S, Sloan JA, Rummans TA. Cancer Caregiver: Perceived Benefits of Technology. Telemed J E Health 2015 Nov;21(11):893-902 [FREE Full text] [doi: 10.1089/tmj.2014.0117] [Medline: 26075800]
59. Costa Stutzel M, Filippo MP, Sztajnberg A, da Costa RMEM, Brites ADS, da Motta LB, et al. Multi-part quality evaluation of a customized mobile application for monitoring elderly patients with functional loss and helping caregivers. BMC Med Inform Decis Mak 2019 Jul 22;19(1):140 [FREE Full text] [doi: 10.1186/s12911-019-0839-3] [Medline: 31331309]
60. Martínez-Alcalá CI, Rosales-Lagarde A, Alonso-Lavernia MDA, Ramírez-Salvador JA, Jiménez-Rodríguez B, Cepeda-Rebollar RM, et al. Digital Inclusion in Older Adults: A Comparison Between Face-to-Face and Blended Digital Literacy Workshops Internet. Frontiers in ICT 2018;5:21 [FREE Full text] [doi: 10.3389/fict.2018.00021]
61. 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 06;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
62. Spann A, Stewart E. Barriers and facilitators of older people's mHealth usage: A qualitative review of older people's views. Human Technology 2018 Nov 30;14(3):264-296. [doi: 10.17011/ht/urn.201811224834]

Abbreviations

ICT: information and communication technologies
mHealth: mobile health
NIV: noninvasive ventilation
SEMSA: Self Efficacy in Sleep Apnea

Edited by G Eysenbach; submitted 25.09.19; peer-reviewed by W Zhang, J Xu, J Aikens; comments to author 25.11.19; revised version received 19.01.20; accepted 04.02.20; published 13.04.20.

Please cite as:
Baltaxe E, Embid C, Aumatell E, Martínez M, Barberan-Garcia A, Kelly J, Eaglesham J, Herranz C, Vargiu E, Montserrat JM, Roca J, Cano I
Integrated Care Intervention Supported by a Mobile Health Tool for Patients Using Noninvasive Ventilation at Home: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e16395
URL: http://mhealth.jmir.org/2020/4/e16395/
doi:10.2196/16395
PMID:32281941

Original Paper

Evaluation of a Mobile Phone App for Patients With Pollen-Related Allergic Rhinitis: Prospective Longitudinal Field Study

Manuela Glattacker[1], PhD; Martin Boeker[2], PhD; Robin Anger[1], MSc; Frank Reichenbach[3], PhD; Adrian Tassoni[3]; Rainer Bredenkamp[4]; Juergen M Giesler[1], PhD

[1]Section of Health Care Research and Rehabilitation Research, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[2]Medical Data Science, Institute of Medical Biometry and Medical Statistics, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[3]Clinical Trials Unit, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[4]Clinical Trials Unit UMG, University Medical Center Göttingen, Georg-August-University, Göttingen, Germany

Corresponding Author:
Manuela Glattacker, PhD
Section of Health Care Research and Rehabilitation Research
Medical Center
Faculty of Medicine, University of Freiburg
Hugstetter Strasse 49
Freiburg, 79106
Germany
Phone: 49 761 27036940
Fax: 49 76127073310
Email: manuela.glattacker@uniklinik-freiburg.de

Abstract

Background: Mobile health apps have great potential to support the self-management of chronic conditions such as allergic diseases, which constitute significant challenges in health care. However, the health app market is confusing for users, as it is vast, dynamic, and lacks scientific evidence regarding the effectiveness of the apps on offer. To our knowledge, no health app for pollen-related allergic rhinitis has been evaluated.

Objective: The aim of our study was to evaluate the Husteblume mobile phone health app, developed in Germany to facilitate the self-management of pollen-related allergic rhinitis.

Methods: We evaluated usability and changes in quality of life, health literacy, and self-efficacy for managing one's chronic disease. We conducted 2 online surveys of registered users of the app, 1 before and 1 after the 2017 pollen season, allowing for the analysis of both cross-sectional and longitudinal data in a field setting.

Results: The sample comprised 661 app users at the first measurement point and 143 users at follow-up. The subgroup of study participants at follow-up rated the usability of the app as good or very good. There were no significant changes in patient-reported outcomes such as quality of life, health literacy, and self-efficacy between the 2 measurement points (P>.05). However, those reached at follow-up perceived subjective improvements due to the app: 55.9% (80/143) reported being subjectively better informed about their allergy, 27.3% (39/143) noted improved quality of life, 33.6% (48/143) reported subjectively better coping with their allergy, and 28.0% (40/143) felt better prepared for the consultation with their physician. Finally, 90.9% (130/143) users did not identify any adverse effects of the app.

Conclusions: Despite some methodological caveats, the results of the evaluation of the Husteblume app are encouraging for the subgroup using the app in the long term. However, further studies evaluating the effectiveness of the app are needed.

Trial Registration: German Clinical Trials Register DRKS00011897; https://tinyurl.com/yxxrg9av

(JMIR Mhealth Uhealth 2020;8(4):e15514) doi:10.2196/15514

KEYWORDS

mobile applications; rhinitis, allergic, seasonal; patient reported outcome measures; prospective studies; longitudinal studies; usability; effectiveness

Introduction

Scientific Background

The prevalence of allergic diseases has increased dramatically over the last few decades in many regions of the world [1], and allergic diseases pose a significant challenge in health care [2]. For example, the lifetime prevalence of asthma is 8.6%, and allergic rhinoconjunctivitis, which is a comorbidity in more than 80% of patients with asthma, has a lifetime prevalence of 14.8% [3]. Pollen-related allergic rhinitis is characterized by symptoms such as sneezing, secretion, and conjunctivitis, and is associated with decreased quality of life and performance [3,4]. Effective disease self-management, such as avoiding allergens, and planning medication and everyday life, reduces the burden of pollen-related allergic rhinitis [5]. However, partly due to low adherence to the prescribed medication and a lack of education, allergic rhinitis control is inadequate for many patients [6].

Mobile health (mHealth) apps are a promising way to support the self-management of chronic diseases [7,8]. They have the potential to optimize access to health information and to health interventions in a low-cost way. They can contribute to the empowerment and participation of patients, change health care in a patient-centered, decentralized way, and support health care professionals to treat patients more efficiently [9-11]. Therefore, health apps might increasingly become a "major source of health guidance" ([12] pg 1051) and have the potential to change existing health care delivery pathways [13]. Supporting this, mHealth interventions are growing in popularity worldwide [14]. Over 100,000 mHealth apps are available [10] and are increasingly accepted as a tool to observe and manage health in everyday life [7]. In Germany, a recent population-based survey of more than 4000 participants showed that 61% of participants used a smartphone and, among these, 21% used health apps primarily focusing on smoking cessation, healthy diet, and weight loss [15]. As the mHealth market is one of the fastest-growing areas in health care [16], these numbers will probably continue to increase [17].

However, given the size of and rate of innovation in the health app market, in combination with a lack of objective and valid criteria to assess the quality of health apps [9,16], it can be difficult for end users to choose effective apps. The technical quality of many health apps is problematic, for example, with respect to transparency and data privacy [9,16]. Many apps are developed without the involvement of experts and do not adhere to medical evidence [18]. Further, the lack of evidence regarding the effectiveness of health apps remains a concern. Several reviews have examined the impact of health apps on a specific behavior, such as physical activity [19], adherence to medication [20], or specific diseases such as diabetes [21,22], depression [23], cardiovascular disease [24], chronic renal disease [25], heart failure [26], or chronic obstructive pulmonary disease [27]. Nevertheless, a recent systematic review of 23 systematic reviews assessing the effectiveness of mHealth interventions for different health conditions concluded that the evidence for the efficacy of mHealth interventions is still limited, despite some moderate-quality evidence for improvement across various outcomes [28]. Most of the studies included in the systematic reviews, as well as the reviews themselves, have been criticized for significant methodological limitations [28].

The same is true for health apps for self-management of allergic diseases and asthma. Although the number of apps is growing, there are very few evaluation studies, and the usefulness of these apps is still uncertain [4]. With respect to the effectiveness of apps to facilitate the self-management of patients with asthma, a Cochrane review from 2013 [29] including 2 randomized controlled trials was unable to draw reliable conclusions due to an insufficient number of studies and the considerable degree of heterogeneity between the studies. In 2015, about 200 asthma-related mobile phone apps were available on the iOS and Android platforms [30]. However, a systematic content assessment found that many apps did not include comprehensive information or offer guidance consistent with evidence for asthma self-management. Indeed, 13% of the apps recommended self-care procedures unsupported by evidence [30]. Applying slightly different quality criteria, such as available functions or general quality, a recent review assessing 38 apps found great variation across all of the investigated criteria [31]. Many apps were of low quality, while the major concern was the absence of clinical validation. Finally, a recent systematic review and meta-analysis of 11 studies evaluating the efficacy of mobile technology interventions on clinical outcomes and adherence in individuals with asthma found strong evidence for at least short-term efficacy for asthma management [8]. However, those authors made the criticism that most studies lacked a theoretical basis for their interventions and did not specify the behavior change technique used in the intervention [8].

Very few studies have evaluated the impact of health apps in allergic rhinitis [4]. However, in 2015, a worldwide consortium proposed a plan for the use of mHealth technology in the management of allergic rhinitis (MACVIA-ARIA Sentinel Network [MASK]) [4]. Following this group, the MASK-rhinitis app is the first app to have been tested in a pilot study [32]. To our knowledge, no study has evaluated a health app for pollen-related allergic rhinitis. Such an app would be a potentially effective way to target and supply many people with pollen allergies with pollen information. This would be of great importance, as it could help people to plan medication and everyday life and contribute to a better quality of life [33].

Objectives

The aim of our study was to evaluate the Husteblume mHealth app for patients with pollen-related allergic rhinitis, with respect to its usability and changes in quality of life, health literacy, and self-efficacy for managing this chronic disease.

Methods

The Husteblume Health App

The Husteblume health app was developed by a German health insurance company (Techniker Krankenkasse, Hamburg, Germany). The app is available for download free of charge from the Apple and Google app stores. It aims to support the self-management of its users and provides functions that allow them to (1) register their allergy-related symptoms and their

medication in a diary, (2) retrieve prognostic information on the type and amount of pollen expected at the user's present location, (3) retrieve information about the relationship between the user's individual symptoms, the pollen load, and the medication the user has been taking and then graphically display this information for a period of time (week or month), (4) access information about available treatments for a specific type of pollen allergy and its symptom burden, (5) access a dictionary providing information on allergens and cross-allergies, and (6) perform a self-test to assess their allergic rhinitis.

As is apparent from these features, the app uses behavior change techniques [8], most notably *information about antecedents* of symptoms, *information about consequences* of behaviors, and *self-monitoring.* It thus allows deliberate planning of behaviors aimed at avoiding or otherwise managing situations that tend to increase rhinitis symptoms. The app was developed on the basis of current medical guidelines and provides information on its functionality and accountability (eg, the security and privacy of user data and the timeliness of the information that the app provides).

Design, Recruitment, and Ethics

We conducted an online survey during the 2017 pollen season using a design with 2 measurement points that combined a cross-sectional and a longitudinal approach in a field setting.

To be eligible, participants had to be registered users of the app, aged 18 years or older, and allergic to birch or grass pollen, or both, as these most frequently elicit allergic rhinitis. The first measurement was taken before the allergen season (T0) started and the second one, after it had ended (T1). Since the birch and grass pollen seasons cover the period from late March to mid-May and from mid-May to July, respectively, we set the first measurement point to April 1 and the second to August 31. We chose this time interval to ensure that app users would have a sufficiently broad basis of experience to rate the usability of the app and possible changes in patient-reported outcomes.

Potential participants were contacted via 3 routes. During the release of an update of the app by the provider, users were notified of the study via a teaser within the app and asked to participate. Potential users were notified of the study through a push message when browsing the provider's website that generally aims to provide insurees with information on insurance benefits. In addition, the app provider referred potential participants to the study via various print materials.

If potential users agreed to participate from within the app, they immediately were linked to the survey website at our institution. There, they were presented with detailed information about the study. If consenting to participate, users had to indicate this by ticking 3 consent boxes stating they were at least 18 years old, agreed with the data protection statement, and agreed to participate in the study. Thereafter, they were immediately referred to the survey. Multimedia Appendix 1 shows the challenges of programming the survey.

The study received approval from the Ethics Committee of the University of Freiburg (reference number 33/17) and was registered with the German Clinical Trials Register (DRKS00011897).

Measures

We measured all variables using either self-constructed items or validated questionnaires. For this paper, 2 of the authors (JMG, MG) translated the survey questions and responses used from German to English.

To measure participants' *access to the app* and their *previous use* of health apps, we used 4 items at T0. These covered (1) access routes, (2) previous use of the app, (3) if used, frequency and duration of previous use of the app, and (4) use of other health apps.

Sociodemographic variables (assessed at T0) included age, sex, nationality, marital status, whether living with a partner, education, and occupational status.

Allergy- and treatment-related variables (measured at T0) covered physician-certified allergy diagnosis, type of allergy, allergen(s) with hyperreactivity, time since allergy onset, degree of impairment during pollen season, use of medication, comorbidities (eg, asthma or sinusitis), and smoking status.

We measured *usability* at T1 using 13 items that were based on 2 established instruments: the System Usability Scale (SUS) [34] and the Modular Evaluation of Components of User Experience (meCUE) [35]. Further, the qualitative work of Grindrod and colleagues [36] provided useful information on dimensions of mobile app user experiences. Therefore, we used items addressing the dimensions of personal usefulness, simplicity, accessibility, functionality, and design of the app. However, since we found the wording of the German translations of both the SUS [37] and the meCUE [35] not entirely satisfactory, we decided to write new items addressing content that was in part covered by the SUS (4 items) or the meCUE (5 items). Finally, we added 1 item asking for a summary evaluation of the app. All the usability items were answered on 5-point scales ranging from "not at all true" to "completely true" with a middle category of "partly true, partly not true".

We measured *user behavior during the pollen season* at T1 using 2 self-constructed items. These required participants to rate their average frequency of app usage per week during the pollen season and to report whether they had used the app in relation to 3 levels of symptom burden (low, medium, and high, multiple responses possible).

We measured *perceived effects* of using the app *on self-management and illness behavior* by self-constructed items at T1 that focused on changes participants might have perceived as resulting from using the app during the pollen season. These items covered the following aspects: (1) participants' knowledge about the allergy, (2) frequency of health service consultations due to the disease, (3) form of preparation for health service consultations, (4) experience of negative effects due to the app, (5) adherence to physicians' advice, (6) management of the allergy, (7) perceived improvements to their condition, and (8) perceived improvements to their quality of life in general. Responses to these items were measured on 5-point scales ranging either from "not at all true" to "completely true" or from "deteriorated" to "improved".

We measured *health-related quality of life* at T0 and T1 using the Quality of Life in Allergic Rhinitis (FL-Heu) questionnaire [38]. The FL-Heu consists of 32 items combined into 7 scales measuring quality of life in terms of impairment in various domains and 1 generic item addressing the respondent's current health in general. The scales cover the domains of sleep, eyes, nose, general symptoms, social relationships, being affected by the disease, and emotional impairment. Items are answered on a 7-point scale. Higher item and scale scores represent greater impairment and thus lower quality of life. Scale consistencies (Cronbach alphas) have been reported to range from .74 to .90. The instrument has been found to be sensitive to change [38,39].

We measured *health literacy* at both time points with the Health Education Literacy of Patients With Chronic Musculoskeletal Diseases (HELP) [40] questionnaire. HELP consists of 18 items combined into 3 scales measuring comprehension of medical information, applying medical information, and communicative competence in patient-provider interactions. Scale internal consistencies (Cronbach alpha) range from .88 to .95. The scales are compatible with a Rasch model and preliminary evidence of their validity is available [40]. Higher scale scores represent higher levels of health literacy.

We measured *self-efficacy for managing one's chronic condition* at T0 and T1 with a scale by Lorig [41,42]. Its 6 items ask how confident one feels that one can do various things without interference from one's chronic condition. Items are answered on a 10-point scale ranging from "not at all confident" to "totally confident." An individual's scale score is represented by the mean of their item responses. The Cronbach alphas reported for this scale exceed .90 [41,42]. Higher scores indicate higher self-efficacy.

In total, the survey included 92 items at T0 and 85 items at T1. At each time point, questions were presented in a linear order across a total of 17 pages at T0 and 16 pages at T1. In some instances, we used adaptive questioning (ie, in regard to nonemployment).

Data Analysis

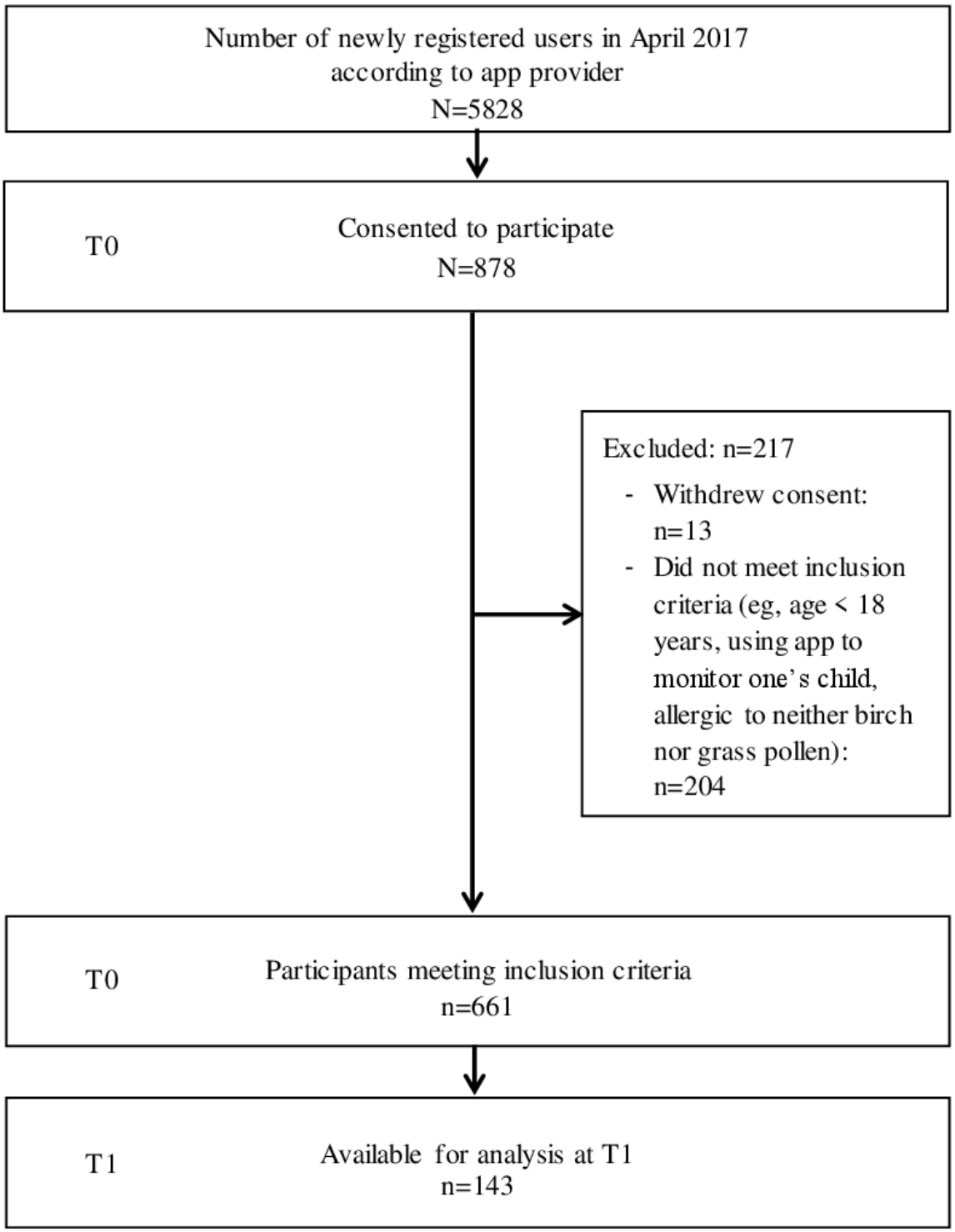
Data analysis included all participants who had provided data at either the first or the first *and* second measurement points. First, we computed descriptive statistics for sociodemographic, medical, and app access parameters measured at T0. We also determined descriptive statistics for user behaviors and perceived effects and changes in health and illness behaviors measured at T1. Missing data were not imputed. When computing scale scores, we handled missing data in compliance with the recommendations of the authors of the respective questionnaire. To estimate whether completers of the survey differed from dropouts in respect to the characteristics measured at T0, we computed chi-square analyses and *t* tests for independent groups. To determine changes from T0 to T1, we performed paired *t* tests (2-tailed) for quality of life, health literacy, and self-efficacy for managing one's chronic condition. We set type I error probability to P=.05 throughout. All computations were performed with the statistical software IBM SPSS version 25 (IBM Corporation).

Results

Participant Flow, Sample Characteristics, and Dropout Comparisons

Figure 1 presents the participant flow from the first (T0) to the second point of measurement (T1). At T0, 5828 persons had registered as users of the app. Of these, 878 (15.1%) consented to participate; 13 withdrew their consent and thus were excluded from the study. In addition, 204 persons did not meet the inclusion criteria and were excluded. Thus, at T0 data from 661 individuals were available for analysis.

Figure 1. Participant flow from first (T0) to second measurement point (T1).



At the second point of measurement, 143 persons completed the study (21.6% of those participating at T0 and 2.45% of those who initially had registered for the app). It should be noted, however, that the total number of active users of the app decreased drastically from registration to T1, that is, from 5828 in April to 191 in August 2017. From this perspective, the proportion of participants at T1 was 74.9% of those who were still actively using the app at that time.

It took participants a mean of 11 (SD 6.65) minutes at T0 and 10 (SD 6.01) minutes at T1 to complete the survey. At T0, 91.8% (607/661) of those starting the survey completed it; at T1 this proportion was 88.8% (127/143).

Participants were a mean of 39 years old, and 58.6% (387/660) were female (see Table 1 for more information on sociodemographic and medical characteristics). A total of 73.2% (484/661) rated their impairment during the pollen season as strong or very strong (results not displayed in Table 1). Of the

sample, 24.7% (163/661) had not used the app before, and 71.9% (358/498) of those who had previously used the app indicated they had done so for 3 to 4 or for 5 or more days per week. A total of 53.7% (355/661) reported not currently using any other health apps.

Comparisons between dropouts (n=518) and completers (n=143) of the study indicated that the 2 groups were comparable with regard to most of the variables. The only significant difference was that completers were older on average than dropouts (mean age 42.0, SD 21.0 vs mean 38.6, SD 12.7 years, respectively; t_{659}=2.82; P<.01).

Table 1. Sociodemographic and medical sample characteristics (N=661, unless otherwise indicated).

Characteristics	Values
Age (years), mean (SD)	39.4 (12.6)
Sex (n=660), n (%)	
Male	273 (41.4)
Female	387 (58.6)
Marital status, n (%)	
Single	290 (43.9)
Married	327 (49.5)
Divorced or separated	40 (6.1)
Widowed	4 (0.6)
Years of education, n (%)	
9	23 (3.5)
10	144 (21.8)
11	14 (2.1)
12	100 (15.1)
13	371 (56.1)
0	2 (0.3)
Other	7 (1.1)
Employed, n (%)	
Yes	579 (87.6)
No	82 (12.4)
Allergy diagnosis by physician, n (%)	
Yes	648 (98.0)
No	13 (2.0)
Allergen, n (%)	
Birch pollen (only)	165 (25.0)
Birch and grass pollen	398 (60.2)
Grass pollen (only)	98 (14.8)
Duration of allergy (years), n (%)	
<1	5 (0.8)
1-4	88 (13.3)
5-10	131 (19.8)
>10	437 (66.1)
Use of medication against allergic rhinitis (n=657), n (%)	
Yes	589 (89.6)
No	68 (10.4)

App Usage During Pollen Season and Usability

Of those participating in the study at T1, 73.4% (105/143) reported having used the app during the pollen season for 3 to 4, or for 5 or more days per week. A total of 87.4% (125/143) rated the app as easy to use and 1.4% (2/143) rated it as too complicated or requiring too much prior knowledge (Table 2). Access to the app was rated positively or very positively by 93.0% (133/143) of the participants, and a similar proportion provided an overall positive evaluation. The functionality of the app, its design, and the personal benefit of using it also received positive ratings; for example, 84.6.% (121/146) of participants at T1 intended to use the app in the future. However, there was more variation between these items and lower proportions of respondents agreed with them.

Table 2. Descriptive statistics of app usability items (137≤n≤143; in descending order of means).

Dimensions and items (abbreviated)	Answer scores			
	Mean (SD)	(mostly) not true, n (%)[a]	partly true, partly not true, n (%)[a]	(mostly) true, n (%)[a]
Simplicity				
Easy to see how to operate the app[b]	4.42 (0.68)	2 (1.4)	9 (6.3)	126 (88.1)
App easy to handle[b]	4.33 (0.73)	2 (1.4)	16 (11.2)	125 (87.4)
App too complicated[b]	1.67 (0.86)	126 (88.1)	10 (7.0)	7 (4.9)
App requires too much prior knowledge to be operated effectively[b]	1.53 (0.73)	126 (88.1)	7 (4.9)	4 (2.8)
Functionality				
Results reporting clearly arranged[b]	3.87 (0.83)	9 (6.3)	33 (23.1)	101 (70.6)
Functions well integrated[b]	3.80 (0.75)	6 (4.2)	36 (25.2)	95 (66.4)
Personal benefit				
Will use app in the future[b]	4.15 (0.78)	22 (15.4)	16 (11.2)	121 (84.6)
App is very useful[b]	4.01 (0.79)	4 (2.8)	31 (21.7)	108 (75.5)
Functions appropriate for my goals[b]	3.74 (0.93)	14 (9.8)	36 (25.2)	87 (60.8)
App provides much useful information[b]	3.63 (0.85)	12 (8.4)	48 (33.6)	77 (53.8)
Design				
App design attractive[b]	3.89 (0.75)	5 (3.5)	34 (23.8)	104 (72.7)
Access				
App easy to download and install[b]	4.79 (0.57)	2 (1.4)	2 (1.4)	133 (93.0)
Overall evaluation[a]				
All in all, app is running well[b]	4.48 (0.63)	2 (1.4)	4 (2.8)	131 (91.6)

[a]Percentages were computed based on all 143 respondents at T1. This includes respondents who had missing data for individual variables. Thus, row percentages do not always add up to 100%.

[b]Response categories and scores: “not true at all” (1), “mostly not true” (2), “partly true, partly not true” (3), “mostly true” (4), “completely true” (5).

Perceived Changes

A total of 2.1% (3/143) of participants reported negative effects of the app (Table 3). By contrast, 55.9% (80/143) felt better informed about their rhinitis, and between 20.3% (29/143) and 33.6% (48/143) indicated that they felt supported with respect to adherence to medication, preparing for medical visits, or coping. The vast majority at least partially agreed that using the app had led to improvements in their quality of life or their allergic rhinitis.

Table 3. Descriptive statistics of items assessing perceived changes in information, coping, medical visits, adherence, quality of life, and negative effects as a consequence of app use (n=133).

Using the app...	Answer scores			
	Mean (SD)	(mostly) not true, n (%)[a]	partly true, partly not true, n (%)[a]	(mostly) true, n (%)[a]
makes me feel better informed about my allergic rhinitis[b]	3.59 (1.00)	22 (15.4)	31 (21.7)	80 (55.9)
helps me cope better[b]	3.11 (1.04)	36 (25.2)	49 (34.3)	48 (33.6)
helps me prepare better for my medical visits[b]	2.88 (1.11)	46 (32.2)	47 (32.9)	40 (28.0)
helps me adhere to my doctor's recommendations[b]	2.62 (1.09)	64 (44.8)	40 (28.0)	29 (20.3)
improved my quality of life[c]	3.32 (0.53)	0 (0.0)	94 (65.7)	39 (27.3)
improved my allergic rhinitis[c]	3.10 (0.30)	0 (0.0)	120 (83.9)	13 (9.1)
has negative effects on me, too[b]	1.36 (0.62)	140 (90.9)	1 (0.7)	2 (1.4)

[a]Percentages were computed based on all 143 respondents at T1, including those with missing data for individual variables. Thus, row percentages do not always add up to 100%.

[b]Response categories and scores: "not true at all" (1), "mostly not true" (2), "partly true, partly not true" (3), "mostly true" (4), "completely true" (5).

[c]Response categories and scores: "worsened" (1), "worsened somewhat" (2), "neither worsened nor improved" (3), "improved somewhat" (4), "improved" (5).

Changes in Quality of Life, Health Literacy, and Self-Efficacy from T0 to T1

As Table 4 shows, we detected almost no significant changes between the 2 measurement points for quality of life variables, health literacy, or self-efficacy for managing one's chronic condition; the only exception was less impairment of quality of life by nasal symptoms.

Table 4. Changes in rhinitis-related quality of life, health literacy, and self-efficacy of coping with chronic disease across the observation period (116≤n≤127).

Subscale and attribute	T0 score, mean (SD)	T1 score, mean (SD)	*r*	*t*	*df*	*P* value
Impairment of quality of life						
Sleep[a]	34.48 (19.23)	34.48 (22.68)	.538	0.00	120	>.99
Eyes[a]	39.38 (21.57)	35.85 (20.94)	.490	1.79	118	.08
Nose[a]	62.18 (21.11)	57.45 (24.26)	.483	2.23	119	.03
General symptoms[a]	41.32 (18.11)	40.70 (18.14)	.625	0.43	118	.67
Social relationships[a]	39.05 (19.77)	37.75 (20.93)	.681	0.88	118	.38
Impairment through disease[a]	54.35 (20.66)	53.81 (20.19)	.530	0.29	115	.77
Emotional impairment[a]	36.68 (17.12)	35.19 (18.21)	.642	1.08	117	.28
General health[b]	3.58 (1.32)	3.33 (1.35)	.281	1.67	117	.10
Health literacy[c]						
Understand medical information	82.28 (16.26)	80.54 (17.76)	.722	1.54	126	.13
Apply medical information	85.28 (13.92)	83.19 (15.50)	.551	1.68	126	.10
Talk to clinicians	80.88 (19.33)	78.74 (21.03)	.686	1.50	126	.14
Self-efficacy[d]						
Managing chronic disease	7.29 (1.90)	7.36 (1.75)	.636	–0.48	125	.63

[a]Scale scores range from 0 to 100. Higher scores indicate greater impairment.

[b]Scores range from 0 to 7. Higher scores indicate greater impairment.

[c]Scores range from 0 to 100. Higher scores indicate higher health literacy.

[d]Scores range from 1 to 10. Higher scores indicate higher self-efficacy beliefs.

Discussion

Principal Findings

We are not aware of any studies evaluating a health app in the context of pollen-related allergic rhinitis, and there are few German health app studies in the literature [43]. Therefore, the aim of our study was to evaluate the Husteblume mobile app among patients with pollen-related allergic rhinitis with respect to its usability and changes in patient-reported outcomes.

The Husteblume app meets many of the quality criteria that have been suggested to determine the quality of health apps [12,44]. It was developed on the basis of current medical guidelines and provides functionality, credibility, and accountability information. Pollen-specific quality criteria such as providing comprehensive information on pollination, guiding management of the pollen allergy, allowing the documentation of symptoms, and informing the user about the developer of the app [33] are implemented. The app uses several behavior change techniques [8] with a focus on self-monitoring, one of the most common behavior change techniques applied in apps across a broad range of health issues [11].

The major results of our study showed that the usability of the app was largely rated positively by its users. While we observed few significant changes in patient-reported outcomes over time, participants indicated subjectively perceived changes of varying degree in relation to being better informed about their condition, to better coping with it, or to their quality of life.

Usability, which is a critical factor for the continuous application and the effectiveness of health apps [43] and is therefore part of several taxonomies for assessing health apps [12,44], was rated as good or very good in the subgroup of study participants who were followed up. This result might also be responsible for their comparatively high level of adherence to the app. Of those we followed up, 51.0% (73/143) used the app for more than 6 months, and 84.6% (121/143) of users who were reached at the second measurement time point were motivated to use the app in the future. However, in line with studies showing that many stop using a health app shortly after downloading it [45], a high number of participants dropped out of the study. The number of active users of the app decreased from 5828 in April to 191 in August 2017. This resulted in only 2.45% (143/5828) of those who had registered for the app still being involved in its evaluation at follow-up. While the proportion of ongoing users who were followed up was relatively high (143/191, 74.9%), findings regarding the usability of the app were gained in a relatively small subgroup of long-term active users. We do not know whether those who stopped using the app were discouraged by a perceived lack of usability or whether there were other reasons for dropping out from active use, such as the timing of the T1 assessment. The T1 assessment took place well after the end of the pollen season, and these people with allergy might have had little reason to continue using the app after the pollen season had finished. Thus, in future studies it would be useful to analyze users' reasons for ceasing to use the app.

With respect to patient-reported outcomes, such as quality of life, health literacy, and self-efficacy, there were no significant changes between the 2 measurement points. Nevertheless, those participants who were followed up perceived subjective improvements due to the app; half felt better informed about the allergy, one-quarter reported improved quality of life, and one-third reported subjectively better allergy self-management and being better prepared for the consultation with their physician. Finally, most users could not identify any adverse effects of the app. However, the high dropout rate should be kept in mind when interpreting these findings.

Strengths and Limitations

The results of our study must be interpreted in the light of 3 major limitations. The first limitation is that selection bias may reduce the generalizability of the results. Initially, only 15.07% (878/5828) of all registered Husteblume app users could be reached as study participants, so the results cannot be generalized to all users. By follow-up, the number of active users of the app had decreased drastically since registration, and the participation rate of 21.6% (143/661) of the initially included study participants further limits the generalizability of the results. Having said that, with the exception of younger age, there were no significant differences between dropouts and those who were followed up. Moreover, we were able to reach 74.9% (143/191) of the active users at the time of the follow-up assessment. Nonetheless, our results predominantly represent the situation in a relatively small subgroup of long-term active users.

Compared with the MASK sample of more than 2500 users from 20 countries [32], our sample comprised a comparatively high proportion of women (59% versus 43% in the MASK study), and slightly older users (mean age 39 years versus 33 years in the MASK study). Our sample was characterized by high educational status and long-term patients who had allergy for over 10 years. While these characteristics may have affected the results of the evaluation to some extent, the direction of the influence is not clear. On the one hand, the literature shows that selection bias is a problem for many studies using information technology tools [32], as well as for the technology tools themselves. On the other hand, the role of patient characteristics such as age, sex, or disease severity in the use and effectiveness of health apps is still unclear [46]. Therefore, further studies focusing on underrepresented patient groups are needed.

The second major limitation relates to design aspects of our study. We applied a single-arm, noncontrolled study design. Similar study designs are common in evaluations of, for example, asthma-related health apps [47], and there are several challenges associated with designing evaluation studies for health apps [9]. Barriers to evaluation can be seen in the mismatch between the rapid pace of mHealth innovation and rather rigid research designs; in the difficulty of applying characteristics of gold-standard research designs (eg, randomized controlled trial) such as blinding; in the selection of the appropriate app-related outcome variables such as patient autonomy, transparency, or satisfaction with information; and in the lack of psychometrically sound measures of many of these outcome domains [9,48]. While trials of higher

methodological quality are needed [11], the appropriate research standard in this area is still being debated [49]. Taken together, external or ecological validity needs to be maximized without reducing the study's internal validity [50].

Concerning the design of our study, that we refrained from reporting power analyses due to practical aspects of the study can also be criticized. Our potential sample was a priori limited to those who registered for the app during the 2017 pollen season. However, a power analysis for paired *t* tests showed that a sample of 115 participants would be required to detect small effects (of 0.24 with an alpha of .05 and 1–beta of .80). Therefore, the study was adequately powered.

Furthermore, while we used validated tools for the assessment of changes in quality of life, health literacy, self-efficacy, and—in part—usability, we also used self-constructed single items, for example, to assess perceived changes. The development of these items was based on our experience with similar studies and on clinical expertise; however, we did not pilot test or validate the items beforehand.

A challenge in the context of the evaluation of a pollen-related health app is to select the optimal measurement time points, which ensure that disease burden or quality of life (which were outcomes in our study) do not decrease simply due to seasonal differences in pollination. With our study design, this cannot be completely excluded. Applying the selected measurement points, we attempted to cover the start and the end of the pollen season for both birch and grass pollen allergies, ensuring that app users had a sufficiently broad basis of experience to rate the usability of the app and perceived changes in health-related outcomes. However, the third major limitation of this study is that the timing of the T1 assessment may have been too late. The vast majority of users had stopped using the app by that time, and in addition to the possibility that they didn't find the app to be useful, it is also plausible that they stopped using the app because the pollen season was over and they no longer needed the app.

The strengths of our study include the relatively large sample, the assessment of the majority of active users at follow-up, and the combination of a cross-sectional and longitudinal study design focusing on different outcome measures, such as usability and patient-reported outcomes.

Bearing in mind that health apps with even a small positive effect on health might still be a valuable intervention if the population-level reach is high [11], we conclude that the results of the evaluation of the Husteblume app are encouraging. However, further studies addressing the abovementioned limitations are needed.

Conclusion

Despite the obvious potential of health apps, high-quality apps are still rare. Evidence is still lacking for their usability, integration into treatment processes, and effectiveness.

To our knowledge, this is the first study to evaluate a health app for pollen-related allergic rhinitis. Despite limitations due to methodological problems, the study showed that the subgroup of study participants at follow-up rated the usability of the Husteblume app as good, and that these users perceived many subjective improvements due to the app. Therefore, we conclude that the results of the evaluation of the Husteblume app are encouraging, but that further studies evaluating the effectiveness of the app are needed.

Acknowledgments

The project was funded by the German health insurance company Techniker Krankenkasse.

We would like to thank the participating users of the Husteblume mobile app. Furthermore we thank Prof Dr med Karl-Christian Bergmann, who was involved as a clinical expert in the development of the content of the app and who supported the construction of disease-related evaluation variables. We thank Andrea Feigel and James Balmford for proofreading and their critical review of the manuscript. The article processing charge was funded by the German Research Foundation (DFG) and the Albert Ludwigs University of Freiburg in the Open Access Publishing funding program.

Authors' Contributions

MG, MB, and RB developed the study design and supervised project management, data management, and statistics. FR was responsible for project management. RA, FR, and JMG were responsible for data management and statistics. AT was responsible for development of the online survey. All authors were involved in the preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Challenges associated with programming the survey.
[DOCX File , 17 KB - mhealth_v8i4e15514_app1.docx]

References

1. Langen U, Schmitz R, Steppuhn H. [Prevalence of allergic diseases in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May;56(5-6):698-706. [doi: 10.1007/s00103-012-1652-7] [Medline: 23703488]
2. Bergmann K, Heinrich J, Niemann H. Current status of allergy prevalence in Germany: position paper of the Environmental Medicine Commission of the Robert Koch Institute. Allergo J Int 2016;25:6-10 [FREE Full text] [doi: 10.1007/s40629-016-0092-6] [Medline: 27069844]
3. Vosteen KH. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). Unfallchirurgie 1992 Feb;18(1):55-56. [doi: 10.1007/bf02588242]
4. Huang X, Matricardi PM. Allergy and asthma care in the mobile phone era. Clin Rev Allergy Immunol 2019 Apr;56(2):161-173. [doi: 10.1007/s12016-016-8542-y] [Medline: 27209270]
5. Gerhards C, Gregoriano C, Moser C, Dieterle T, Leuppi J. Allergisches Asthma und allergische Rhinitis. Ther Umsch 2016;73(6):349-355. [doi: 10.1024/0040-5930/a000803] [Medline: 27646433]
6. Hellings PW, Fokkens WJ, Akdis C, Bachert C, Cingi C, Dietz de Loos D, et al. Uncontrolled allergic rhinitis and chronic rhinosinusitis: where do we stand today? Allergy 2013 Jan;68(1):1-7. [doi: 10.1111/all.12040] [Medline: 23025484]
7. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632. [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
8. Miller L, Schüz B, Walters J, Walters EH. Mobile technology interventions for asthma self-management: systematic review and meta-analysis. JMIR Mhealth Uhealth 2017 May 02;5(5):e57 [FREE Full text] [doi: 10.2196/mhealth.7168] [Medline: 28465281]
9. Albrecht UV. Kapitel Kurzfassung. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Hannover, Germany: Medizinische Hochschule Hannover; 2016:14-47.
10. mHealth. Brussels, Belgium: European Commission; 2018 Jun 12. URL: https://ec.europa.eu/digital-single-market/en/mhealth [accessed 2020-03-09]
11. 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]
12. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
13. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
14. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies. Global Observatory for eHealth series -- Volume 3. Geneva, Switzerland: World Health Organization; 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2020-03-09]
15. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
16. Bork U, Weitz J, Penter V. Apps und mobile health: Viele Potenziale noch nicht ausgeschöpft. Deutsches Ärzteblatt 2018;115(3):A62.
17. Jahns RG, Houck P. Mobile Health Market Report 2013-2017. The Commercialization of Mhealth Applications (Vol. 3). Berlin, Germany: research2guidance; Mar 04, 2013.
18. Subhi Y, Bube SH, Rolskov BS, Skou TAS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
19. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
20. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
21. Bonoto BC, de Araujo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
22. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
23. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]

24. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
25. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
26. Cajita MI, Gleason KT, Han H. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22. [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
27. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Abu AW. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 Sep 01;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
28. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
29. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;11:CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
30. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med 2015;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]
31. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
32. Bousquet J, Caimmi DP, Bedbrook A, Bewick M, Hellings PW, Devillier P, et al. Pilot study of mobile phone technology in allergic rhinitis in European countries: the MASK-rhinitis study. Allergy 2017 Jun;72(6):857-865. [doi: 10.1111/all.13125] [Medline: 28072463]
33. Berger U, Bastl K, Berger M, Bergmann K, Kmenta M. [Pollen information applications for pollen allergy sufferers – requirements for an useful app ad advantages for persons concerned]. Atemwegs- und Lungenkrankheiten 2017;43(02):63-67. [doi: 10.5414/ATX02171]
34. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
35. Minge M, Riedel L. meCUE - Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In: Boll S, Maass S, Malaka R, editors. Mensch & Computer 2013. Munich, Germany: Oldenburg Verlag; 2013:89-98.
36. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
37. Reinhardt W, Ruegenhage E, Rummel B. SAP User Experience Community. 2015 Jan 13. System Usability Scale – jetzt auch auf Deutsch URL: https://experience.sap.com/skillup/system-usability-scale-jetzt-auch-auf-deutsch/ [accessed 2020-03-13]
38. Kupfer J, Roetting U, Niemeier V, Brosig B, Gieler U. Lebensqualität bei Patienten mit allergischer Rhinitis. Entwicklung und Validierung eines Fragebogens. In: Bullinger M, Siegrist J, Ravens-Sieberer U, editors. Jahrbuch der Medizinischen Psychologie; Bd 18; Lebensqualitätsforschung aus medizinpsychologischer und -soziologischer Perspektive. Göttingen, Germany: Hogrefe; 2000:194-205.
39. Kupfer J, Brosig B, Gottwald B, Niemeier VV, Gieler U. Überprüfung der Sensitivität eines Fragebogens zur Lebensqualität bei Patienten mit allergischer Rhinitis. Allergologie 2001;24:300-308. [doi: 10.5414/ALP24300]
40. Farin E, Ullrich A, Nagl M. Health education literacy in patients with chronic musculoskeletal diseases: development of a new questionnaire and sociodemographic predictors. Health Educ Res 2013 Dec;28(6):1080-1091. [doi: 10.1093/her/cyt095] [Medline: 24113084]
41. Stanford Patient Education Research Center. Chronic Disease Self-Management Program Questionnaire Book. Palo Alto, CA: Stanford University School of Medicine; 2007.
42. Freund T, Gensichen J, Goetz K, Szecsenyi J, Mahler C. Evaluating self-efficacy for managing chronic disease: psychometric properties of the six-item Self-Efficacy Scale in Germany. J Eval Clin Pract 2013 Feb;19(1):39-43. [doi: 10.1111/j.1365-2753.2011.01764.x] [Medline: 21883720]
43. Kuhn B, Amelung V. Kapitel 4: Gesundheits-Apps und besondere Herausforderungen. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Hannover, Germany: Medizinische Hochschule Hannover; 2016:100-114.
44. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. J Med Syst 2017 Aug;41(8):128. [doi: 10.1007/s10916-017-0776-6] [Medline: 28735372]
45. 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]
46. Wildenbos GA, Peute LW, Jaspers MWM. Impact of patient-centered eHealth applications on patient outcomes: a review on the mediating influence of human factor issues. Yearb Med Inform 2016 Nov 10(1):113-119. [doi: 10.15265/IY-2016-031] [Medline: 27830238]

47. Cook KA, Modena BD, Simon RA. Improvement in asthma control using a minimally burdensome and proactive smartphone application. J Allergy Clin Immunol Pract 2016;4(4):730-737.e1 [FREE Full text] [doi: 10.1016/j.jaip.2016.03.005] [Medline: 27107690]
48. Wakefield BJ, Turvey CL, Nazi KM, Holman JE, Hogan TP, Shimada SL, et al. Psychometric properties of patient-facing eHealth evaluation measures: systematic review and analysis. J Med Internet Res 2017 Oct 11;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
49. Roess A. The promise, growth, and reality of mobile health - another data-free zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]
50. Albrecht U, Kuhn B, Land J, Amelung VE, von Jan U. [Assessing the benefits of digital health solutions in the societal reimbursement context]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar;61(3):340-348. [doi: 10.1007/s00103-018-2696-0] [Medline: 29368121]

Abbreviations

FL-Heu: Quality of Life in Allergic Rhinitis
HELP: Health Education Literacy of Patients With Chronic Musculoskeletal Diseases
MASK: MACVIA-ARIA Sentinel Network
meCUE: Modular Evaluation of Components of User Experience
mHealth: mobile health
SUS: System Usability Scale

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by F Alvarez, M Bennie; comments to author 03.10.19; revised version received 10.12.19; accepted 10.02.20; published 17.04.20.

Please cite as:
Glattacker M, Boeker M, Anger R, Reichenbach F, Tassoni A, Bredenkamp R, Giesler JM
Evaluation of a Mobile Phone App for Patients With Pollen-Related Allergic Rhinitis: Prospective Longitudinal Field Study
JMIR Mhealth Uhealth 2020;8(4):e15514
URL: http://mhealth.jmir.org/2020/4/e15514/
doi:10.2196/15514
PMID:32301735

Original Paper

Needs, Experiences, and Views of People With Rheumatic and Musculoskeletal Diseases on Self-Management Mobile Health Apps: Mixed Methods Study

Aurelie Najm[1,2], MD; Heidi Lempp[3], PhD; Laure Gossec[4,5], MD, PhD; Francis Berenbaum[6,7], MD, PhD; Elena Nikiphorou[3], MBBS/BSC, MD (Res)

[1]Department of Rheumatology, Nantes University Hospital, Nantes, France

[2]INSERM UMR 1238, Nantes, France

[3]Centre for Rheumatic Diseases, Department of Inflammation Biology, School of Immunology and Microbial Sciences, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom

[4]Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM UMR S1136, Paris, France

[5]Rheumatology department, AP-HP, Pitié Salpêtrière hospital, Paris, France

[6]Department of Rheumatology, AP-HP, Saint Antoine Hospital, Paris, France

[7]Sorbonne Université and INSERM CRSA, Paris, France

Corresponding Author:
Aurelie Najm, MD
Department of Rheumatology, Nantes University Hospital
1, Place Alexis Ricordeau
Nantes
France
Phone: 33 633444793
Email: aurelie.najm@gmail.com

Abstract

Background: Despite the growing interest and exponential popularity of mobile health (mHealth) apps for long-term conditions such as rheumatic and musculoskeletal diseases (RMDs) and their self-management, patients are rarely directly consulted and involved in the app development process.

Objective: This study aims to explore the needs, experiences, and views of people diagnosed with RMDs on mHealth apps.

Methods: The study used a mixed methods approach: (1) an initial qualitative phase via a patient focus group in the UK and (2) a survey disseminated through national organizations for patients with RMDs across European countries, the United States, Canada, and Australia.

Results: The focus group included six patients with life-long musculoskeletal conditions. Half had used a self-management app at least once. The use of existing apps was reported as time-consuming due to a lack of functionality. The need for bespoke apps was voiced by all participants. Among 424 patients across European countries, the United States, Canada, and Australia, the main age group was 45 to 54 years (122/424, 28.7%), and 86.8% (368/424) were women. Half of the respondents were aware of the existence of apps to support self-management of their RMDs (188/355, 53%), with 42% (79/188) of them currently using such devices. Patients were mostly interested in an app to self-monitor their health parameters (259/346, 74.9%) and disease activity (221/346, 63.9%) or communicate directly with their health care provider (200/346, 57.8%).

Conclusions: Patients considered that using an app could help them to self-manage their RMD condition if it was tailored to their needs and co-developed with health professionals. The development of such apps will require standardization and regular quality control.

(JMIR Mhealth Uhealth 2020;8(4):e14351) doi:10.2196/14351

KEYWORDS

mHealth; mixed methods; mobile health apps; rheumatic and musculoskeletal disease; smartphone; apps; rheumatoid arthritis, digital health, mobile health

Introduction

The role of self-management in rheumatic and musculoskeletal diseases (RMDs) has become increasingly recognized, especially in support of holistic and patient-centered care [1]. A growing number of patients seek out various sources of information to better understand and manage their disease as part of their quest to take a more active role in their care [2,3]. Self-management refers to "the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic condition" [4]. In line with this, the past decade has witnessed an exponential growth of mobile phones and other electronic apps. A cumulative willingness of healthy individuals to improve their lifestyle has led to the demand for new digital lifestyle management solutions. The use of apps could support people taking personal responsibility for their well-being and contribute to their disease management in a more proactive way. Indeed, the number of health-related devices downloaded from various app stores has more than doubled from 1.7 billion in 2013 to 3.7 billion in 2017 [5].

Most apps on the market cater to healthy individuals, with a focus on physical activity and diet [6]. In parallel, other electronic devices tailored to persons living with long-term conditions have emerged. Their potential as disease management tools has been described in various conditions, such as asthma [7], hypertension [8], diabetes mellitus [9], and other life-long conditions [10]. Moreover, mHealth apps can support self-management in long-term conditions, such as RMDs, but can also help empower individuals to take a more active role in their health and be part of shared decision making [11-13] between the patient and health care staff.

The existence of apps may enable patients who live with long-term physical or mental conditions to improve their self-management. However, the evidence is lacking [14-16] on the development and evaluation of apps for their use in routine clinical care. To guide the development of further apps for patient self-management, it is essential to gain insight into the perceptions and experiences of patients living with long-term conditions, such as RMDs, who are current or potential users of available apps. Through a mixed-methods approach, this study explores the needs, experiences, and views of people with RMDs about their knowledge and use of mHealth apps. The aim was to obtain direct patient feedback on preferences for content, functionalities, structure, navigation, and relevance for their long-term condition and an evaluation of mHealth apps for self-management of their RMDs.

Methods

Overview

This was a mixed-methods study starting with one patient focus group to gather direct patient insights on the subject. An online, multinational survey followed, informed by the themes emerging from the focus group in the UK.

Patient Focus Group

One audio-recorded patient focus group in the UK was facilitated by an experienced qualitative researcher (HL) in March 2018; a cofacilitator was also present and took notes during the focus group. Patient research partners affiliated with the department of rheumatology at King's College London were invited by email. After they agreed to participate, an information sheet (developed by AN and EN) was sent to all to explain the purpose of the study.

Broad and open-ended questions were developed for the semistructured focus group guide along with specific questions about the use of mobile health apps for the management of their RMDs (see focus group schedule in Multimedia Appendix 1). The focus group lasted 75 minutes and took place in a private room in a medical school building. The analysis was performed through an inductive approach, carrying out three coding phases and identifying emerging key themes.

Online Open Survey

Key themes that emerged from the patient focus group informed the design of the next phase, namely the patient survey to obtain further detailed information on the subject of mHealth apps. The survey was created by a panel of four rheumatologists (AN, EN, LG, FB). A draft survey was shared with five patient research partners who live with RMDs from across five different countries (UK, Slovakia, Germany, the Netherlands, Cyprus) for direct feedback on the structure, relevance, and comprehension of the content, validation with pretest, and appropriate rephrasing when needed before its finalization and dissemination. The results are reported according to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist [17]. Consent was collected before entering the study. The final survey contained 40 questions, divided into five main sections: (1) sociodemographic information about the patients, (2) current mobile phone use, (3) current knowledge and use of apps for RMDs, (4) past use of apps for RMDs, and (5) development of the ideal app for self-management. Most questions were multiple choice, with a few open-ended ones.

Survey Dissemination

For dissemination of the survey, the sample frame was defined as people living with RMDs. Individual patient organizations from 45 EU countries, the United States, Canada, and Australia were approached.

For each EU country, dedicated People with Arthritis and Rheumatism (PARE) representatives linked to each European League Against Rheumatism (EULAR) country were contacted directly via email by the study team. Each country representative was responsible for contacting registered patient associations. Patient organizations then disseminated the survey through their websites and social media channels (Twitter). In the countries not linked to EULAR, patient associations or scientific societies were contacted to support the dissemination process.

The survey was translated in Slovakian, Portuguese, Spanish, and French as requested by national patient research partners in charge of dissemination. The survey dissemination was conducted in 2018. Informed consent was collected at the

beginning of the survey and was mandatory to start the survey; participants were informed of the expected duration of the survey, purpose, and data storage modalities. The survey contained 41 items on nine pages. It was made available for completion from mid-May to late June through Survey Monkey.

Data Analysis

The information provided by participants during the focus group was transcribed verbatim by AN. The data were thematically analyzed manually [18] by one researcher and one rheumatologist in the research team (AN, EN). A further cross-check of the emerging themes was carried out by HL and a patient involved in the focus group; single counting was applied (number of events, phenomena identified during the analysis) [19]. Final agreement among the research team (AN, EN, LG, FB) was reached for the key themes to be presented within an analytical framework. The survey results were collected and analyzed by AN through descriptive statistics using Excel (Microsoft Office) and Graphpad software [20].

Results

Patient Focus Group

Three of six focus group participants were female, with an age range between 32 and 69 years, and backgrounds of diverse disease duration and level of disability related to their disease (Multimedia Appendix 1). All attended the same tertiary outpatient clinic in London and lived in different parts of England. All had a diagnosis of a long-term autoimmune inflammatory RMD condition (five had rheumatoid arthritis [RA], one had myositis). Disease duration ranged from 3 to 47 years. All were medically retired, meaning the participants left their jobs before they had reached the official retirement age due to a health condition or sickness. They had diverse ethnic and socioeconomic backgrounds.

Summary of Key Emerging Themes

The analytical framework consisted of five key themes: (1) knowledge and previous use of mHealth apps in general, (2) experience of mHealth apps for self-management in rheumatology, (3) positive and negative features of the apps, (4) need for improvement in self-management apps, and (5) content and functionalities of ideal self-management apps for RMDs.

Knowledge and Previous Use of mHealth Apps in General

Three of six patients had heard of the term *eHealth* at least once and knew what the phrase referred to. eHealth was defined as "an emerging field in the intersection of medical informatics, public health, and business, referring to health services and information delivered or enhanced through the internet and related technologies" [21]:

> *I remember when I talked to my consultant, and he let me know that there was an app apparently that you could go into and basically put all your appointments.* [patient 1, female]

> *I would find this [app] useful if you want to know where to go in the hospital, and it has phone numbers from the departments.* [patient 2, male]

Most patients (4/6) were aware of the actual existence or the current development of apps for patients with RMDs: "My consultant told me about [the developments of apps]" (patient 2, male).

Two of six patients used general health apps (not designed specifically for patients living with RMDs), such as activity trackers or pedometers. Activity trackers and pedometers were defined as devices that use sensors to help users automatically track step counts while aiming for a particular step count or activity goal. Pedometers help promote self-awareness and self-monitoring of activity levels [22]. Such devices were perceived positively by the group: "It [addition of an app] makes me more conscious of the steps I do" (patient 4, female).

Half of the informants (3/6) reported that the pedometers could also be discouraging:

> *I had the feeling to do thousands of steps, and it [pedometer] was telling me not [participant had the impression of having done a lot of activity; however, the pedometer displayed a low number of counted steps]. It was actually demotivating, and I put it in the bin.* [patient 1, female]

Experience of mHealth Apps for Self-Management in Rheumatology

A third of patients had used mHealth apps for their rheumatic disease:

> *I have used it [app], I felt it was pretty time-consuming as you had to put so much in information) it [app], especially for the blood test.* [patient 6, male]

> *I am aware of a couple [of apps] through the NRAS [National Rheumatoid Arthritis Society, National Charity UK], the DAS [Disease Activity Score] app ...I also was involved in trying "Rheumopady," it is an app where you daily mark [input] your fatigue, sleep, pain, and stiffness, and it [app] marks a graph for you.* [patient 5, female]

Most (5/6) expressed enthusiasm about the incorporation of apps for self-management into their daily lives:

> *That is why I think it is app] very relevant, because I usually write these [symptoms] in a diary, how I feel.* [patient 2, male]

> *A graph would be great to show trends...a graph is a simple image you could show to the consultant.* [patient 4, female]

> *I think it [app] would be useful for people who are still working to see...because often they [clinicians, friends, families] don't realize how much time people have pain, even though they [patients] seem fine all day, they [patients] are probably not well-controlled.* [patient 5, female]

One patient voiced reluctance to use an app due to lack of knowledge about the function and therefore decided not to buy one:

> *"I was never going to put bloods [tests] into it [app] because, you know, the people who need to know got the info in the medical notes, I don't look at mine [blood tests] anyway".* [patient 1, female]

None of the participants were using bespoke RA self-management apps at the time of the interview.

Positive and Negative Features of the Apps

Positive and negative features were discussed by the participants, as well as the pros and cons of using self-management apps. Some participants (2/6) reported that RA self-management apps were helpful in principle, such as to remind them of their disease progress and their symptoms over time (eg, between clinic appointments). The majority (5/6) mentioned that apps could help them to remember and describe their symptoms more accurately to the clinicians as a "symptom diary":

> *I think it depends what [the information on the app] is going to be used for, because I think sometimes it would be useful, to put flares in [the app], because it seems so far, we are always fine on the day we go to clinics.* [patient 5, female]
>
> *I think it depends on the quantity of information asked [to be included in the app]...it was easy to type from 0-10 [for the VAS] for different health parameters.* [patient 4, female]

However, others (3/6) stated that self-management apps tend to take too much time to enter data every day and were rarely accessible or easy to use:

> *It's too much hassle, I mean, it was interesting, but I found it [app] was such a *** some days, it would save the data for you and some days it wouldn't. It was a trial [used the app during a trial], but if it was in real life I would not have carried on. If you put things [data] in that [app], fine, that's great, the principle is amazing. I think it is really interesting, you could see the graphs...but the fact is that it [app] did not work as well as it should have...The functionality is a big issue.* [patient 4, female]
>
> *It [the process] was so time-consuming...taking five minutes to go online.* [patient 3, male]
>
> *That kind of apps frustrate me generally; I don't find that they [apps] usually work as much as they should.* [patient 5, female]

Need for Improvement in Self-Management Apps

Participants (2/6) expressed a need for an improvement in the intuitive functionality of apps, especially regarding the time needed for data and information entry into the app.

An app allowing data collection and description of the symptoms would facilitate more relevant outpatient visits as patients tend not to remember how they have been feeling or experiencing symptoms over the past few weeks:

> *Why I think is very relevant, is because every time I go to see the consultant...and they ask you about not only how you be feeling over the last 7 days, which maybe you could remember...but then you maybe miss the kind of patterns, especially over months...and then that is not always indicative.* [patient 2, male]
>
> *It has to be easy too, for the older people, very often you hear this in clinics: "I can't do that."* [patient 1, female]

Content and Functionalities of Ideal Self-Management Apps for Rheumatic and Musculoskeletal Diseases

According to most focus group participants (4/6), the key domains to be addressed by an "ideal" app should include the following features: assessment and measurements of fatigue, sleep, pain, mood, activity, symptoms, Disease Activity Score (DAS 28), and appetite. All patients (6/6) suggested that such features in an app could help them in their self-management of their long-term conditions within the context of their daily activities (eg, planning social events, traveling) (Textbox 1). Moreover, they discussed features, functionalities, and potential links for inclusion into an "ideal app" as detailed in Textbox 1: "It would be quite nice to have a link with explanations on why you are taking this drug" (patient 6, male).

Textbox 1. Domains suggested by patients for how an app could be helpful in the self-management of their rheumatic and musculoskeletal diseases, including special features that result in an "ideal app."

Features

- The ability to analyze and draw correlations between symptoms (eg, fatigue, pain, activities), blood tests, medication, and disease activity
- The possibility of identifying specific "patterns" of symptom features that could predict flares
- Display graphs and trends of the inputted data/info
- The ability to quantify and share fatigue levels with the rheumatologist
- Information on medication adherence (reminders and other information about the importance of medication adherence)
- Blood test and outpatient clinic appointment reminders
- Records of blood monitoring results

Ease of use

- Availability on a tablet (easier to navigate with a bigger screen, easy use of keyboard with painful fingers due to RMD)
- Ease of use (intuitive functionality) especially in the case of "older people"

Disease monitoring (self-management)

- Therapeutic and management advice, subject to health parameters (eg, high disease activity score=advice to increase steroids or other treatment)
- Visual personalized health information through graphs and trends

Useful information and website links

- Information about medication side effects and interactions
- Links to websites to explain the purpose of each medication
- Links to national organization/helpline (national RMDs Society or Charity)
- Explanations and advice about the disease and some aspects of lifestyle changes (eg arthritis and fatigue, arthritis and sex)

Reminders (self-management)

- Option to set reminders to enter health parameters (eg, fatigue, sleep, pain, mood, activity, symptoms, DAS 28, blood tests)
- Notifications/alerts (eg, when new medication supply is needed)

Online Open Survey

General Demographics

Among the 429 respondents, 394 (91.8%) participants provided complete responses (all mandatory questions completed). The majority (86.9%, 342/394) of respondents were female, and almost half of the participants were 18 to 44 years of age. Sociodemographic information and diagnoses of the participants are detailed in Table 1.

Country representation varied, with the highest number of responses from Portugal (84/394, 21.3%), followed by Germany (50/394, 15%), Australia (33/394, 8.4%), United States (36/394, 9.1%), and Cyprus (25/394, 6.3%). Other participating countries are shown in Multimedia Appendix 2.

Table 1. Demographics and disease-related information of participants (N=429).

Demographic	n (%)
Age range (years)	
18-24	15 (3.8)
25-34	50 (12.7)
35-44	98 (24.9)
45-54	113 (28.7)
55-64	87 (22.1)
65 and older	31 (7.9)
Gender	
Female	342 (86.8)
Male	50 (12.7)
Transgender male	1 (0.3)
Disease[a]	
Rheumatoid arthritis	152 (38.8)
Fibromyalgia	80 (20.4)
Psoriatic arthritis	76 (19.4)
Ankylosing spondylitis or spondyloarthritis	67 (17.1)
Osteoarthritis	64 (16.3)
Sjögren syndrome	55 (14.0)
Spine or back disease	36 (9.2)
Systemic lupus erythematosus	31 (7.9)
Juvenile idiopathic arthritis	30 (7.7)
Osteoporosis	29 (7.4)
Vasculitis	11 (2.8)
Scleroderma or systemic sclerosis	14 (3.6)
Polymyositis/dermatomyositis	4 (1.0)
Other	65 (16.6)

[a]The participant percentage total is greater than 100% because multiple choices were possible.

Patients were diagnosed mostly with RA (152/394, 38.6%), fibromyalgia (80/394, 20.3%), psoriatic arthritis (76/394, 19.3%), and ankylosing spondylitis (67/394, 17%). Almost half of the patients (191/394, 48.5%) had their main diagnosis for more than 10 years.

Current and Past Use of Mobile Phone and mHealth Apps for Rheumatic and Musculoskeletal Diseases

Most patients used a mobile phone (355/394, 90.1%), mostly for text messages and calls (341/355, 95.1%) or internet access (328/355, 92.4%) and social media use (eg, Facebook, Twitter, Instagram, Snapchat; 289/355, 81.4%) (Table 2). A majority (341/358, 95.3%) reported searching for information on the internet about their disease. Half of the respondents were aware of the existence of apps to support them in self-managing RMDs (188/355, 52.3%). Among them, almost half of the respondents were currently using a self-management app (79/188, 42%), a third of them on a weekly basis (Table 2).

Table 2. Current and past use of mobile phone apps for rheumatic and musculoskeletal diseases (N=358).

Use of mobile phone apps[a]	n (%)
Purposes for the use of a mobile phone[b] (n=358)	
Text messages and calls	341 (95.3)
Internet access	328 (91.6)
Games	115 (32.1)
Social networking sites (eg, Facebook, Twitter, Snapchat, Instagram)	289 (80.78)
Other apps	196 (54.7)
Knowledge about mHealth apps for self-management of RMDs (n=355)	
Yes	188 (53.0)
No	167 (47.0)
Current use of an app for RMD self-management (n=188)	
Yes	79 (42.0)
No	109 (58.0)
Frequency of use of an app for RMD self-management (n=67)	
On a daily basis	20 (29.9)
On a weekly basis	21 (31.3)
On a monthly basis	8 (11.9)
Less than once monthly	3 (4.6)
Intermittently	15 (22.4)
Past use of an app (n=106)	
Yes	44 (41.5)
No	62 (58.5)
Duration of use (n=41)	
Less than 3 months	31 (75.6)
3-6 months	8 (19.5)
More than 6 months	2 (4.9)
Reason for discontinuing use[b] (n=37)	
I did not find it/them useful for my condition	24 (64.9)
I feel it did not benefit my overall health	16 (43.2)
I found it/them too time-consuming	15 (40.5)
I got bored	11 (29.7)
I did not like the design or user interface	10 (27.0)

[a]The n in parentheses represent the total number of participants who answered each question

[b]Total is greater than 100% because multiple choices were possible.

The reported app usage was mainly for disease management (42/67, 62.7%) and coping with arthritis-related symptoms and consequences (28/67, 41.8%) (Multimedia Appendix 2). Medication intake monitoring (24/67, 35.8%) was the third most common reported use.

The patients received recommendations about the apps by patient associations (29/67, 43.3%), advertisements (eg, print media, online social media; 24/67, 35.8%), and friends or relatives (13/67, 19.4%). Apps were recommended by physicians or health care professionals only in a few cases (10/67, 14.9%).

Almost half of the nonusers described a previous trial of a self-management app for less than three months. They mostly stopped using it because they did not find them helpful (24/37, 64.9%), saw no benefit for their health (16/37, 43.2%), found the device too time-consuming (15/37, 40%), got bored of using the app (11/37, 29.7%), or they did not like the design or user interface (10/37, 27%).

For respondents who declared no usage of apps for disease management, the reasons provided were the following: no benefit for their health (33/105, 31.4%), fear they would spend

too much time on filling the app (27/105, 25.7%), or concerns about health data protection (25/105, 23.8%) (Multimedia Appendix 3).

Features and Functionalities of Ideal Self-Management Apps for Rheumatic and Musculoskeletal Diseases

The majority of patients were interested in an app that helped with self-monitoring of health parameters (259/346, 74.9%), disease activity (221/346, 63.9%), communication with their health care providers (221/346, 57.8%), and information about their disease (200/346, 53.5%). Multimedia Appendix 4 details other reported features and functionalities of "ideal" self-management apps for RMDs. The collection of anonymized health data for research purposes seemed acceptable by most patients (n=176/320, 57.9%), as shown in Multimedia Appendix 5. The aspects to be addressed and information to be collected by the ideal app are described in Table 3.

Table 3. Features to be collected by the ideal App (N=345).

Features	n (%)
Pain [a]	*186 (84.2)*
Fatigue	*153 (69.2)*
Physical Activity	*149 (67.4)*
Sleep	118 (53.4)
Disease activity score	118 (53.4)
Well-being	119 (53.8)
Medication, adherence	111 (50.2)
Morning stiffness	112 (50.7)
Blood tests	112 (50.7)
Nutrition	106 (47.9)
Depression/Anxiety	97 (43.9)
Infections	95 (42.9)
Mood	86 (38.9)
Social support	70 (31.7)
Fears	61 (27.6)
Work	64 (28.9)
Other[b]	15 (6.8)

[a]Top responses are highlighted in italic.

[b]Other: Included Flares, Heart Rate, Blood Pressure, Temperature, a place to keep track of how the weather impacts pain level, and X-ray results.

Patients suggested they would like the information collected by the app to be shared with their rheumatologist (277/345, 80.1%), their general practitioner (240/345, 69.4%), and health professionals (eg, physiotherapist, nurse; 254/345, 73.4%). A majority (325/346, 93.9%) stated that they prefer to have control over which health care professionals have access to their personal data through the app because this could make clinic visits more efficient (322/346, 93.1%). Interesting concepts emerged related to either self-management strategies (eg, pain, fatigue, physical activity) or other aspects such as work and social support.

The majority of respondents (197/338, 58.3%) expressed an unwillingness to pay for an app. From those prepared to pay, most (85/141, 60.3%) expressed their preference for a one-off payment (Multimedia Appendix 6).

Patients expressed highest confidence in an app developed by an RMD scientific society (very confident: 207/331, 62.5%). Poorest confidence was expressed for an app developed by a pharmaceutical company (152/331, 45.9%).

Regarding the operating mode of the app, the majority (280/320, 87.5%) were in favor of the app to send reminders, with half of them in favor of receiving reminders for medication intake (149/280, 53.2%) and health appointments (162/280, 57.8%). Most respondents (295/320, 92.2%) were keen to include a display of visual health parameters with images (95/320, 29.7%), tables (71/320, 22.2%), or graphics (129/320, 40.3%).

A majority (306/320, 95.6%) of the respondents wanted to be given the opportunity to rate the app: 18.4% (59/320) wanted to do so by voting, 50.9% (163/320) by using a satisfaction scale, and 22.8% (73/320) by providing free-text comments.

Discussion

This study assessed patients' perceptions of and experiences with mHealth apps for self-management in rheumatology. It also explored people's views on what they perceive as an ideal app. Our study highlights the willingness of people with RMDs to use such apps to improve their disease management. The latter could potentially be through patient empowerment and

encouraging patients to take a more active role in decisions relating to their health. Our study also highlights the need for improvement in app functionalities and content, a desire that was consistently expressed across both the patient focus group and the survey.

Despite the recent rapid growth in the development and dissemination of apps for disease self-management, our focus group and survey results demonstrated that half the respondents did not know about the existence of mHealth apps in rheumatology. This might be explained in part by the wide range of ages represented by our survey participants, with 58.7% older than 45 years (231/394). These results are in line with another recent survey on cancer apps, in which the age of the patient (age range 18-39 years) was significantly correlated with the use of the app [23], although the correlation was weak. However, a recent survey performed with type 1 diabetes patients, who represent a younger patient population by definition, showed that 61% of the patients reported no awareness or knowledge of the existence of self-management apps [24]. These observations may indicate that other reasons may be at play that are not related to age, such as lack of health care provider awareness or willingness to encourage app use or directly related to the underlying condition (eg, joint swelling and deformities in rheumatic diseases preventing easy use of apps). The focus group revealed the wish of patients for apps to be tailored for use by "older patients." From a financial point of view, patients expressed their preferences for a free app.

Interestingly, patients who had used available mHealth apps for RMD self-management had an overall negative opinion, and most stopped using them for various reasons (eg, too time-consuming to log in or to complete their health parameters and with no direct positive impact on their health). For example, patients mentioned they were more likely to use an app when they were feeling well and when not undergoing a flare. Moreover, poor functionality of certain apps rendered them unreliable, which caused frustration. These findings are in line with a similar survey conducted in other medical specialties (eg, in type 1 diabetes [24]). Patients did not think the app benefited their health because data, especially information on glycemia across the day, could not be shared with their doctor. The lack of immediate feedback on the blood results and without therapeutic action was perceived as unsatisfactory. Indeed, the creation of apps designed for people living with RMDs represents a relatively recent change [25]. Previous work that assessed the quality of existing apps in Google Play, iTunes, or Android app stores showed that most of the available ones did not meet high quality standards [26-28]. Moreover, based on a previous systematic literature review from our group [29], the majority of apps and devices for patients living with RMDs were not developed with any input from patients. This is a major drawback. This oversight of the direct inclusion of relevant stakeholders might lead to irrelevant and inappropriate development processes.

Data sharing with health care providers does raise ethical issues [30,31]. Based on our survey, health data protection was also a reason that was frequently stated for not using mHealth apps (25/105, 23.8%). A majority of patients from both our multinational survey and focus group expressed a wish to be able to choose if and which health care providers could access their health data. Previous work has proposed or summarized existing regulations on the safety and security of mobile apps [32-35]. Moreover, apps and any other e-device should follow the EU General Data Protection Regulation (GDPR) and US Health Insurance Portability and Accountability Act (HIPAA). However, to our knowledge, no specific regulation on which health care providers should access the data gathered by the apps exists yet.

Our study specifically raised questions about the content and functionalities of apps that people with RMDs wished to see in an ideal app. A broad range of features was described, such as pain, fatigue, physical activity, sleep, and morning stiffness. However, many of these details were not yet included or only partially included in existing apps currently available on the market [30,31]. We acknowledge the challenges of developing an app that is inclusive of all functions desired by study participants also from a technical point of view. Furthermore, we acknowledge that people with RMDs can have significant limitations in their function, including fine movement of the hands and fingers, for example, which poses practical challenges in the use of apps. The app needs to be developed to support and encourage a proactive role of patients in the management of their health. The elaboration of a single app might not necessarily be able to provide all described features. A good way to make the app straightforward and practical is to involve patients in the development process.

Therefore, our results are important to provide further guidance on how best to tailor newly developed apps to patient needs in the field of RMDs.

Our study provides an in-depth approach to studying views, needs, preferences, and previous experiences of mHealth apps in patients living with RMDs. Using a mixed-methods approach enabled the research team to identify key aspects of mHealth app features and functionalities that have the potential to positively influence future development of apps. Moreover, many of the emerging themes were in line with the five core self-management skills (pain education, self-efficacy building, self-monitoring, social support, and goal setting) [36].

Our online, multinational survey, translated into five different languages, made it possible to reach out to the wider patient community with RMDs and to obtain first-hand patient feedback. The survey was disseminated and completed by participants with a broad range of age categories, diseases, and countries across three different continents. Limitations of our study include the single focus group performed in one country. In addition, because of the qualitative arm of the study results and that the online survey was open to everyone, the results might not necessarily be generalizable. Furthermore, most of the participants in the focus group had a diagnosis of RA, which may affect the generalizability of the results to other RMDs.

However, our survey provided greater granularity of information on specific themes that emerged from the focus group, as well as more generic themes confirmed in the systematic review of the literature and expert opinion. The higher number of female patients participating in the survey is in line with existing epidemiological data, which supports that the gender ratio is in

favor of women in highly prevalent RMDs, such as fibromyalgia and RA [37,38].

In conclusion, despite the reluctance of some people with RMDs to incorporate apps for reasons such as relevance of content, reliable functionality, and data safety concerns, there was acknowledgment that apps may be suitable for patient self-management. If apps are bespoke and developed in close collaboration with the key stakeholders (ie, the patients and carers), this inclusive approach may have potential benefits for patient health. The development of such apps will require standardization and quality control processes to be in place [16]. There is an unmet need to develop helpful and tailored mHealth apps. It is essential that patients can be sure of the accuracy and safety of their personal data, including sociodemographic data. National RMD societies and patient associations have an important role to play in the development, validation, and dissemination of mHealth apps among patients and health care providers.

Acknowledgments

We wish to thank the patient research partners who helped in the survey development, all the PARE liaisons who helped in the survey dissemination, all the patient research partners that contributed to the survey dissemination outside the EULAR countries, all the patients who participated in the focus group and survey, and the native speakers who checked and modified the survey (Javier Carrio, Goncalo Boleto, Petra Balážová). The project was funded by EULAR (project number CLI102).

Authors' Contributions

AN designed the survey, analyzed the qualitative and quantitative data, and wrote the first draft of the manuscript. HL carried out the focus group, helped with the conduct of the study, assisted with the qualitative data analysis, and edited the manuscript. LG and FB designed the survey, participated in the quantitative data analysis, and edited the manuscript. EN designed the survey, helped with the execution of the qualitative and quantitative studies and data analysis, and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient characteristics.
[DOCX File , 13 KB - mhealth_v8i4e14351_app1.docx]

Multimedia Appendix 2
Demographic data: country of origin of the survey participants.
[DOC File , 13 KB - mhealth_v8i4e14351_app2.doc]

Multimedia Appendix 3
Reasons for app use reported by participants.
[DOC File , 13 KB - mhealth_v8i4e14351_app3.doc]

Multimedia Appendix 4
Reasons for not using apps for RMDs.
[DOC File , 13 KB - mhealth_v8i4e14351_app4.doc]

Multimedia Appendix 5
Purposes of the ideal app.
[DOCX File , 13 KB - mhealth_v8i4e14351_app5.docx]

Multimedia Appendix 6
Patients willingness to pay for the app.
[DOCX File , 13 KB - mhealth_v8i4e14351_app6.docx]

References

1. Manning VL, Hurley MV, Scott DL, Coker B, Choy E, Bearne LM. Education, self-management, and upper extremity exercise training in people with rheumatoid arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2014 Feb;66(2):217-227 [FREE Full text] [doi: 10.1002/acr.22102] [Medline: 23925924]

2. Ellis J, Mullan J, Worsley A, Pai N. The role of health literacy and social networks in arthritis patients' health information-seeking behavior: a qualitative study. Int J Family Med 2012;2012:397039 [FREE Full text] [doi: 10.1155/2012/397039] [Medline: 22997575]
3. des Bordes JK, Gonzalez E, Lopez-Olivo MA, Shethia M, Nayak P, Suarez-Almazor ME. Assessing information needs and use of online resources for disease self-management in patients with rheumatoid arthritis: a qualitative study. Clin Rheumatol 2018 Jul;37(7):1791-1797. [doi: 10.1007/s10067-018-4047-9] [Medline: 29500557]
4. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]
5. Statistica. 2017 Nov 08. Number of mHealth app downloads worldwide from 2013 to 2017 URL: https://www.statista.com/statistics/625034/mobile-health-app-downloads/
6. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017 Jan;33:343-347. [doi: 10.1016/j.nut.2016.08.003] [Medline: 27742102]
7. Cook KA, Modena BD, Simon RA. Improvement in asthma control using a minimally burdensome and proactive smartphone application. J Allergy Clin Immunol Pract 2016;4(4):730-737.e1 [FREE Full text] [doi: 10.1016/j.jaip.2016.03.005] [Medline: 27107690]
8. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015 Feb;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
9. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM Self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
10. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
11. Greene JA. Do-it-yourself medical devices--technology and empowerment in American health care. N Engl J Med 2016 Jan 28;374(4):305-308. [doi: 10.1056/NEJMp1511363] [Medline: 26816009]
12. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
13. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
14. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2014 Dec 5;38(1):23-38. [doi: 10.1007/s13246-014-0315-4]
15. Buijink AW, 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]
16. 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 Mar 17:i1174. [doi: 10.1136/bmj.i1174]
17. 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. [doi: 10.2196/jmir.6.3.e34]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Seale C. Introducing Qualitative Methods: The Quality of Qualitative Research. London: Sage Publications Ltd; Jul 1999.
20. Swift ML. GraphPad prism, data analysis, and scientific graphing. J Chem Inf Comput Sci 1997 Mar 24;37(2):411-412. [doi: 10.1021/ci960402j]
21. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):e20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
22. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
23. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile health in oncology: a patient survey about app-assisted cancer care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
24. Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z. Exploration of users' perspectives and needs and design of a type 1 diabetes management mobile app: mixed-methods study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11400 [FREE Full text] [doi: 10.2196/11400] [Medline: 30249580]
25. de Souza S, Galloway J, Simpson C, Chura R, Dobson J, Gullick NJ, et al. Patient involvement in rheumatology outpatient service design and delivery: a case study. Health Expect 2017 Jun;20(3):508-518 [FREE Full text] [doi: 10.1111/hex.12478] [Medline: 27345769]
26. Grainger R, Townsley H, White B, Langlotz T, Taylor W. 2016 ACR/ARHP annual meeting abstract supplement. Arthritis Rheumatol 2016 Oct;68 Suppl 10:1-4550 [FREE Full text] [doi: 10.1002/art.39977] [Medline: 27870508]

27. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for people with rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
28. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019 Apr;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
29. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile Health Apps for Self-Management of Rheumatic and Musculoskeletal Diseases: Systematic Literature Review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730-e14227 [FREE Full text] [doi: 10.2196/14730] [Medline: 31769758]
30. Hussain M, Al-Haiqi A, Zaidan AA, Zaidan BB, Kiah ML, Anuar NB, et al. The landscape of research on smartphone medical apps: coherent taxonomy, motivations, open challenges and recommendations. Comput Methods Programs Biomed 2015 Dec;122(3):393-408. [doi: 10.1016/j.cmpb.2015.08.015] [Medline: 26412009]
31. Capon H, Hall W, Fry C, Carter A. Realising the technological promise of smartphones in addiction research and treatment: an ethical review. Int J Drug Policy 2016 Oct;36:47-57. [doi: 10.1016/j.drugpo.2016.05.013] [Medline: 27455467]
32. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
33. Armstrong S. What happens to data gathered by health and wellness apps? BMJ 2016 Jun 23;353:i3406. [doi: 10.1136/bmj.i3406] [Medline: 27338775]
34. Hordern V. Data protection compliance in the age of digital health. Eur J Health Law 2016 Jun;23(3):248-264. [doi: 10.1163/15718093-12341393] [Medline: 27491248]
35. Sharp M, O'Sullivan D. Mobile medical apps and mHealth devices: a framework to build medical apps and mHealth devices in an ethical manner to promote safer use - a literature review. Stud Health Technol Inform 2017;235:363-367. [Medline: 28423815]
36. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
37. Kvien T, Uhlig T, Ødegård S, Heiberg MS. Epidemiological aspects of rheumatoid arthritis: the sex ratio. Ann NY Acad Sci 2006 Jun;1069:212-222. [doi: 10.1196/annals.1351.019] [Medline: 16855148]
38. Jones GT, Atzeni F, Beasley M, Flüß E, Sarzi-Puttini P, Macfarlane GJ. The prevalence of fibromyalgia in the general population: a comparison of the American College of Rheumatology 1990, 2010, and modified 2010 classification criteria. Arthritis Rheumatol 2015 Feb;67(2):568-575. [doi: 10.1002/art.38905] [Medline: 25323744]

Abbreviations

DAS: Disease Activity Score
EULAR: European League Against Rheumatism
RA: rheumatoid arthritis
RMD: rheumatic and musculoskeletal diseases

Edited by C Dias; submitted 15.04.19; peer-reviewed by C Jacquemin, M Abdulnabi, R Grainger, J des Bordes; comments to author 24.05.19; revised version received 12.07.19; accepted 23.08.19; published 20.04.20.

Please cite as:
Najm A, Lempp H, Gossec L, Berenbaum F, Nikiphorou E
Needs, Experiences, and Views of People With Rheumatic and Musculoskeletal Diseases on Self-Management Mobile Health Apps: Mixed Methods Study
JMIR Mhealth Uhealth 2020;8(4):e14351
URL: https://mhealth.jmir.org/2020/4/e14351
doi:10.2196/14351
PMID:32310151

Review

Methods and Measures Used to Evaluate Patient-Operated Mobile Health Interventions: Scoping Literature Review

Meghan Bradway[1,2*], MBA; Elia Gabarron[1*], PhD; Monika Johansen[1,3], PhD; Paolo Zanaboni[1,3], PhD; Patricia Jardim[4], MA; Ragnar Joakimsen[5,6], MD, PhD; Louise Pape-Haugaard[7], PhD; Eirik Årsand[1,2], PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[2]Department of Clinical Medicine, Faculty of Health Science, University of Tromsø The Arctic University of Norway, Tromsø, Norway

[3]Telemedicine and eHealth Research Group, Department of Clinical Medicine, University of Tromsø The Arctic University of Norway, Tromsø, Norway

[4]Norwegian Institute of Public Health, Oslo, Norway

[5]Tromsø Endocrine Research Group, Department of Clinical Medicine, University of Tromsø The Arctic University of Norway, Tromsø, Norway

[6]Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway

[7]Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[*]these authors contributed equally

Corresponding Author:
Meghan Bradway, MBA
Norwegian Centre for E-health Research
University Hospital of North Norway
PO Box 35
Tromsø, 9038
Norway
Phone: 47 91193393
Email: mbradway90@gmail.com

Abstract

Background: Despite the prevalence of mobile health (mHealth) technologies and observations of their impacts on patients' health, there is still no consensus on how best to evaluate these tools for patient self-management of chronic conditions. Researchers currently do not have guidelines on which qualitative or quantitative factors to measure or how to gather these reliable data.

Objective: This study aimed to document the methods and both qualitative and quantitative measures used to assess mHealth apps and systems intended for use by patients for the self-management of chronic noncommunicable diseases.

Methods: A scoping review was performed, and PubMed, MEDLINE, Google Scholar, and ProQuest Research Library were searched for literature published in English between January 1, 2015, and January 18, 2019. Search terms included combinations of the description of the intention of the intervention (eg, self-efficacy and self-management) and description of the intervention platform (eg, mobile app and sensor). Article selection was based on whether the intervention described a patient with a chronic noncommunicable disease as the primary user of a tool or system that would always be available for self-management. The extracted data included study design, health conditions, participants, intervention type (app or system), methods used, and measured qualitative and quantitative data.

Results: A total of 31 studies met the eligibility criteria. Studies were classified as either those that evaluated mHealth apps (ie, single devices; n=15) or mHealth systems (ie, more than one tool; n=17), and one study evaluated both apps and systems. App interventions mainly targeted mental health conditions (including Post-Traumatic Stress Disorder), followed by diabetes and cardiovascular and heart diseases; among the 17 studies that described mHealth systems, most involved patients diagnosed with cardiovascular and heart disease, followed by diabetes, respiratory disease, mental health conditions, cancer, and multiple illnesses. The most common evaluation method was collection of usage logs (n=21), followed by standardized questionnaires (n=18) and ad-hoc questionnaires (n=13). The most common measure was app interaction (n=19), followed by usability/feasibility (n=17) and patient-reported health data via the app (n=15).

Conclusions: This review demonstrates that health intervention studies are taking advantage of the additional resources that mHealth technologies provide. As mHealth technologies become more prevalent, the call for evidence includes the impacts on patients' self-efficacy and engagement, in addition to traditional measures. However, considering the unstructured data forms, diverse use, and various platforms of mHealth, it can be challenging to select the right methods and measures to evaluate mHealth

technologies. The inclusion of app usage logs, patient-involved methods, and other approaches to determine the impact of mHealth is an important step forward in health intervention research. We hope that this overview will become a catalogue of the possible ways in which mHealth has been and can be integrated into research practice.

(JMIR Mhealth Uhealth 2020;8(4):e16814) doi:10.2196/16814

KEYWORDS

mobile health; apps; self-management; chronic disease; noncommunicable diseases; interventions; patient-centered approach; patient-operated intervention

Introduction

Need for Mobile Health Evaluation

Health research is yet to agree upon a framework for evaluating mobile health (mHealth) interventions. This is especially true for tools, such as apps and wearables, that are intended primarily to aid patients in health self-management. Traditionally, the evaluation of mobile medical devices has been based on clinical evidence, and it can take years to bring these devices to the market. The continuous glucose monitor first came onto the market in 1999, but it was not until 2006 that the next version was available [1]. Similarly, the pulse oximeter struggled for decades to become a standard mobile tool for measuring blood oxygenation [2]. Because there are increasingly easy-to-use patient-operated mHealth technologies available on the market, patients are no longer willing to wait for a lengthy evaluation process. Instead, patients often use apps without assurance of quality or guidance from their health care providers [3].

Always-Available Self-Management Technologies

Individuals are more empowered to take greater responsibility for their health, and currently, they enthusiastically seek out mHealth apps and other devices for self-management. For chronic conditions in particular, health challenges occur continuously, not just when it is convenient or at a doctor's office. Technologies for self-management must allow individuals to register and review the measurements that they input into the app or system at any time. Connectivity to devices, such as medical or commercial sensors and wearables, adds to the utility of an app. A report by Research2Guidance [4], an organization that provides market research on digital health, emphasized the central role of patient-operated mHealth apps in the "connectivity landscape" of electronic health technologies [5]. However, their diverse functionalities and intended uses pose great challenges to researchers.

Challenges of mHealth Evaluation: Single Apps Versus Multiplatform Interventions

The amount of assessment and testing that is necessary for health technology is directly related to its potential risks and benefits [6,7]. For example, medications based on patient-gathered health data are associated with higher health risks than those in patients with type 2 diabetes who seek motivation from an activity tracker for weight management. Although multiplatform (ie, system) interventions serve to increase the benefits (eg, automatic and less burdensome operations), they increase the risks related to data safety, integrity, and reliability [8,9]. Researchers must adapt their approaches, methods, and measures for patient self-management interventions involving single mHealth apps and those involving multiplatform systems.

Evaluation Framework: Coverage

There are two main categories of mobile medical or mHealth devices associated with the amount of oversight health authorities will show; those that are "actively regulated" and those that fall under "enforcement discretion." These categories are described in the 2015 Guidance for Industry and Food and Drug Administration Staff [10] and are echoed in the updated 2019 Guidance [11] and included in the terms of The European Economic Area Certification (CE) Mark [12]. Devices that are actively regulated are required to undergo an evaluation and meet security and effectiveness standards for use in health care. On the other hand, many patient-operated technologies fall under "enforcement discretion," and they pose less risk to patient safety and health. For individuals aiming to assess the usefulness or safety of these technologies, there are no evaluation frameworks or guidelines to follow. The year 2015 marked a relevant change in the mHealth arena, which we are still exploring today (connectivity between different device types, development on different platforms, and marked focus on mHealth integration into clinical practice) [13].

Although there have been many strategies [14-17] for the evaluation of this subset of mHealth (eg, National Institute for Health and Care Excellence [18]), there is no agreement about which qualitative or quantitative measures should be addressed or how they should be evaluated [19]. Evaluation frameworks, such as the World Health Organization (WHO) mHealth evidence reporting and assessment (mERA) checklist [20], suggest that traditional health research measures and methods are not sufficient. For assessing the comprehensive impacts of such patient-operated mHealth approaches, research needs to look into additional factors. This can be achieved by producing evidence that is relevant for both patients and clinicians.

Additional Factors for mHealth Evaluation

Although clinical evidence is essential for the evaluation of any health aid, the two major concepts of time and human behavior must also be addressed in mHealth evaluation. As "always available" technologies are being used continuously and uniquely by patients, it is uncertain how much time is needed to produce an effect and what changes in self-management behavior will occur. Traditionally, medical devices rely on established biological knowledge, have fewer alternatives in the market, and do not offer frequent updates. However, patient-operated mHealth approaches require the consideration of patients' motivation, health beliefs, and resources for self-management. They must also compete with hundreds of

other mHealth apps and devices that are continuously developed and updated. In recent years, clinical research has attempted to keep pace with mHealth by employing methods that aim to expedite the research process and produce more tailored knowledge for the field of mHealth [21].

Stakeholders associated with chronic health and care (researchers, individuals, health care providers, and health care authorities) have been calling for evidence related to the personal use of mHealth technologies for many years [22-24]. Regardless of the beneficial or harmful outcomes, we need to know their potential. Without such evidence, people in the health care field will not be able to effectively support and guide individuals in the use of these technologies for health self-management. This evidence must be obtained with appropriate questions and methods.

Recent scoping reviews of mHealth technologies for chronic conditions focused on evidence as it relates to a specific age group [25], the development process [26], or clinical outcomes [27] and not on how the research was performed or which resources were used in the evaluation. The purpose of this scoping review was to identify which methods were used and which qualitative and quantitative data were measured to assess patient-operated mHealth devices for the self-management of chronic noncommunicable diseases (NCDs). As evidence for health authorities and health care providers, quantitative clinical outcomes have historically been considered the primary target for evaluation [28]; however, given the growing trend of mHealth, we included qualitative measures of participants' use of and experiences with the technology.

Research Questions

The research questions were as follows: (1) What methods are used to evaluate patient-operated mHealth apps and systems for self-management of chronic NCDs? (2) Which qualitative and quantitative measures are used to evaluate the impact of patient-operated mHealth apps and systems for self-management of chronic NCDs?

Methods

Scoping Review Objective

We performed a scoping review to document how researchers have evaluated mHealth interventions for self-management of chronic NCDs. Munn et al [29] stated that scoping reviews are favored over other review types in cases in which researchers are using an evolving set of methods owing to the novelty of the field or where the purpose of the review is to inform future questions about the field. We intended to provide an overview of what methods researchers use and which qualitative and quantitative measures were adopted to evaluate mHealth self-management interventions. This review reports information according to the Preferred Reporting Items for Systematic review and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist (Multimedia Appendix 1).

Search Strategy and Databases

The scope of the search and definitions of mHealth were discussed among the coauthors (MB, EG, EÅ, and MJ). The databases searched for scientific literature were PubMed, MEDLINE, Google Scholar, and ProQuest Research Library. PubMed and MEDLINE were both included because PubMed includes citations that are not yet indexed in MEDLINE [30]. We searched for articles published in English between January 1, 2015, and January 18, 2019, which were related to the evaluation of patient-operated mHealth interventions for self-management of chronic NCDs. The search string included key terms describing the intervention's intended use (ie, self-efficacy, self-assessment, self-management, or self-monitoring) and the intervention's platform (ie, mobile phones, wearables, sensors, or apps). The full search string was used for titles and abstracts, and the format was adapted to the database being searched (Multimedia Appendix 2).

Medical Subject Headings (MeSH) terms were not considered because our search included articles published recently, which may contain terminology that has not yet been indexed within the MeSH database. The identified abstracts and titles were collected in EndNote [31] and then uploaded into Rayyan [32], an online "library systematic review service" that allows researchers to collaborate on the organization, inclusion, and exclusion of articles for literature review.

Eligibility Criteria

We aimed to include research efforts that may have addressed new guidelines for mobile medical devices. Within our broad search criteria for low-risk mHealth apps and systems, articles were eligible for inclusion if they described low-risk technologies consistent with the FDA and CE Markings' description of mobile medical devices under "enforcement discretion" [10-12]. Multimedia Appendix 3 describes the specificities of this subcategory.

A preliminary search was performed, and a random selection of 10 articles was reviewed for inclusion or exclusion by two authors (MB and EG). Refinements were made to the review criteria.

For this review, we included studies that evaluated interventions involving (1) mHealth technologies for chronic NCDs, including the primary NCDs listed by the WHO [33] (ie, diabetes, cancer, cardiovascular diseases, chronic respiratory diseases, and chronic mental health conditions); (2) mHealth technologies for self-management (tasks which a person must perform in order to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic condition, and efficacious self-management was considered to encompass the ability to monitor one's condition and to affect the cognitive, behavioral, and emotional responses necessary to maintain a satisfactory quality of life) [34]; and (3) mHealth technologies that allow the patient to choose which measures to register and review.

The details of the inclusion and exclusion criteria are described in Multimedia Appendix 4, and they were used during the main review search.

Data Extraction and Synthesis

After removing duplicate articles, reviews, and protocol articles without evaluation results, two authors (MB and PJ)

independently screened the titles and abstracts for eligibility according to the inclusion and exclusion criteria. In case of disagreement regarding eligibility, another author (EG) was called to join the discussion until an agreement was reached. Author MB reviewed the full-text articles and performed data extraction.

The identified studies were classified as either those that evaluated mHealth apps or mHealth systems. Interventions that included a single app were grouped as mHealth apps, whereas those that included services or devices connected to a central app were grouped as mHealth systems. In this way, we could more clearly assess the different approaches taken by researchers when addressing the various impacts of these two mHealth intervention types.

Abilities of Studies to Produce Results

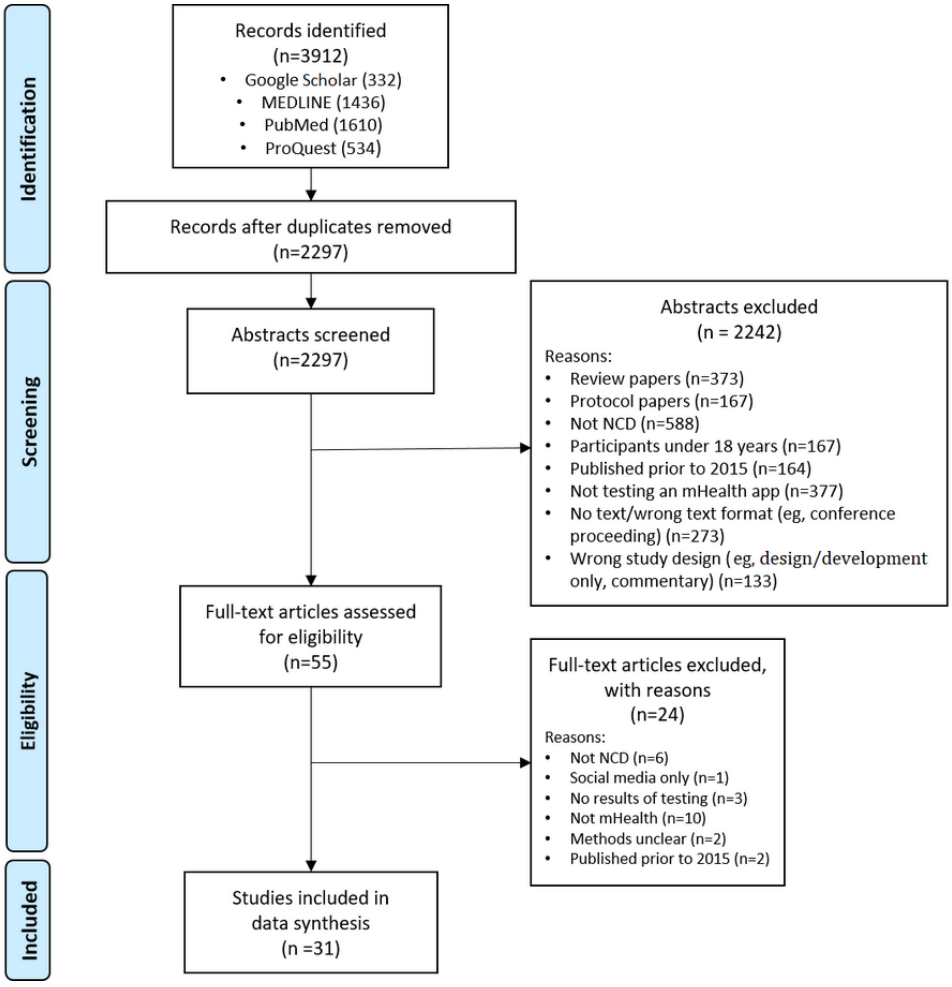
For both groups, one author (MB) assessed whether a study was able to produce the evidence that it aimed to obtain, using the selected methods. This was performed by comparing the objectives as stated by the authors of the identified articles to the methods and reported results. The studies were judged according to their ability to produce the information, and the findings were reported as yes, yes and more than expected, no, and cannot tell. The results of these comparisons are detailed in Multimedia Appendix 5.

Results

Overview

Among 3912 records identified by the search criteria, we reviewed 55 full-text articles and included 31 studies for data extraction and synthesis. Figure 1 illustrates the process of identifying the relevant articles for inclusion in data extraction.

Figure 1. Flow diagram illustrating the selection of studies for inclusion in data synthesis. NCD: noncommunicable disease.



Summary of Studies: Apps Versus Systems

Among the 31 studies chosen for data extraction, 15 were categorized as those that evaluated mHealth apps and 17 were categorized as those that evaluated mHealth systems. One study evaluated both apps and systems [35] and was therefore included in both categories. General information about the selected studies that evaluated mHealth apps are summarized in Table 1 [35-49] and those that evaluated mHealth systems are summarized in Table 2 [35,50-65].

Table 1. Information about the studies that evaluated mHealth apps.

Reference	App name	Year	Country	Study design	Duration	Health condition	Patient participants	Health care provider and caregiver participants	Intended secondary users
[36]	Diet and Activity Tracker (iDAT)	2015	Singapore	Prospective study	8 weeks	Type 2 diabetes	Patients (n=84)	N/A[a]	N/A
[37]	Diabetes Notepad	2015	Korea	Cross-sectional study	Single evaluation	Diabetes	Patients (n=90)	N/A	N/A
[38]	Personal Life-chart app	2015	Germany	Prospective study	72 weeks	Bipolar disorder	Patients (n=54)	N/A	N/A
[39]	HeartKeeper	2015	USA	Cross-sectional study	Single evaluation	Heart diseases	Patients (n=24) and researchers	N/A	N/A
[40]	HeartKeeper	2016	Spain	Retrospective study	36 weeks	Heart diseases	Patients (n=32)	N/A	N/A
[41]	PTSD Coach	2015	USA	Retrospective study	Duration of availability of the app on app stores	Post-traumatic stress disorder	Current users (n=156)	N/A	N/A
[42]	PTSD Coach	2015	USA	RCT[b]	16 weeks	Post-traumatic stress disorder	Patients (n=10)	Health care providers (n=3)	Health care providers
[43]	PTSD Coach	2016	USA	RCT	4 weeks	Post-traumatic stress disorder	Patients (n=49)	N/A	N/A
[44]	PTSD Coach	2017	USA	RCT	24 weeks	Post-traumatic stress disorder	Patients (n=120)	N/A	N/A
[45]	Hypertension management app (HMA)	2016	Korea	—[c]	Single event evaluation	Hypertension	Patients (n=38)	Nurses (n=3) and experts (n=5)	N/A
[35][d]	Multiple commercial apps for heart failure	2016	USA	Cross-sectional study	Single evaluation	Heart failure	Apps (n=34)	N/A	Family, friends, and health care providers (not all apps)
[46]	Multiple commercial apps (n=11)	2016	USA	Cross-sectional study	Single evaluation	Multiple	Patients (n=20)	Caregivers (n=9)	N/A
[47]	I-IMR intervention	2017	USA	Cross-sectional study	Single evaluation	Serious mental health conditions[e]	Patients (n=10)	N/A	N/A
[48]	Serenita	2017	Israel	Prospective study	16 weeks	Type 2 diabetes	Patients (n=7)	Health care providers	N/A
[49]	Sinasprite database	2018	USA	Retrospective study	6 weeks	Depression and anxiety	Patients (n=34)	N/A	N/A

[a]N/A: not applicable.

[b]RCT: randomized controlled trial.

[c]Not available.

[d]Study evaluated both apps and systems and therefore will appear in both categories.

[e]Combination of cardiovascular disease, obesity, diabetes, high blood pressure, high cholesterol, osteoporosis, gastroesophageal reflux disease, osteoarthritis, chronic obstructive pulmonary disease, congestive heart failure, coronary artery disease, and bipolar disorder, major depressive disorder, schizophrenia, or schizoaffective disorder [47].

Table 2. Information about the studies that evaluated mHealth systems.

Reference	Intervention name	Year	Country	Study design	Duration	Health condition	Participants	Intended secondary users	Others involved in the intervention	Medical device included (Y/N)	Other devices included
[50]	SUPPORT-HF Study	2015	UK	Cross-sectional study	45 weeks	Heart failure	Patients (n=26)	Health care providers	Health care providers and informal care givers	Y	Blood pressure monitor, weight scales, and pulse oximeter
[51]	__[a]	2015	USA	Cross-sectional study	Single evaluation	Diabetes	Patients (n=87) and health care providers (n=5)	Health care providers	Health care providers	Y	Glucose meter
[52]	Multiple commercial technologies for activity tracking	2015	USA	Prospective study	80-100 days (mean 12.5 weeks)	Serious mental health condition[b]	Patients (n=10)	Health care providers and peers (optional)	N/A[c]	N	Wearable activity monitoring devices
[53]	Diabetes Diary app	2015	Norway	Prospective study	2 weeks	Type 1 diabetes	Patients (n=6)	N/A	N/A	Y	Smartwatch app and glucose meter
[54]	Diabetes Diary app	2015	Norway	RCT[d]	23 weeks	Type 1 diabetes	Patients (n=30)	N/A	N/A	Y	Glucose meter
[55]	Diabetes Diary app	2016	Norway	RCT	48 weeks	Type 2 diabetes	Patients (n=151)	Health care providers	N/A	Y	Glucose meter
[56]	SnuCare	2016	Korea	Prospective study	8 weeks	Asthma	Patients (n=44)	N/A	Research team	Y	Peak flow meter
[57]	HealthyCircles Platform	2016	USA	RCT	24 weeks	Hypertension	Patients (n=52)	Health care providers	Health care providers	Y	Withings blood pressure monitor
[58]	Multiple commercial technologies for activity tracking	2016	USA	Prospective study	24 weeks	Serious mental health condition[b]	Patients (n=11)	N/A	N/A	N	Fitbit Zip
[35][e]	Multiple commercial apps for heart failure	2016	USA	Cross-sectional study	Single evaluation	Stroke	Apps (n=34)	Family, friends, and health care providers (not all apps)	N/A	N	Y
[59]	Electronic Patient Reported Outcome tool (ePRO)	2016	Canada	Prospective study	4 weeks	Multiple	Patients (n=8) and health care providers (n=6)	Health care providers	Health care providers	N	N
[60]	STARFISH	2016	UK	Prospective study	6 weeks	Stroke	Patients (n=23)	Peers (automatic)	N/A	N	ActivPAL™ activity monitor

Reference	Intervention name	Year	Country	Study design	Duration	Health condition	Participants	Intended secondary users	Others involved in the intervention	Medical device included (Y/N)	Other devices included
[61]	HeartMapp	2016	USA	Cross-sectional study	Single evaluation	Heart failure	Patients (n=25) and health care providers (n=12)	Health care providers	Health care providers	Y	Zephyr Bioharness or Biopatch
[62]	EDGE digital health system	2017	UK	RCT	48 weeks	Chronic obstructive pulmonary disease	Patients (n=110) and research nurses (n=2)	Health care providers (automatic)	Informal care givers	N	N
[63]	IBGStar Diabetes Manager Application	2017	Germany	Prospective study	12 weeks	Diabetes	Patients (n=51)	N/A	N/A	Y	iBGStar blood glucose meter
[64]	MyHeart	2017	USA	Prospective study	24 weeks	Heart failure	Patients (n=8) and nurses	Nurses (automatic)	Nurses	Y	Weight scale, blood pressure monitor, and glucose meter
[65]	—	2018	UK	Cross-sectional study	4 weeks	Cancer	Patients (n=23)	Peers and health care providers	N/A	N	N

[a]Not available.

[b]Schizophrenia spectrum disorder, bipolar disorder, or major depressive disorder [52,58].

[c]N/A: not applicable.

[d]RCT: randomized controlled trial.

[e]Study evaluated both apps and systems and therefore will appear in both categories.

App interventions mainly targeted mental health conditions (n=7), followed by diabetes (n=3) and cardiovascular and heart diseases (n=4), with one study evaluating multiple apps that were used to self-manage multiple health conditions (Table 1).

Patients were included in all studies, and the studies had between 3 and 156 participants (median 36, IQR 15-87, maximum 156). The exception was one study in which only researchers evaluated patient-operated apps according to Google recommendations and quality standards [35,39]. Although studies tested single apps intended to be used primarily by patients, two studies also explored the impact of patients sharing their collected data with health care providers [35,42].

Six studies utilized single evaluations, either through a cross-sectional design [35,37,39,45-47] or an analytic service to analyze data available through the app store [41]. The remaining studies evaluated the impacts of app use over time, lasting between 4 and 72 weeks, with a mean period of 22.75 weeks (median 16 weeks, IQR 6-36, maximum 72). Of these, four utilized prospective study designs, three were randomized controlled trials (RCTs), and two used a retrospective design.

Among the 17 studies that described mHealth systems, most involved patients diagnosed with cardiovascular and heart disease (n=6), followed by diabetes (n=5), respiratory disease (n=2), mental health conditions (n=2), cancer (n=1), and multiple illnesses (n=1; Table 2).

As with mHealth app studies, all system studies, except one [35], involved patients. The 16 studies had between 6 and 151 patients (median 30, IQR 14.5-51.5, maximum 151), with eight studies involving health care providers. In these cases, health care providers either provided input on the suitability of an app for patient use or reviewed patient-gathered data during consultations.

In 12 studies, patients were required to share data (n=6) [50,51,57,60,62,64] or encouraged to share data (n=6) [35,53,55,59,61,65] with their health care providers or peers as part of the study. Data were also collected and transmitted to the main app by medical devices [50,51,53-57,61,63,64] and commercial wearables [35,52,53,58,60], demonstrating the prevalence of connectivity in modern mHealth systems.

Few studies (n=3) used single evaluations. RCTs (n=4) lasted longer (35.75 weeks on average) than cross-sectional studies (mean 24.5 weeks, n=2) and prospective studies (mean 12.93 weeks, n=7). Overall system evaluations lasted a mean of 20.32 weeks, which is very close to that for app interventions, but with a higher median number of 23 weeks.

Methods and Measures

Most studies included a combination of qualitative and quantitative methods of evaluation. Evaluation of usage logs was the most commonly adopted method (21 studies), followed by standardized questionnaires (17 studies; Table 3). Only two studies adopted quality guidelines to evaluate mHealth interventions; the Mobile Application Rating Scale was used to evaluate multiple apps [35], and compliance with Google standards for Android systems, in addition to other approaches, was used to evaluate the HeartKeeper app [39].

Table 3. Categories of methods used to evaluate mHealth interventions.

Methods (adopted approaches)	Studies that evaluated mHealth apps	Studies that evaluated mHealth systems
Evaluation of usage logs	[36,38,40-42,44,48,49]	[50,52,54,56-59,62-64]
Standardized questionnaires	[35-39,41-45,48,49]	[35,55-57,60,64]
Ad-hoc questionnaires	[36,37,40,42-44,47]	[51,53,55-58,61-63]
Interviews	[40,45,46]	[50,52,58,59,65]
Clinical outcomes	[36,48]	[54-56,63,64]
Open feedback (ie, oral or written)	[35,41,43,45]	[35,53,62]
Collection of additional device data (eg, medical device data)	N/A[a]	[54,56,57,60,62,64]
Field study and observation	[46,47]	[61,65]
Focus groups	N/A	[59,64]
Observational tests (in a lab setting)	[45,47]	N/A
Quality guidelines	[35,39]	[35]
Medical record entries	[42]	[63]
Attendance (intervention assigned activities/meetings)	[42,48]	N/A
Download count	[41]	N/A

[a]N/A: not applicable.

Among the 14 ad-hoc questionnaires used, four were developed according to concepts or questions from standardized questionnaires [47,58,61,62]. Similarly, two studies included interviews, where the interview guides were based on standardized questionnaires [40,45]. Some standardized questionnaires were used in more than one study. Multimedia Appendix 6 lists these questionnaires and illustrates the combination of questionnaires used in each study. Compared with traditional medical device testing, relatively few studies included information gathered from medical record entries (n=2), clinical outcomes (n=9), or observational tests (n=2).

Of note, some studies inferred more information from usage logs than the count and type of app interactions and patient-gathered data. For example, Triantafyllidis et al [50] interpreted information from the evaluation of usage logs on the usability of the device and participants' engagement in the study. The complete set of the types of data that were measured and collected by the mHealth app and system intervention studies are listed in Table 4.

Table 4. Categories of qualitative and quantitative data that were measured to evaluate mHealth interventions.

Types of data measured	Studies that evaluated mHealth apps	Studies that evaluated mHealth systems
Interactions (via app)	[36,37,40-42,44,45,49]	[50,52,53,56-59,62-65]
Usability/feasibility	[35,37,39-42,45,47]	[35,52,53,56,58,59,61,62,65]
Patient-gathered self-management data (via app)	[36-38,41,45,49]	[50,54,55,57,59,62-64]
Efficacy/effectiveness	[35-37,40,42,43,45,48]	[35,50,51,53,56,58,59,64,65]
Physical well-being	[36,40,42,48]	[54-57,60,62-64]
Perceptions, opinions, and suggestions	[35,40,41,45-47]	[35,51-53,58,64,65]
Intervention experiences	[39,41,46,47]	[50,52,58,59,64,65]
Psychological well-being	[38,41,42,44,49]	[55,60,62]
Patient-reported health	[40-44]	[56,63]
Self-efficacy	[36,44,47,49]	[55,57,61]
Engagement/motivation in self-management	[36,41]	[50,52,56,63]
Health care utilization and impact	[42]	[56,59,62-64]
Task performance	[45-47]	[50,61,65]
Study engagement	[35,41,42,48,49]	[35]
Patient-reported app use	[43,44]	[53,58,59]
Patient-reported self-management	[36,37]	[52,57,60]
Quality of life	[48]	[55,56,60,64]
App features and quality	[35,39,41,47]	[35]
Efficiency	N/A[a]	[62,65]
Security	[39]	[51]
Lifestyle	[48]	N/A

[a]N/A: not applicable.

Although a single method can often provide information regarding more than one measure, over one-third of the studies in this review used more than one method to collect information on one type of measure [40,42,45,48,50,55-60]. For example, two studies used both the collection of additional device data and clinical outcomes to report physical well-being [54,64]. Multimedia Appendix 7 includes a description of which measures were produced by each method. Several of the studies collected information on twice as many types of data measured as methods used to collect them (n=9) [35,41,44,49,58-60,65], with two studies collecting three [51,52] and one collecting four [39] times the number of types of data measured as methods used to collect them. Only one study used four methods to evaluate the most unique data types that were measured (n=10) by utilizing information resources that mHealth technologies make available (eg, automatically collected data from current users in the Android app store) [41].

Conversely, measures can be reported using more than one method. For example, usability/feasibility was the most common measure (22 times in 17 studies), followed by efficacy/effectiveness (20 times in 16 studies), interactions (via app; 19 times in 19 studies), physical well-being (18 times in 13 studies), and patient-gathered self-management data (via app; 15 times in 14 studies; Multimedia Appendix 7).

The study by Possemato et al [42] described the only app intervention that measured health care utilization and impact from these methods. Kim et al [56], Alnosayan et al [64], and Sieber et al [63] described system interventions that measured health care utilization or impact (ie, hospitalizations reported by participating health care providers and hospitalizations recorded retroactively). The remaining studies (n=5) collected information regarding physical well-being from clinical outcomes measured by researchers or health care providers during follow-up [36,48,54,55,61].

More comprehensive mapping of methods and measures revealed that the methods that were used to produce the most diverse set of data were, as expected, interviews (n=9), standardized questionnaires (n=16), and study-specific questionnaires (n=13; Multimedia Appendix 7). However, evaluation of usage logs produced nearly as many different types of measures (n=8).

Objectives and Methods Versus Results

A comparison of the study objectives with the results demonstrated that 30 of the 31 studies reported the results that they intended. One study reported all but one of the intended results described in the original objectives (ie, whether the reviewed apps and systems had been previously validated) [35]. Ten studies reported more than they anticipated, some of which included the assessment of app [42,48] and system [50] usage

patterns, as well as comparisons with other outcomes [41,44]. Other unforeseen outcomes included the accuracy of the app's knowledge base, as evaluated by nurses [45]; usability according to patients' performance of predetermined tasks with the app [47]; usability of connected devices in an mHealth system [53]; health care utilization [56]; and patient-reported symptoms [63]. Two studies stated that the objective was to develop mHealth systems; however, their outcomes also included evaluation results [50,51]. None of the studies phrased their goals as research questions and some reported what they intended, but the objective was not explicitly stated or detailed [40,63]. For example, Velardo et al [62] stated their intention to evaluate their intervention at scale. However, it was not clear how they intended to "evaluate" their intervention.

Discussion

Principal Findings

We identified 31 studies that described evaluations of mHealth apps or systems, with one describing evaluation of both intervention types [35]. Our findings show that studies relied mostly upon more continuous measures. Except for the collection of additional device data used by system interventions but not app interventions, there were no significant differences between apps and systems with regard to their ability to produce the intended outcomes, health conditions, or types of methods or measures used within the studies. Overall, medical record entries [42], attendance of meetings or activities assigned by the intervention [63], and download count [41] were the least used methods for gathering information about an intervention's impact on patients and providers. On the other hand, evaluation of usage logs [36,38,40-42,44,48-50,52,54,56-59,62-64] and standardized questionnaires [35-39,41-45,48,49,55-57,60,64] were the most commonly used methods. These two approaches (ie, one traditional and one mHealth) were also commonly used together in the same studies, demonstrating that mHealth is supplementing, not replacing, traditional research approaches.

mHealth Trends Versus Methods and Measures Used

Although clinical integration of mHealth technologies is on the rise, only two studies described app interventions that were meant to be used by secondary users (ie, health care providers and family and friends) [35,42], with three involving health care providers in the evaluation process [42,45,48]. Despite the focus on data safety and security, as well as patient privacy, as described by the new General Data Protection Regulation [66] and established FDA [10,11] and CE marking [12] expectations for health-related technologies, only two studies included measures regarding security [39,51].

Need to Reassess Evaluation Standards

Health evaluation studies are meant to produce evidence and understanding of how various interventions could affect patients and providers in real-world health care settings. Traditionally, studies have been classified within a hierarchy based on their designs, methods, and measures used to evaluate health interventions [67]. Health professionals consider high-level studies to be those that use rigorous and strict study designs, such as RCTs [68]. These studies provide an objective and quantitative understanding of how an intervention would influence patient clinical health measures, cost, or health care resource use [69]. On the other hand, low-level studies are often those that rely upon subjective and flexible study designs (eg, qualitative studies of participants' perception of the intervention or its impact on their lifestyle) [70].

Challenges of Quality Assessment

Health intervention researchers are not given instructions or guidance about how to evaluate these mHealth apps or which additional evidence is needed to determine their comprehensive impacts on patients and providers. The recent addition of connected technologies, such as wearables and sensors, has introduced even more factors to the evaluation context. Interventions now vary from recording exercise, to decision support for patient self-management, to providing evidence of a patients' actions for health care providers, to review from a variety of data sources. Because of these new information sources, we cannot always anticipate all of the impacts of these diverse networks of mHealth self-management technologies. For example, 10 studies did not intend to obtain results related to certain factors, such as usage logs and patient-reported outcomes [41,42,44,50,53,63].

The assessment of a study's success, validity, or quality presents another challenge to traditional research practice. mHealth resources consist of factors that make standard quality assessments inconclusive for intervention studies. For example, identifying patterns of patient self-management habits and progress describes the impact of an mHealth intervention on a patient's behavior. However, the analysis of usage logs, as a measure of intervention effectiveness, patient engagement, or self-management practices, has been minimally investigated as an appropriate method. As demonstrated by some of the reviewed articles, usage logs, download counts, and online ratings of apps were interpreted as indications of patient engagement, self-management behavior, intervention reach [41], effectiveness, and intervention utility [40] or feasibility.

Comparing Objectives and Results to Determine Successful Use of Methods

As opposed to completing a formal quality assessment, we chose to determine whether a study was able to produce the evidence that it aimed to provide, using selected methods. Some studies that performed usage log analysis were able to produce more information than they anticipated. Possemato et al [42] stated their intention to assess the fidelity of the PTSD Coach intervention by comparing health care utilization and health outcomes between those who used the app with and without clinical support. They were able to provide evidence for the effectiveness and fidelity of the intervention among health care providers, symptoms, and clinical health parameters from questionnaires. Moreover, they provided evidence for participants' patterns of intervention use from usage logs. Thereby, they were able to discuss the relationship between health care provider involvement and reinforced use of the app, as patients may have felt more accountable for using the app to self-manage their post-traumatic stress disorder.

Among the 31 studies identified, one did not obtain all of the intended information (missing one of the intended outcomes) [35] and one was found to be inconclusive [53]. We found that it was challenging to determine the specific objective of a study when objectives were not stated as such or when they were vague. This made it difficult to determine if a study was successful in the use of its selected methods and study design to reach its goals. For example, Velardo et al [62] stated that they intended to evaluate the EDGE digital health system intervention at scale; however, they did not state how they intended to do so or provide a research question that they intended to answer. Sieber et al [63] did not state the objective of their study. Instead, they stated simply what was done (ie, investigated the effects of usage profiles on hemoglobin A1c). Without a stated objective, we are unable to judge the reliability of intervention studies, whether it be through standard traditional means or an alternative approach. Clear objectives must be included in order to validate mHealth resources as trustworthy and relevant measures for evaluating mHealth interventions.

Relevance

mHealth must work for health care providers as well as patients. Patients are more engaged in their health, and they incorporate mHealth into their self-management. Thus, patients are aware of and can even influence how an mHealth intervention should or could be used to influence the kind of impact that is relevant for them. Understanding the potential risks and benefits of patient-operated mHealth requires more continuous evidence of not only technical and clinical outcomes but also personal and psychological impacts. This review demonstrates, through the use of such measures as mHealth interactions and patient-gathered data via an app, that we as researchers have the resources at our disposal and are beginning to use them.

A 2016 study by Pham et al [71] called for alternative or additional methods and measures for mHealth clinical trials that address the additional needs of mHealth. As most mHealth technologies for chronic health self-management are intended to be always available and continuously used by the patient, research questions, approaches, and designs need to reflect the real-world situations in which patients use these apps and systems.

Several studies within the presented scoping review demonstrated an attempt to meet this call by including more flexibility in their intervention design. For example, the EDGE digital health system [62], PTSD Coach app [42,43], and HeartKeeper app [40] made the patient the "decision maker" by allowing the patient to choose which data are relevant for them to gather and share with their health care providers. Further, two studies focused on reporting that patient engagement improved as a result of using mHealth apps [36,52]. User engagement is a necessity for the success of any intervention. It is paramount to consider patients' intentions when using these apps outside of the clinic; we should deem an app's ability to engage patients with their health as necessary as clinical evidence. There are individuals who do not choose to manage their chronic illnesses at all, for example, those deemed "hard to reach," who may benefit from merely acknowledging their health challenge by using an app primarily for education, without the expectation of performing complicated and time-consuming self-management. Therefore, when judging the success, usefulness, or potential benefit of an evaluated mHealth intervention, there should be less of a hierarchical gap between clinical health change or improvement and patients' experiences and change in self-efficacy.

Limitations

We believe our review covers most of the articles that were published during the established period and dealt with mHealth interventions for chronic conditions. This review reported on patient-operated mHealth self-management and did not include other potentially relevant interventions, such as SMS-based interventions.

We chose to focus on self-management of chronic NCDs, as defined by the WHO, in addition to severe mental health conditions, according to the demand for solutions from two fields (the medical system and public app development market) [4,13,33,72]. As such, these health cases represented the most potential for including state-of-the-art technology studies, with chronically ill people consistently being the leading market. However, exclusion of preventive treatments and other chronic health challenges (eg, musculoskeletal diseases) may have excluded a large proportion of cases that both involve the use of self-management options and represent a relevant portion of the chronic disease burden for individuals and health care systems worldwide [73]. As such, this noninclusion may have omitted conditions that could have provided relevant insights into methods and measures used to assess motivational, educational, and empowering mHealth technologies for self-management.

Because we did not collect data on reported results for this scoping review and did not perform a systematic methodological quality assessment, we cannot comment on the usefulness or effectiveness of the mHealth app and system interventions presented in these studies.

Conclusion

Researchers are now using several mHealth resources to evaluate mHealth interventions for patient self-management of select NCDs. This is evident as studies relied mostly on more continuous measures, including usage logs [36,38,40-42,44,48-50,52,54,56-59,62-64] and patient-collected data from medical devices [54,56,57,60,62,64], in addition to pre-post measures, such as clinical health measures [36,40,48,54-56,63,64] and standardized questionnaires [35-39,41-45,48,49,55-57,60,64]. In doing so, they evaluated the health status, engagement, and feasibility of mHealth apps and systems. In this review, which focused on mHealth, we found that only 20% of the included studies relied solely on traditional study designs (eg, RCTs) and methods that measure only pre- and postintervention health changes. The findings illustrate that the tradition of focusing on "clinical effectiveness, cost-effectiveness, and safety" [74] or health-related quality of life and the use of health care resources [75] is not being replaced, but is instead being expanded by taking advantage of additional resources that mHealth provides to evaluate interventions.

There is still no clear standard for the evaluation of mHealth interventions for patient self-management of chronic conditions. However, because mHealth presents additional challenges, needs, and resources to the field of health intervention research, we have the opportunity to expand and maintain our relevance to patients, providers, and health authorities. mHealth provides new types of information that we can and should gather to determine the impact of the interventions.

The presented results demonstrate that health studies have started to take advantage of additional mHealth resources, such as app usage logs and other patient-involved research methods, to determine the comprehensive impacts of mHealth on patients and other stakeholders. We are able to not only answer questions, such as which tasks patients choose to perform during interventions that may affect their clinical outcomes, but also say more about the relevance of mHealth for various types of users. This is essential in health intervention research, as the call for evidence on mHealth continues to push for not only traditional clinical health measures but also impacts on patients' self-efficacy and engagement. We believe that to achieve a compromise between the rigidity of traditional quality standards and the push for more patient-relevant outcomes, the definition of quality or meaningful impact, as well as available and appropriate evidence should be reassessed.

Acknowledgments

As a PhD candidate, the primary author is grateful for the input and guidance of the coauthors, who include all of the supervisors as part of the multidisciplinary Full Flow Project. This work was conducted as part of the Full Flow Project, which is funded by the Research Council of Norway (number 247974/O70). The publication charges for this article have been funded by a grant from UiT-The Arctic University of Norway's publication fund.

Authors' Contributions

MB, EG, and EÅ developed the search and inclusion criteria. MB and PJ performed the literature search, article screening, and data collection. EG served as a third reviewer when disputes surrounding the inclusion of an article arose. MB performed data synthesis and drafting of the manuscript. PZ contributed to the planning and editing of the manuscript. EG and EÅ additionally contributed to the editing of the text. MJ and RJ provided quality assurance of the manuscript and the necessary details within the description of the literature search and article selection. LPH guided article content. All authors have read and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 2500 KB - mhealth_v8i4e16814_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 127 KB - mhealth_v8i4e16814_app2.docx]

Multimedia Appendix 3
Scope of included technologies.
[DOCX File , 81 KB - mhealth_v8i4e16814_app3.docx]

Multimedia Appendix 4
Inclusion and exclusion criteria by category.
[DOCX File , 15 KB - mhealth_v8i4e16814_app4.docx]

Multimedia Appendix 5
Comparison of study objectives to reported results.
[DOCX File , 26 KB - mhealth_v8i4e16814_app5.docx]

Multimedia Appendix 6
List of questionnaires and scales used in mHealth intervention studies.
[DOCX File , 18 KB - mhealth_v8i4e16814_app6.docx]

Multimedia Appendix 7
Mapping of measures to methods.
[DOCX File , 15 KB - mhealth_v8i4e16814_app7.docx]

References

1. Olczuk D, Priefer R. A history of continuous glucose monitors (CGMs) in self-monitoring of diabetes mellitus. Diabetes Metab Syndr 2018;12(2):181-187 [FREE Full text] [doi: 10.1016/j.dsx.2017.09.005] [Medline: 28967612]
2. Pole Y. Evolution of the pulse oximeter. International Congress Series 2002 Dec;1242:137-144 [FREE Full text] [doi: 10.1016/s0531-5131(02)00803-8]
3. Omer T. Empowered citizen 'health hackers' who are not waiting. BMC Med 2016 Aug 17;14(1):118 [FREE Full text] [doi: 10.1186/s12916-016-0670-y] [Medline: 27530970]
4. Research2Guidance. Berlin, Germany: Research2Guidance; 2018. mHealth Developer Economics: Connectivity in Digital Health URL: https://research2guidance.com/product/connectivity-in-digital-health/ [accessed 2019-05-15]
5. Research2Guidance. Berlin, Germany: Research2Guidance; 2017. mHealth app economics 2017: current status and future trends in mobile health URL: https://tinyurl.com/y6urgf2x [accessed 2019-06-14]
6. Silvis L. US Food and Drug Administration.: US Food and Drug Administration; 2018 Oct 25. The Long Run Is Now: How FDA is Advancing Digital Tools for Medical Product Development URL: https://tinyurl.com/y9ssrspf [accessed 2019-06-14]
7. nyemetoder.no.: The Regional Health Authorities, The Norwegian Medicines Agency, The Norwegian Knowledge Centre for Health Services, The Norwegian Directorate of Health; 2014 Jan 23. The national system for the introduction of new health technologies within the specialist health service – For better and safer patient care URL: https://nyemetoder.no/Documents/Administrativt%20(brukes%20kun%20av%20sekretariatet!)/System%20Description%20(23012014).pdf [accessed 2020-04-11]
8. Gurupur VP, Wan TT. Challenges in implementing mHealth interventions: a technical perspective. Mhealth 2017;3:32. [doi: 10.21037/mhealth.2017.07.05] [Medline: 28894742]
9. Kotz D. A threat taxonomy for mHealth privacy. : IEEE; 2011 Feb 17 Presented at: 2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011); 2011; Bangalore, India p. 4-8 URL: https://ieeexplore.ieee.org/document/5716518 [doi: 10.1109/comsnets.2011.5716518]
10. FDA.gov. Rockville, MD: US Food & Drug Administration; 2015 Sep 05. Humanitarian Use Device (HUD) Designations: Guidance for Industry and Food and Drug Administration Staff URL: https://tinyurl.com/y8md9el6 [accessed 2018-06-12]
11. FDA.gov.: The U.S. Food and Drug Administration; 2019 May 11. Device Software Functions Including Mobile Medical Applications URL: https://tinyurl.com/y93jtst8 [accessed 2019-10-03]
12. Berensmann M, Gratzfeld M. Requirements for CE-marking of apps and wearables. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar;61(3):314-320. [doi: 10.1007/s00103-018-2694-2] [Medline: 29368120]
13. Research2Guidance.com.: Research2Guidance; 2016 Oct. mHealth Economics 2016 – Current Status and Trends of the mHealth App Market URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/ [accessed 2017-08-01]
14. Vallespin B, Cornet J, Kotzeva A. Ensuring Evidence-Based Safe and Effective mHealth Applications. Stud Health Technol Inform 2016;222:248-261. [Medline: 27198107]
15. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
16. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
17. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
18. nice.org.uk. UK: NICE; 2019 Mar. Evidence Standards Framework For Digital Health Technologies URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2019-06-14]
19. Ferretti A, Ronchi E, Vayena E. From principles to practice: benchmarking government guidance on health apps. The Lancet Digital Health 2019 Jun;1(2):e55-e57. [doi: 10.1016/s2589-7500(19)30027-5]
20. New checklist published to help improve reporting of mHealth interventions. WHO 2018 Apr 22 [FREE Full text]
21. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
22. Cheryl A, Jose FF, Christopher B, Joaquin B, Hamish F, Richard G. WHO. Bellagio, Italy: World Health Organization; 2011 Sep. Call to Action on Global eHealth Evaluation: Consensus Statement of the WHO Global eHealth Evaluation Meeting, Bellagio, Italy, September URL: https://tinyurl.com/yc97wzyh [accessed 2019-06-11]

23. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
24. Charani E, Castro-Sánchez E, Moore LS, Holmes A. Do smartphone applications in healthcare require a governance and legal framework? It depends on the application!. BMC Med 2014 Feb 14;12:29 [FREE Full text] [doi: 10.1186/1741-7015-12-29] [Medline: 24524344]
25. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, Implementing, and Evaluating Mobile Health Technologies for Managing Chronic Conditions in Older Adults: A Scoping Review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
26. Woods L, Duff J, Cummings E, Walker K. Evaluating the Development Processes of Consumer mHealth Interventions for Chronic Condition Self-management: A Scoping Review. Comput Inform Nurs 2019 Jul;37(7):373-385. [doi: 10.1097/CIN.0000000000000528] [Medline: 31135468]
27. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
28. FDA.org. Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Device Evaluation; 1998 Nov 04. General/Specific Intended Use - Guidance for Industry URL: https://tinyurl.com/y7c7dqws [accessed 2019-06-03]
29. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
30. nlm.nih.gov. Bethesda, MD: National Library of Medicine; 2019 Sep 09. MEDLINE, PubMed, and PMC (PubMed Central): How are they different? URL: https://www.nlm.nih.gov/bsd/difference.html [accessed 2017-07-01]
31. Thomson R. Endnote.com.: Clarivate; 2020. EndNote X7 URL: https://endnote.com/ [accessed 2019-06-04]
32. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
33. euro.who.int. Copenhagen, Denmark: World Health Organization; 2016. Action Plan for the Prevention and Control of Noncommunicable Diseases in the WHO European Region 2016–2025 URL: https://tinyurl.com/y9mjbk57 [accessed 2018-03-15]
34. ncbi.nlm.nih.gov. Bethesda, MD: US National Library of Medicine; 2018. Self-Management URL: https://www.ncbi.nlm.nih.gov/mesh/?term=self-management [accessed 2019-07-08]
35. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
36. Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015 Feb 03;17(2):e33 [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]
37. Kim YJ, Rhee SY, Byun JK, Park SY, Hong SM, Chin SO, et al. A Smartphone Application Significantly Improved Diabetes Self-Care Activities with High User Satisfaction. Diabetes Metab J 2015 Jun;39(3):207-217 [FREE Full text] [doi: 10.4093/dmj.2015.39.3.207] [Medline: 26124991]
38. Schärer LO, Krienke UJ, Graf S, Meltzer K, Langosch JM. Validation of life-charts documented with the personal life-chart app - a self-monitoring tool for bipolar disorder. BMC Psychiatry 2015 Mar 14;15:49 [FREE Full text] [doi: 10.1186/s12888-015-0414-0] [Medline: 25885225]
39. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Experiences and Results of Applying Tools for Assessing the Quality of a mHealth App Named Heartkeeper. J Med Syst 2015 Nov;39(11):142 [FREE Full text] [doi: 10.1007/s10916-015-0303-6] [Medline: 26345452]
40. de Garibay VG, Fernández MA, de la Torre-Díez I, López-Coronado M. Utility of a mHealth App for Self-Management and Education of Cardiac Diseases in Spanish Urban and Rural Areas. J Med Syst 2016 Aug;40(8):186. [doi: 10.1007/s10916-016-0531-4] [Medline: 27329050]
41. 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;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
42. 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 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
43. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: A pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]

44. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
45. Kang H, Park H. A Mobile App for Hypertension Management Based on Clinical Practice Guidelines: Development and Deployment. JMIR Mhealth Uhealth 2016 Feb 02;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
46. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of Commercially Available Mobile Applications for Diverse Patients. J Gen Intern Med 2016 Dec;31(12):1417-1426. [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
47. Fortuna KL, Lohman MC, Gill LE, Bruce ML, Bartels SJ. Adapting a Psychosocial Intervention for Smartphone Delivery to Middle-Aged and Older Adults with Serious Mental Illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828 [FREE Full text] [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]
48. Munster-Segev M, Fuerst O, Kaplan SA, Cahn A. Incorporation of a Stress Reducing Mobile App in the Care of Patients With Type 2 Diabetes: A Prospective Study. JMIR Mhealth Uhealth 2017 May 29;5(5):e75 [FREE Full text] [doi: 10.2196/mhealth.7408] [Medline: 28554881]
49. Silva Almodovar A, Surve S, Axon DR, Cooper D, Nahata MC. Self-Directed Engagement with a Mobile App (Sinasprite) and Its Effects on Confidence in Coping Skills, Depression, and Anxiety: Retrospective Longitudinal Study. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e64 [FREE Full text] [doi: 10.2196/mhealth.9612] [Medline: 29549066]
50. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: The SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
51. Park HS, Cho H, Kim HS. Development of Cell Phone Application for Blood Glucose Self-Monitoring Based on ISO/IEEE 11073 and HL7 CCD. Healthc Inform Res 2015 Apr;21(2):83-94 [FREE Full text] [doi: 10.4258/hir.2015.21.2.83] [Medline: 25995960]
52. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216 [FREE Full text] [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]
53. Årsand E, Muzny M, Bradway M, Muzik J, Hartvigsen G. Performance of the first combined smartwatch and smartphone diabetes diary application study. J Diabetes Sci Technol 2015 May;9(3):556-563 [FREE Full text] [doi: 10.1177/1932296814567708] [Medline: 25591859]
54. Skrøvseth SO, Årsand E, Godtliebsen F, Joakimsen RM. Data-Driven Personalized Feedback to Patients with Type 1 Diabetes: A Randomized Trial. Diabetes Technol Ther 2015 Jul;17(7):482-489 [FREE Full text] [doi: 10.1089/dia.2014.0276] [Medline: 25751133]
55. Holmen H, Wahl A, Torbjørnsen A, Jenum AK, Småstuen MC, Ribu L. Stages of change for physical activity and dietary habits in persons with type 2 diabetes included in a mobile health intervention: the Norwegian study in RENEWING HEALTH. BMJ Open Diabetes Res Care 2016;4(1):e000193 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000193] [Medline: 27239317]
56. Kim M, Lee S, Jo E, Lee S, Kang M, Song W, et al. Feasibility of a smartphone application based action plan and monitoring in asthma. Asia Pac Allergy 2016 Jul;6(3):174-180 [FREE Full text] [doi: 10.5415/apallergy.2016.6.3.174] [Medline: 27489790]
57. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
58. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable Devices and Smartphones for Activity Tracking Among People with Serious Mental Illness. Ment Health Phys Act 2016 Mar;10:10-17 [FREE Full text] [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]
59. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The Electronic Patient Reported Outcome Tool: Testing Usability and Feasibility of a Mobile App and Portal to Support Care for Patients With Complex Chronic Disease and Disability in Primary Care Settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
60. Paul L, Wyke S, Brewster S, Sattar N, Gill JM, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
61. Athilingam P, Labrador MA, Remo EF, Mack L, San Juan AB, Elliott AF. Features and usability assessment of a patient-centered mobile application (HeartMapp) for self-management of heart failure. Appl Nurs Res 2016 Nov;32:156-163. [doi: 10.1016/j.apnr.2016.07.001] [Medline: 27969021]

62. Velardo C, Shah SA, Gibson O, Clifford G, Heneghan C, Rutter H, EDGE COPD Team. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19 [FREE Full text] [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
63. Sieber J, Flacke F, Link M, Haug C, Freckmann G. Improved Glycemic Control in a Patient Group Performing 7-Point Profile Self-Monitoring of Blood Glucose and Intensive Data Documentation: An Open-Label, Multicenter, Observational Study. Diabetes Ther 2017 Oct;8(5):1079-1085 [FREE Full text] [doi: 10.1007/s13300-017-0306-z] [Medline: 28913822]
64. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Houston Feenstra L. Design and Usability of a Heart Failure mHealth System: A Pilot Study. JMIR Hum Factors 2017 Mar 24;4(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.6481] [Medline: 28341615]
65. Brett J, Boulton M, Watson E. Development of an e-health app to support women prescribed adjuvant endocrine therapy after treatment for breast cancer. Patient Prefer Adherence 2018;12:2639-2647 [FREE Full text] [doi: 10.2147/PPA.S187692] [Medline: 30587936]
66. Council of the European Union, European Parliament. Publications Office of the EU. Brussels, Belgium: Official Journal of the European Union; 2016 Apr 27. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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 (General Data Protection Regulation) (Text with EEA relevance) URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2019-06-12]
67. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011 Jul;128(1):305-310 [FREE Full text] [doi: 10.1097/PRS.0b013e318219c171] [Medline: 21701348]
68. Winona State University. Winona, MN: Winona State University; 2019 Sep 12. Evidence Based Practice Toolkit URL: https://libguides.winona.edu/c.php?g=11614&p=61584 [accessed 2019-10-15]
69. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. JAMA 2000 Sep 13;284(10):1290-1296. [doi: 10.1001/jama.284.10.1290] [Medline: 10979117]
70. Giacomini M. The rocky road: qualitative research as evidence. ACP J Club 2001;6(1):4-6 [FREE Full text] [doi: 10.1136/ebm.6.1.4]
71. Pham Q, Wiljer D, Cafazzo JA. Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
72. WHO.: World Health Organization; 2020 Jan 20. Depression URL: https://www.who.int/en/news-room/fact-sheets/detail/depression [accessed 2020-01-30]
73. WHO.: World Health Organization; 2019 Nov 26. Musculoskeletal conditions URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2020-01-30]
74. Bidonde J, Fagerlund B, Fronsdal K, Lund U, Robberstad B. fhi.no. Oslo, Norway: Norwegian Institute of Public Health; 2017 Aug. FreeStyle Libre Flash Glucose Self‐Monitoring System: A Single‐Technology Assessment URL: https://tinyurl.com/y8mymfru [accessed 2019-06-24]
75. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: An overview of reviews. Chron Respir Dis 2017 Aug;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]

Abbreviations

MeSH: Medical Subject Headings
mHealth: mobile health
NCD: noncommunicable disease
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.10.19; peer-reviewed by R Grainger, S Baptista, D Gunasekeran, W Zhang; comments to author 19.11.19; revised version received 10.02.20; accepted 25.03.20; published 30.04.20.

Please cite as:
Bradway M, Gabarron E, Johansen M, Zanaboni P, Jardim P, Joakimsen R, Pape-Haugaard L, Årsand E
Methods and Measures Used to Evaluate Patient-Operated Mobile Health Interventions: Scoping Literature Review
JMIR Mhealth Uhealth 2020;8(4):e16814
URL: https://mhealth.jmir.org/2020/4/e16814
doi:10.2196/16814
PMID:32352394

Original Paper

Physician Anxiety and Burnout: Symptom Correlates and a Prospective Pilot Study of App-Delivered Mindfulness Training

Alexandra Roy[1], BA; Susan Druker[2], MS; Elizabeth A Hoge[3], MD; Judson A Brewer[1], MD, PhD

[1]Brown University, Providence, RI, United States

[2]University of Massachusetts Medical School, Worcester, MA, United States

[3]Georgetown University, Washington, DC, United States

Corresponding Author:
Judson A Brewer, MD, PhD
Brown University
Box 1875
Brown Mindfulness Center
Providence, RI,
United States
Phone: 1 401 863 2826
Email: judson_brewer@brown.edu

Abstract

Background: Physician burnout is on the rise, yet little is known about its relationship to anxiety. Mindfulness-based stress reduction has demonstrated decreases in anxiety, yet physicians have reported reluctance to engage in it due to significant time commitments.

Objective: The aims of this study are to assess whether app-based mindfulness training can reduce anxiety in physicians and to explore if anxiety and burnout are correlated, thus leading to a reduction in both anxiety and burnout.

Methods: This was a nonrandomized pilot study comprised of 34 physicians who worked in a large US health care network and reported having anxiety. The intervention was an app-based mindfulness program. The main outcome measure was anxiety, measured by the Generalized Anxiety Disorder-7 (GAD-7). The secondary outcome measures assessed burnout: cynicism and emotional exhaustion items from the Maslach Burnout Inventory.

Results: GAD-7 scores decreased significantly at posttreatment (1 month after treatment initiation, 48% reduction, $P<.001$) and at the 3-month follow-up (57% reduction, $P<.001$). There was a significant correlation between anxiety and burnout (cynicism: $r=.43$; $P=.01$; emotional exhaustion: $r=.71$; $P<.001$). There was also a significant decrease in cynicism (50% reduction, $P=.003$ at posttreatment; 50% reduction, $P=.009$ at follow-up) and emotional exhaustion at both time points (20% reduction, $P<.001$ at posttreatment; 20% reduction, $P=.003$ at follow-up).

Conclusions: This pilot study is the first to test an app-based mindfulness training program targeted at reducing anxiety with physicians and to demonstrate that in physicians, anxiety is correlated with burnout. These findings suggest that this may be an effective tool to reduce anxiety and burnout in physicians.

Trial Registration: ClinicalTrials.gov NCT04137081; https://www.clinicaltrials.gov/ct2/show/NCT04137081

(JMIR Mhealth Uhealth 2020;8(4):e15608) doi:10.2196/15608

KEYWORDS

anxiety; burnout; mindfulness; app; mHealth; physician; smartphone; digital therapeutics

Introduction

Physician burnout has gained increasing attention and concern over the past several years due to its effects on physicians, the direct impact it has on patient care, and the increase in prevalence. A 2011 study found that 38% of physicians reported significant symptoms of burnout *on a weekly basis*, which is 10% higher than the general working population in the United States [1]. By 2014, burnout had increased to 48% among physicians, nearly double that of the general working population, which had not shown increases in the same time period [2].

Symptoms of physician burnout include exhaustion, irritability, inability to concentrate, and cynicism, among other symptoms [3,4]. Burnout has been reported to be associated with sleep

disturbances, marital difficulties, depression, and anxiety [4]. Of note, while intuitive at face value, reported links between burnout and anxiety have largely been anecdotal. Burnout and anxiety may share similar presenting symptoms such as sleep disturbance and irritability; however, to date, no studies have directly assessed the correlation between these variables in physicians.

Both institutional and individual factors are theorized to contribute to the significant increase in physician burnout over the past few years. Institutional contributors include an increase in patient loads, the introduction of electronic medical record systems, and an emphasis on cost management, among other factors [5,6]. Additionally, as the health care landscape has moved increasingly to larger corporate structures, physicians have found themselves progressively removed from their practices' decision-making processes, including decisions pertaining to the length of patient visits and treatment approaches [7]. This trend in burnout seems to be gaining momentum despite long-standing knowledge of key related and possible causal factors; a study of physicians practicing in two Kaiser Permanente regions in the 1990s found that the single most important predictor of physician burnout was a lack of "perceived control over the practice environment" [8]. This reduced sense of control has been associated with a tendency to perceive clinical ineffectiveness and "give up more easily" [4,7].

Interestingly, the link between a perceived lack of control and tendency to give up has been extensively studied in cognitive neuroscience contexts, aptly termed "learned helplessness" (for a review, see Maier et al) [9]. Ironically, for physicians, this learned behavior may begin in medical school (or earlier), directly contributing to the development (or worsening) of anxiety and predisposing them to burnout [10]. In particular, medical students are subject to environments and situations that trigger stress and anxiety responses that can be reinforced over time through well-described operant conditioning learning pathways (reinforcement learning, positive and negative reinforcement) that can lead to feelings of helplessness and anxiety. Necessary components for this type of learning include a trigger, behavior, and reward [11,12]. Positive reinforcement learning might happen when a medical student is asked a question during hospital rounds (historically referred to as a *pimping* – trigger); if she provides the correct response (behavior), she may be rewarded with praise (reward). If she replies incorrectly, she may learn through negative reinforcement by being encouraged to look up the answer, or castigated in front of the team. Both positive and negative reinforcement can perpetuate stress and anxiety as students compete to have "the best" answers and avoid humiliation, self-imposed or otherwise.

On top of environmental causes and conditions, the high demands and inherent uncertainty in the practice of medicine may contribute to excessive worry, a core aspect of anxiety disorders [13]. From a psychological standpoint, worry represents an attempt to engage in mental problem solving for an issue with an uncertain outcome [14,15]. It has been suggested that a feeling of control over a situation can lead to the formation of a "habit loop" in which worry can be reinforced and perpetuated (ie, be positively reinforced, see Figure 1); though ironically, in this modern landscape where physician autonomy and shared decision-making are waning, this habit loop may compound anxiety and burnout [16-19]. The many pressures and institutional factors that contribute to burnout need to be addressed by health care systems; however, in the meantime, it may be possible to provide some tangible support to physicians.

Mindfulness training (MT) is gaining evidence and interest as a potential treatment for anxiety [20-25]. Mindfulness can be defined as the awareness that arises when paying attention in the present moment, on purpose and nonjudgmentally [26]. Through helping individuals observe emotions and bodily sensations instead of getting caught up in anxiety, MT has been theorized to directly target key reinforcement learning pathways [18,27]. Specifically, mindfulness helps individuals learn to identify perseverative worry thought patterns that reinforce anxiety habit loops and, importantly, to notice thoughts and emotions as mental and physical events and sensations instead of propagating the cycle. In other words, this objective observation decreases the degree to which individuals are identified with thoughts and emotions, effectively deconditioning or extinguishing the reinforcement learning process that perpetuates anxiety.

It is also unclear whether anxiety predisposes physicians to burnout or is exacerbated by burnout, or if the two interact to feed off of each other. Several studies have found that MT specifically decreases burnout in physicians, though whether components of burnout are mediated through a reduction of anxiety or another mechanism remains unknown [28-30]. Regardless, these promising findings are suggestive that MT may effectively address at least some components of burnout; however, multi-month programs such as mindfulness-based stress reduction (MBSR) are time-intensive and difficult for many physicians to incorporate into their already busy schedules. For example, in one study, 44% of health care professionals randomized to an MBSR intervention could not complete the program due to a "lack of time" [31]. This suggests a need for targeted mindfulness-based interventions that address anxiety yet are tailored to fit within physicians' already busy lives.

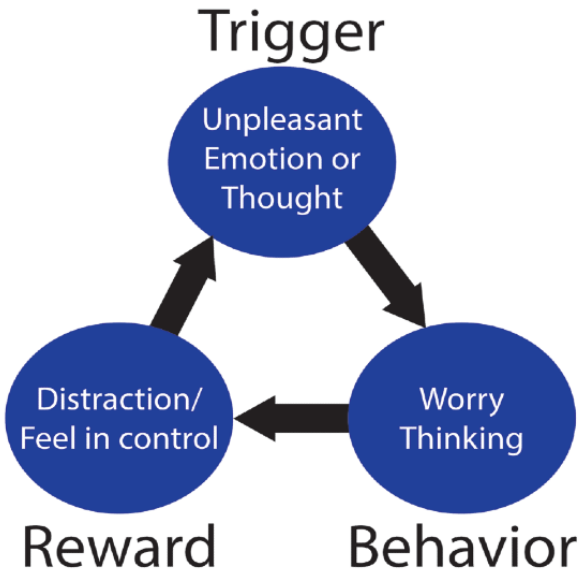
App-based smartphone interventions are increasingly used to deliver behavioral treatments because of their relative low cost, high fidelity, and broad availability. This new class of "digital therapeutics" may also be an effective way to deliver MT to physicians. For example, didactic training can be tailored to be delivered in short, daily modules (eg, 10 minutes/day) rather than weekly 2-hour sessions; brief, just-in-time mindfulness tools can be accessed on demand and in context when triggered by stress or anxiety. With these factors in mind, we developed an app-based MT program for anxiety that targets the underlying reinforcement learning pathways where anxiety develops and is perpetuated. From a path model perspective, MT could reduce anxiety via a reduction in worry or an increase in nonreactivity.

Previously, we used the Obesity-Related Behavioral Intervention Trials framework to develop an app to target anxiety as a behavioral risk factor of interest (Phase I) [32]. This study (Phase II) is the next logical step to perform a "preliminary,

proof-of-concept test [in] a cost-effective way to determine if a treatment package can achieve benefit on a clinically significant target in a small, select sample."

The primary goal of this nonrandomized pilot study was to examine whether a mechanistically-based MT program delivered via app would reduce anxiety in physicians. In addition, the following questions were explored: are anxiety and burnout in physicians correlated, and can an app-based MT reduce physician burnout [33]? The primary hypothesis of this study was that app-based MT would decrease anxiety, and the secondary hypotheses were that the anxiety would be correlated with burnout and that burnout would also be reduced.

Figure 1. Development of a "habit loop" via positive and negative reinforcement [15,18,19].



Methods

Participants

Physicians in the University of Massachusetts Memorial Health Care system (N=1100) were invited via email to a 30-day app-based MT program, Unwinding Anxiety (UA), geared toward anxiety. Inclusion criteria included: direct patient interaction; currently employed as a physician; owned a smartphone; answered yes to the following questions: do you feel nervous, anxious, or on edge, and do you feel you worry too much about different things?; endorsed willingness to use a mindfulness smartphone app for 10 minutes per day for 30 days; endorsed willingness to complete two surveys at 1 month and 3 months. There were no explicit exclusion criteria. Consenting participants were provided the UA app and received a $25 Amazon gift card for completion of each follow-up survey ($50 total). A control condition such as treat-as-usual was not used due to concerns about participant retention. The study was approved by the University of Massachusetts Medical School Institutional Review Board.

Intervention

The app-based MT program teaches individuals to understand how anxious worry is developed and perpetuated through reinforcement learning, how to recognize these anxiety habit loops, and how to bring mindful awareness to moments of stress and worry, so they can uncouple feelings of anxiety from reactive worry thinking and mindfully work with habitual mind states that perpetuate and reinforce anxiety. This process helps individuals "unlearn" or extinguish worry at a core mechanistic level. This experiential education is delivered via a smartphone-based platform**,** which includes a progression of more than 30 daily modules of brief didactic and experience-based MT (videos and animations, approximately 10 min/day, see Textbox 1), app-triggered check-ins, user-initiated guided meditations (5-15 minutes), and brief (30 seconds) on-demand mindfulness exercises to help disrupt anxiety cycles in vivo. The content for this intervention was developed based upon a combination of clinical experience for individuals with anxiety and previously developed in-person and app-based MT protocols for habit change that have yielded clinically-meaningful outcomes such as cessation of smoking or overeating [18,34-36].

Textbox 1. Overview of Unwinding Anxiety themes and content.

Modules 1-7: goals; curiosity; reinforcement learning; body scan; self-monitoring

Sets goals and introduces how habits are formed around worry (eg, reinforcement learning, distraction). Introduces curiosity to foster the nonjudgmental aspects of mindfulness and basic mindfulness practices including the body scan. Unpacks worry and fear both from a brain and behavior perspective.

Modules 8-14: noting practice; RAIN; barriers to change; reinforce concepts

Introduces how to mindfully work with worry cues and affective states using RAIN (Recognize, Accept, Investigate, and Note what emotions feel like as they arise and pass away) [34,37]. These also build on basic mindfulness using noting practice (the N of RAIN) during everyday life, and introduce additional animations to reinforce mindfulness concepts that show how we feed our anxiety by worry thinking and distraction.

Modules 15-21: noting practice (cont'd); RAIN (cont'd); thinking vs knowing; (un)resistance

Reinforces noting practice and continues to train and support self-kindness. Specifically addresses the difference between trying to think our way out of uncertainty (or anxiety), and resting in a kind curious awareness of it. Modules also focus on not resisting experience and not getting tripped up by worry thinking.

Modules 22-30: noting practice (cont'd); RAIN (cont'd); working with uncertainty and change

Help individuals reflect on their own "evidence base" for working with worry to solidify their shift from reactivity to mindfully being with emotions as a new habit.

Modules 30+: Reinforcing concepts via "theme weeks" + individual customization via "personal week"

8 or more themed weeks and unlimited personalization of content by picking modules to develop a custom "week" for review.

Outcome Measures

Participants completed self-administered surveys at three different time points: baseline, one month (primary endpoint), and three months (secondary endpoint) after treatment initiation. Data related to age and program engagement, as measured by the number of modules completed, was obtained directly from the UA app. Each survey contained the following measures.

Generalized Anxiety Disorder-7

The Generalized Anxiety Disorder-7 (GAD-7) is a 7-item questionnaire that is clinically used for measuring and tracking anxiety severity [38]. A 4-point Likert scale of frequency ranging from "not at all" to "nearly every day" is used to measure each item.

Maslach Burnout Inventory

Two single-item measures of emotional exhaustion and cynicism (originally called depersonalization) were included from the Maslach Burnout Inventory (MBI)-based on West's [39] research, which demonstrated that these measures provided information on likelihood of burnout among medical professionals equivalent to the complete 22-item questionnaire (single emotional exhaustion question, r_s=0.76-0.83; single cynicism question, r_s=0.61-0.72) [40]. Furthermore, West [41] confirmed the concurrent validity of the single items in relation to the complete MBI in a follow-up study with physicians in 2011. These were measured using a 7-point Likert scale of frequency that ranged from "never" to "every day".

Participant Satisfaction

Participants were asked their likelihood to recommend the program to a friend on an 11-point Likert scale that ranged from "not at all likely" to "extremely likely".

Statistical Analysis

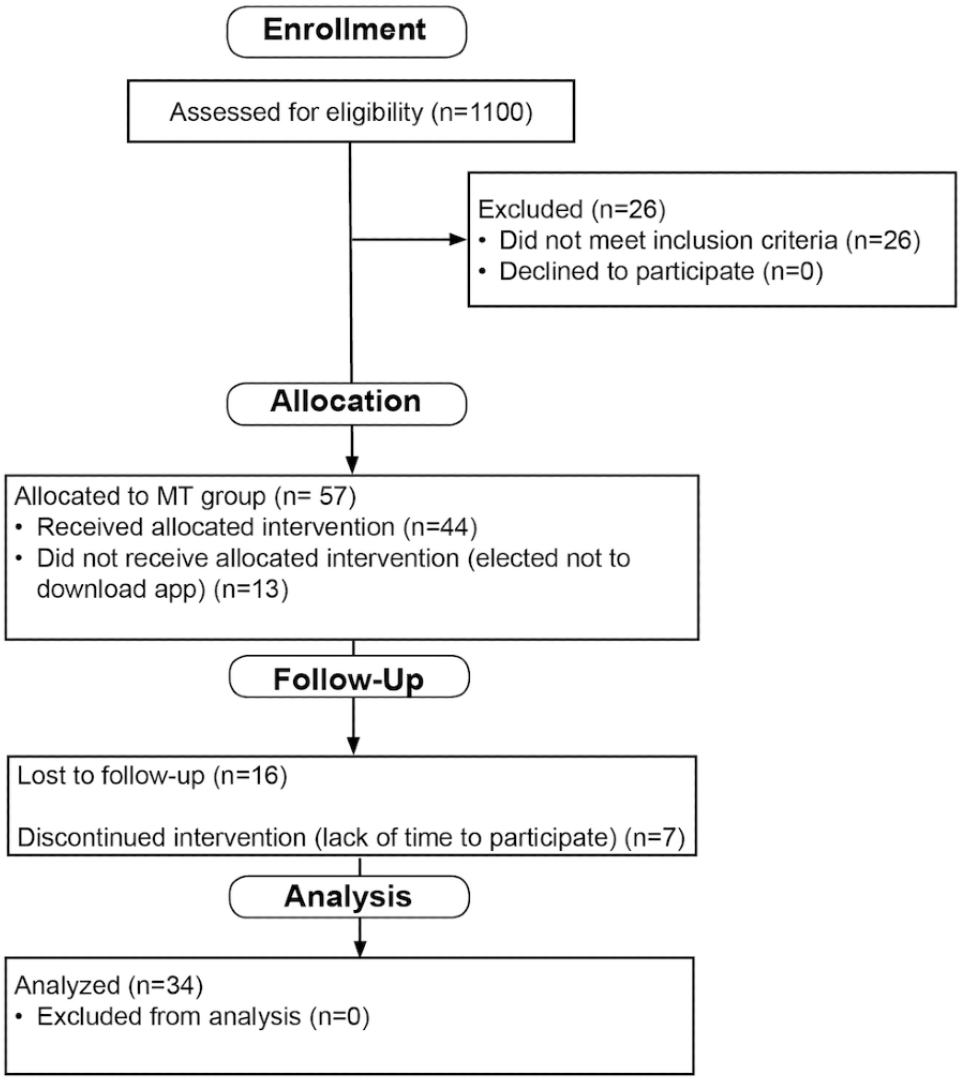
The data was analyzed using R version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria). Friedman's analysis of variance (ANOVA), a nonparametric test, was used to analyze the overall change in GAD-7 scores (primary outcome) and the single-item MBI scores (secondary outcomes) at the three time points due to the data having a nonnormal distribution. Post hoc analyses between the individual time points were analyzed using Wilcoxon signed-rank test and were corrected for multiple comparisons (Bonferroni). The relationship between GAD-7 and the single-item MBI scores was evaluated using Spearman's correlation coefficient. Effect sizes were determined by dividing the z-score by the square root of the sample size to find Pearson's *r* [42]. This uses Cohen's criteria for *r* where 0.1 is a small effect, 0.3 is a medium effect, and 0.5 is a large effect [43].

Results

Profile of Participants

A total of 57 participants met eligibility criteria and consented to participate during the spring and summer of 2018. Out of this group, 44 registered and downloaded the MT app. Out of 57 physicians, 34 (60%) were included in the final analysis (see Consolidated Standards of Reporting Trials diagram, Figure 2). The participant population was comprised of 25 women and 9 men who worked in health care for an average of 18 years (SD 10.25). The average age was 45 years (SD 8.89); although age data was not obtained for two individuals who completed the online surveys but did not download the smartphone app.

Figure 2. Consolidated Standards of Reporting Trials diagram. MT: mindfulness training.



Correlations Between Anxiety and Burnout

There was a significant correlation between anxiety and burnout at baseline, 1 month, and 3 months (Table 1) posttreatment initiation.

Table 1. Correlations between anxiety and burnout scores at baseline, 1 month, and 3 months after initiating mindfulness training.

Variables	GAD-7[a] x emotional exhaustion	*P* value	GAD-7 x cynicism	*P* value
Baseline	0.71	<.001	0.43	.01
1 month	0.67	<.001	0.53	.001
3 months	0.53	.001	0.55	<.001

[a]GAD-7: Generalized Anxiety Disorder-7.

Changes in Anxiety and Burnout

Friedman's ANOVA showed there were significant changes across the three time points for GAD-7 (x^2_2=40.14, *P*<.001), emotional exhaustion (x^2_2=16.70, *P*<.001), and cynicism (x^2_2=10.13, *P*=.006).

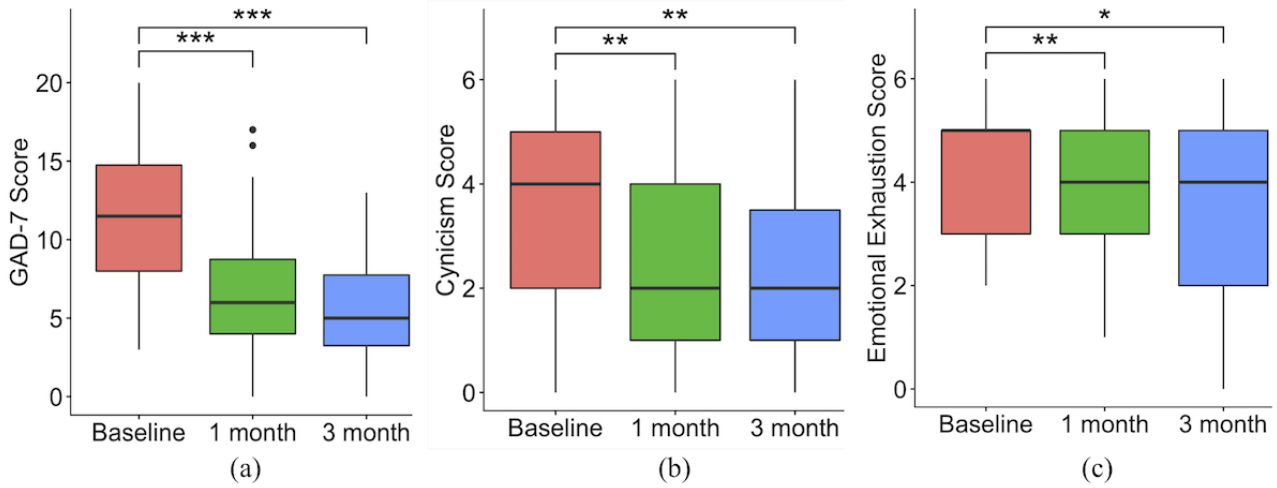
Figure 3 shows box and whisker plots for each time point; the horizontal line inside the box indicates the median while the top of the box indicates the third quartile, and the bottom represents the first quartile. The whiskers indicate the maximum and minimum values and the dots represent outliers. Participants demonstrated a 48% reduction in GAD-7 scores from baseline (median 11.50, interquartile range [IQR] 8-14.75) to 1 month (median 6, IQR 4-8.75, *P*<.001, effect size 0.81; see Figure 3a)

and a 57% reduction at 3 months (median 5, IQR 3.25-7.75, *P*<.001, effect size 0.85).

Participants reported a 50% reduction in cynicism scores from baseline (median 4, IQR 2-5) to 1 month (median 2, IQR 1-4, *P*=.003, effect size 0.54, Figure 3b) and a 50% reduction at 3 months (median 2, IQR 1-3.75, *P*=.009, effect size 0.49).

Participants reported a 20% reduction in emotional exhaustion scores from baseline (median 5, IQR 3-5) to 1 month (median 4, IQR 3-5, *P*<.001, effect size 0.63, Figure 3c) and a 20% reduction at 3 months (median 4, IQR 2-5, *P*=.003, effect size 0.56).

Figure 3. Box and whisker plots at baseline, one month, and three months for: (a) Generalized Anxiety Disorder-7 scores; (b) cynicism scores from Maslach Burnout Inventory; and (c) emotional exhaustion scores from Maslach Burnout Inventory. Significance level is denoted by asterisks: *=.05; **=.01; ***=.001. GAD-7: Generalized Anxiety Disorder-7.



Participant Engagement and Satisfaction

At 1 month, participants had completed an average of 11.03 (SD 9.51) modules of the 30-module program, and the average rating out of 10 for their likelihood to recommend the app to a friend was 7 (SD 2.79). At 3 months, participants had completed an average of 15.29 (SD 12.42) modules, and the average score for recommending the app was 8 out of 10 (SD 2.53).

Discussion

This is the first study to directly test an app-delivered mindfulness-based treatment in anxious physicians and to assess the relationship between physician anxiety and burnout. We found preliminary evidence that app-based MT reduced both anxiety and burnout, demonstrating that this modality and type of training may be an accessible and possibly effective tool for helping busy physicians manage these conditions. We found significant correlations between anxiety and burnout at all time points, confirming the hypothesized link between the two. This is promising, as the majority of participants were midcareer and already experiencing anxiety and burnout. It is important to develop evidence-based tools to support these individuals in combatting burnout, as they are likely to spend at least an additional 15 years in the workforce. This work expands upon previous findings that a correlation exists between the anxiety, as measured by the State-Trait Anxiety Inventory, and burnout, as measured by the MBI, in health care workers in the 1990s [10]. Additionally, this has also been shown in Chinese physicians using the Zung Self-Rating Anxiety scale and the Chinese version of the MBI [44].

This study builds on previous research that indicates that mindfulness-based interventions delivered in-person can decrease anxiety in individuals with moderate to severe anxiety (eg, generalized anxiety disorder) and decrease physician burnout [20,28,45,46]. A large and growing number of app-based mindfulness and meditation trainings have emerged as treatment modalities for anxiety, yet few are theory-driven and even fewer are supported by clinical trials. As the field of digital therapeutics grows, it will be critical for new treatments to demonstrate mechanistic, empirical, and clinically meaningful effects. This study is unique in that it is the first to use app-based MT to directly target a theorized mechanism underlying anxiety (reinforcement learning) and link it to clinical outcomes (anxiety and burnout). These results demonstrate a proof-of-concept that app-based MT may be a promising modality for the treatment of anxiety in physicians and, perhaps, more broadly; however, carefully-designed randomized controlled studies in specific treatment realms are required as next steps.

The link between anxiety and burnout may not be surprising at face value, but it is important to directly establish this correlation in physicians from both a scientific and treatment standpoint, as it informs both observations and outcomes. Specifically for this study, MT was directed toward reinforcement learning pathways that perpetuate anxiety. No modules or even references of burnout, cynicism, or emotional exhaustion were made in the MT program. We theorized that if anxiety and burnout were correlated, we might see a reduction in both symptoms by singularly addressing anxiety. This was indeed the case, and provides further explanatory power as to why MT might reduce burnout in anxious physicians, given the strong correlations that were found between the two.

This study has several novel and notable findings; however, there are a number of limitations, including a relatively small and self-selected sample, a limited geographic area, and a single

intervention arm. Proof-of-concept studies such as this one are a critical first step in exploring potential mechanisms and generating effect size calculations for larger randomized controlled trials. This study demonstrated that anxious physicians were willing to try an app-based MT program and might benefit from it. Another limitation is that the program was not tailored to physicians; it was designed to help anyone with moderate to severe anxiety.

MT has shown promise in helping physicians reduce burnout. This study shows a strong link between anxiety and some personal aspects of burnout, such as cynicism, and that targeting anxiety may help reduce both. This study also suggests that app-based MT may have promise as an accessible, evidence-based tool to combat anxiety and correlated aspects of burnout in physicians; although causal claims cannot definitively be made until future randomized controlled studies are conducted.

Acknowledgments

This study was supported by the Carl Akins Risk Management grant from the University of Massachusetts Memorial Medical Group. Thank you to Dr Eric Dickson, Eric Chang, Cheryl Laproire, and Alan Weston for their support with recruitment.

Conflicts of Interest

JAB owns stock in MindSciences, the company that developed the mindfulness app used in this study. This financial interest has been disclosed to and is being managed by Brown University, in accordance with its Conflict of Interest and Conflict of Commitment policies. All other authors report no biomedical financial interests or potential conflicts of interest.

References

1. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
2. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
3. Gundersen L. Physician burnout. Ann Intern Med 2001 Jul 17;135(2):145-148. [doi: 10.7326/0003-4819-135-2-200107170-00023] [Medline: 11453722]
4. Spickard A, Gabbe SG, Christensen JF. Mid-career burnout in generalist and specialist physicians. JAMA 2002 Sep 25;288(12):1447-1450. [doi: 10.1001/jama.288.12.1447] [Medline: 12243624]
5. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. Lancet 2009 Nov 14;374(9702):1714-1721. [doi: 10.1016/S0140-6736(09)61424-0] [Medline: 19914516]
6. Rothenberger DA. Physician burnout and well-being: a systematic review and framework for action. Dis Colon Rectum 2017 Jun;60(6):567-576. [doi: 10.1097/DCR.0000000000000844] [Medline: 28481850]
7. Southwick FS, Southwick SM. The loss of a sense of control as a major contributor to physician burnout: a neuropsychiatric pathway to prevention and recovery. JAMA Psychiatry 2018 Jul 01;75(7):665-666. [doi: 10.1001/jamapsychiatry.2018.0566] [Medline: 29799941]
8. Freeborn DK. Satisfaction, commitment, and psychological well-being among HMO physicians. West J Med 2001 Jan;174(1):13-18 [FREE Full text] [doi: 10.1136/ewjm.174.1.13] [Medline: 11154654]
9. Maier SF, Seligman MEP. Learned helplessness at fifty: Insights from neuroscience. Psychol Rev 2016 Jul;123(4):349-367 [FREE Full text] [doi: 10.1037/rev0000033] [Medline: 27337390]
10. Turnipseed DL. Anxiety and burnout in the health care work environment. Psychol Rep 1998 Apr 31;82(2):627-642. [doi: 10.2466/pr0.1998.82.2.627] [Medline: 9621740]
11. Brewer J. The Craving Mind: From Cigarettes To Smartphones To Love – Why We Get Hooked And How We Can Break Bad Habits. New Haven, CT: Yale University Press; 2017.
12. Skinner B. Operant behavior. Am Psychol 1963 Aug;18(8):503-515. [doi: 10.1037/h0045185]
13. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition (DSM-5). Washington, DC: American Psychiatric Publishing; 2019.
14. Borkovec TD, Robinson E, Pruzinsky T, DePree JA. Preliminary exploration of worry: some characteristics and processes. Behav Res Ther 1983;21(1):9-16. [doi: 10.1016/0005-7967(83)90121-3] [Medline: 6830571]
15. Borkovec TD, Ray WJ, Stober J. Worry: a cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. Cognitive Therapy and Research 1998;22(6):561-576. [doi: 10.1023/A:1018790003416]
16. Freeston M, Rhéaume J, Letarte H, Dugas M, Ladouceur R. Why do people worry? Pers and Individ Dif 1994 Dec;17(6):791-802. [doi: 10.1016/0191-8869(94)90048-5]
17. Brewer JA, Elwafi HM, Davis JH. Craving to quit: psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. Psychol Addict Behav 2013 Jun;27(2):366-379 [FREE Full text] [doi: 10.1037/a0028490] [Medline: 22642859]

18. Brewer JA, Ruf A, Beccia AL, Essien GI, Finn LM, van Lutterveld R, et al. Can mindfulness address maladaptive eating behaviors? why traditional diet plans fail and how new mechanistic insights may lead to novel interventions. Front Psychol 2018;9:1418 [FREE Full text] [doi: 10.3389/fpsyg.2018.01418] [Medline: 30250438]
19. Brewer J. Mindfulness training for addictions: has neuroscience revealed a brain hack by which awareness subverts the addictive process? Curr Opin Psychol 2019 Aug;28:198-203. [doi: 10.1016/j.copsyc.2019.01.014] [Medline: 30785066]
20. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]
21. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
22. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]
23. Hoge EA, Bui E, Marques L, Metcalf CA, Morris LK, Robinaugh DJ, et al. Randomized controlled trial of mindfulness meditation for generalized anxiety disorder: effects on anxiety and stress reactivity. J Clin Psychiatry 2013 Aug;74(8):786-792 [FREE Full text] [doi: 10.4088/JCP.12m08083] [Medline: 23541163]
24. Goldin P, Ziv M, Jazaieri H, Hahn K, Gross JJ. MBSR vs aerobic exercise in social anxiety: fMRI of emotion regulation of negative self-beliefs. Soc Cogn Affect Neurosci 2013 Jan;8(1):65-72 [FREE Full text] [doi: 10.1093/scan/nss054] [Medline: 22586252]
25. Wong SY, Tang WK, Mercer SW, Kung K, Mak WW, Griffiths SM, et al. Mindfulness-based cognitive therapy for generalised anxiety disorder and health service utilisation among Chinese patients in primary care: a randomised, controlled trial. Hong Kong Med J 2016 Dec;22 Suppl 6(6):35-36 [FREE Full text] [Medline: 27807315]
26. Kabat-Zinn J. Full Catastrophe Living: Using The Wisdom Of Your Body And Mind To Face Stress, Pain, And Illness. New York: Delta Trade Paperbacks; 1991.
27. Brewer J, Pbert L. Mindfulness: an emerging treatment for smoking and other addictions. J Fam Med 2015;2(4):1035.
28. Krasner MS, Epstein RM, Beckman H, Suchman AL, Chapman B, Mooney CJ, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. JAMA 2009 Sep 23;302(12):1284-1293. [doi: 10.1001/jama.2009.1384] [Medline: 19773563]
29. Martín-Asuero A, García-Banda G. The Mindfulness-based Stress Reduction program (MBSR) reduces stress-related psychological distress in healthcare professionals. Span J Psychol 2010 Nov;13(2):897-905. [doi: 10.1017/s1138741600002547] [Medline: 20977037]
30. Goodman MJ, Schorling JB. A mindfulness course decreases burnout and improves well-being among healthcare providers. Int J Psychiatry Med 2012;43(2):119-128. [doi: 10.2190/PM.43.2.b] [Medline: 22849035]
31. Shapiro S, Astin J, Bishop S, Cordova M. Mindfulness-based stress reduction for health care professionals: results From a randomized trial. International Journal of Stress Management 2005 Jan 15;12(2):164-176. [doi: 10.1037/1072-5245.12.2.164]
32. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
33. Eldridge S, Bond C, Campbell M, Hopewell S, Thabane L, Lancaster G, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: using consensus methods and validation to develop a conceptual framework. Trials 2015 Nov 16;16(S2). [doi: 10.1186/1745-6215-16-s2-o87]
34. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. J Behav Med 2018 Apr;41(2):160-173 [FREE Full text] [doi: 10.1007/s10865-017-9884-5] [Medline: 28918456]
35. Garrison K, Pal P, O'Malley S, Pittman B, Gueorguieva R, Rojiani R, et al. Craving to quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2018 Jun 18. [doi: 10.1093/ntr/nty126] [Medline: 29917096]
36. Brewer JA, Mallik S, Babuscio TA, Nich C, Johnson HE, Deleone CM, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. Drug Alcohol Depend 2011 Dec 01;119(1-2):72-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.05.027] [Medline: 21723049]
37. Elwafi HM, Witkiewitz K, Mallik S, Thornhill TA, Brewer JA. Mindfulness training for smoking cessation: moderation of the relationship between craving and cigarette use. Drug Alcohol Depend 2013 Jun 01;130(1-3):222-229 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.015] [Medline: 23265088]
38. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
39. West CP, Dyrbye LN, Sloan JA, Shanafelt TD. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. J Gen Intern Med 2009 Dec;24(12):1318-1321 [FREE Full text] [doi: 10.1007/s11606-009-1129-z] [Medline: 19802645]

40. Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry 2016 Jun;15(2):103-111 [FREE Full text] [doi: 10.1002/wps.20311] [Medline: 27265691]
41. West CP, Dyrbye LN, Satele DV, Sloan JA, Shanafelt TD. Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. J Gen Intern Med 2012 Nov 24;27(11):1445-1452 [FREE Full text] [doi: 10.1007/s11606-012-2015-7] [Medline: 22362127]
42. Field A, Miles J, Field Z. Discovering Statistics Using R. Thousand Oaks, CA: Sage Publications; 2012.
43. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
44. Zhou J, Yang Y, Qiu X, Yang X, Pan H, Ban B, et al. Relationship between anxiety and burnout among Chinese physicians: a moderated mediation model. PLoS One 2016 Aug 1;11(8):e0157013. [doi: 10.1371/journal.pone.0157013] [Medline: 27479002]
45. Hoge EA, Bui E, Palitz SA, Schwarz NR, Owens ME, Johnston JM, et al. The effect of mindfulness meditation training on biological acute stress responses in generalized anxiety disorder. Psychiatry Res 2018 Apr;262:328-332 [FREE Full text] [doi: 10.1016/j.psychres.2017.01.006] [Medline: 28131433]
46. Hoge EA, Bui E, Goetter E, Robinaugh DJ, Ojserkis RA, Fresco DM, et al. Change in decentering mediates improvement in anxiety in mindfulness-based stress reduction for generalized anxiety disorder. Cognit Ther Res 2015 Apr;39(2):228-235 [FREE Full text] [doi: 10.1007/s10608-014-9646-4] [Medline: 28316355]

Abbreviations

ANOVA: analysis of variance
GAD-7: Generalized Anxiety Disorder-7
IQR: interquartile range
MBI: Maslach Burnout Inventory
MBSR: mindfulness-based stress reduction
MT: mindfulness training
UA: Unwinding Anxiety.

Edited by G Eysenbach; submitted 24.07.19; peer-reviewed by U Subnis, Z Huang; comments to author 25.09.19; revised version received 21.10.19; accepted 28.01.20; published 01.04.20.

Please cite as:
Roy A, Druker S, Hoge EA, Brewer JA
Physician Anxiety and Burnout: Symptom Correlates and a Prospective Pilot Study of App-Delivered Mindfulness Training
JMIR Mhealth Uhealth 2020;8(4):e15608
URL: https://mhealth.jmir.org/2020/4/e15608
doi:10.2196/15608
PMID:32234708

Original Paper

Efficacy of the Digital Therapeutic Mobile App BioBase to Reduce Stress and Improve Mental Well-Being Among University Students: Randomized Controlled Trial

Sonia Ponzo[1], BSc, MSc, PhD; Davide Morelli[1,2], BSc, MSc, PhD; Jamie M Kawadler[1], BA, MSc, PhD; Nicola Rose Hemmings[1], BSc, MSc; Geoffrey Bird[3,4], BSc, MA, PhD; David Plans[1,3,5], BA, MSc, PhD

[1]BioBeats Group Ltd, London, United Kingdom

[2]Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom

[3]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

[4]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]Initiative in the Digital Economy, Department of Science, Innovation, Technology, and Entrepreneurship, University of Exeter, Exeter, United Kingdom

Corresponding Author:
David Plans, BA, MSc, PhD
Initiative in the Digital Economy
Department of Science, Innovation, Technology, and Entrepreneurship
University of Exeter
Stocker Road
Exeter, EX4 4PY
United Kingdom
Phone: 44 7527016574
Email: D.Plans@exeter.ac.uk

Abstract

Background: University students in the United Kingdom are experiencing increasing levels of anxiety. A program designed to increase awareness of one's present levels of well-being and suggest personalized health behaviors may reduce anxiety and improve mental well-being in students. The efficacy of a digital version of such a program, providing biofeedback and therapeutic content based on personalized well-being metrics, is reported here.

Objective: The aim of this study was to test the efficacy and sustained effects of using a mobile app (BioBase) and paired wearable device (BioBeam), compared with a waitlist control group, on anxiety and well-being in university students with elevated levels of anxiety and stress.

Methods: The study employed a randomized, waitlist-controlled trial with assessments at baseline, 2 weeks, postintervention (4 weeks), and follow-up (6 weeks). Participants were eligible if they were current full-time undergraduate students and (1) at least 18 years of age, (2) scored >14 points on the Depression, Anxiety, and Stress Scale-21 items (DASS-21) stress subscale or >7 points on the DASS-21 anxiety subscale, (3) owned an iOS mobile phone, (4) did not have any previous psychiatric or neurological conditions, (6) were not pregnant at the time of testing, and (7) were able to read and understand English. Participants were encouraged to use BioBase daily and complete at least one course of therapeutic content. A P value ≤.05 was considered statistically significant.

Results: We found that a 4-week intervention with the BioBase program significantly reduced anxiety and increased perceived well-being, with sustained effects at a 2-week follow-up. Furthermore, a significant reduction in depression levels was found following the 4-week usage of BioBase.

Conclusions: This study shows the efficacy of a biofeedback digital intervention in reducing self-reported anxiety and increasing perceived well-being in UK university students. Results suggest that digital mental health interventions could constitute a novel approach to treat stress and anxiety in students, which could be combined or integrated with existing therapeutic pathways.

Trial Registration: Open Science Framework (OSF.io) 2zd45; https://osf.io/2zd45/

(JMIR Mhealth Uhealth 2020;8(4):e17767) doi:10.2196/17767

KEYWORDS

anxiety; depression; mobile apps; biofeedback; mental health; mobile phones; technology

Introduction

Background

Stress and anxiety in university students of the United Kingdom have been steadily rising in the past decade [1]. Research demonstrates that by the midpoint of their course, 9% of previously symptom-free students develop depression and 20% become anxious to clinically significant levels [2]. Nearly half (48%) of the students registered at a UK university-based general practice report high levels of anxiety [3]. Internationally, levels of students' anxiety are also increasing, with university counseling services experiencing increasingly higher demand since 2010 [4-7]. Longitudinal studies report that students experience higher stress on entering university, which continues to increase during their studies, and does not return to previous levels after graduation [8,9].

Strikingly, only 25% to 36% of students with mental health issues seek treatment [10-12], largely due to the perceived stigma associated with these conditions [11,13]. A study investigating self-reported barriers to help-seeking behaviors and engagement in therapeutic pathways in students at risk of suicide found that a lack of time and a preference for self-management were among the main factors contributing to students' choice not to seek treatment [14]. Untreated mental health issues among university students have been shown to have immediate and significant repercussions on the overall quality of life, increasing the likelihood of dropping out of university and committing suicide [1]. Importantly, untreated mental health issues during university years also have negative impact following graduation, affecting relationships, levels of productivity, and the likelihood of substance abuse [15].

Although on-site facilities are crucial for managing students' mental health, their underutilization [16] suggests that novel approaches are needed to overcome accessibility barriers. Studies calling for more timely and preventative therapeutic interventions have highlighted the need for digital interventions [17-19]. The use of digital interventions, such as internet-based self-help resources and mobile apps, have been on the rise in the past decade due to their increased accessibility, availability, and anonymity [20-24], as well as their cost-effectiveness [25]. Owing to the widespread use of mobile phones, mobile apps could constitute effective therapeutic support for periods when students are away from the university, as well as increasing the capacity of on-site counseling services [26,27]. Mobile apps, paired with biosensors and wearable devices, are also effective in gathering passive data (eg, physical activity [28]) and self-report measures (eg, mood journaling). Accordingly, apps are increasingly used as a real-time monitoring tool, with personalized feedback, insights, and therapeutic content offered to users within the context of mental health interventions [29] and illness prevention [30].

A number of these digital interventions have proven effective in treating a variety of mental health disorders, ranging from anxiety and depression, to substance use disorder [31]. For example, an intervention lasting 2 weeks comprising brief, daily conversations and mood tracking with a Cognitive Behavioral Therapy (CBT)-oriented conversational agent (Woebot) found that, in comparison with an information-based digital control group, those in the Woebot group significantly reduced their symptoms of depression, while participants in both groups showed significantly reduced levels of anxiety [32]. Furthermore, an 8-week intervention in US university students with the mobile-app Calm was found to produce a significantly greater degree of stress reduction than that seen in a waitlist control group [23]. Despite these promising results, studies investigating the efficacy of a combined intervention, including both passive data collection and active therapeutic content, are still lacking.

The app BioBase (BioBeats, Ltd) aims at increasing individuals' well-being by combining elements of mindfulness, biofeedback interventions (such as diaphragmatic breathing exercises), CBT, and behavioral activation theory [33-35]. Specifically, its psychoeducational content is based on the job demands-resources model, which has been shown to be associated with students' well-being and stress management [36]. Alongside therapeutic content, data on physical activity, sleep quality, and heart rate are collected via a wrist-worn wearable device (BioBeam) and made available to individuals using the app to foster an increased awareness of users' current well-being. Furthermore, available in-app tools include an ecological momentary assessment tool based on the Circumplex Model of Affect [37], allowing individuals to log their mood in the moment, and reflect back on their entries at a later date to gain insights into longer-term patterns of emotion. The app also includes diaphragmatic breathing exercises and relaxation techniques for in-the-moment stress reduction. In an initial feasibility study conducted with the BioBase app (BioBeats Ltd) in a sample of full-time employees [38], it was found that 4 weeks of usage of BioBase significantly reduced anxiety and increased self-reported mental well-being. The study also found that higher levels of baseline stress were associated with greater reductions in anxiety and increases in mental well-being, suggesting that usage of BioBase could be most beneficial for individuals with increased anxiety. However, the lack of a control group and the specificity of the selected population did not allow us to draw more general conclusions about the effects of using BioBase on self-reported anxiety and stress.

Objectives

Hence, the purpose of this study was to test the efficacy of a 4-week intervention delivered via a mobile app and wearable device (ie, the BioBase program) in comparison with a waitlist control group on anxiety and general mental well-being in university students with elevated anxiety or stress. The study also examined sustained effects (at 6 weeks from baseline) of the intervention on anxiety and well-being. Finally, in the current study, measures of depression were collected to investigate the impact of the BioBase program on depressive symptoms.

We hypothesized that university students in the intervention group, but not in the waitlist control, would have significant improvements in anxiety and well-being following a 4-week intervention with BioBase. We also predicted that anxiety and well-being would have sustained effects in the intervention group, but not in the waitlist control, at 2 weeks following the end of the intervention. Furthermore, it was hypothesized that being enrolled in the BioBase program would reduce depressive symptoms after 4 weeks of usage.

Methods

Ethics Approval

This study was approved by an Institutional Ethics Committee at the University of Exeter (UEBS Research Ethics Committee, ethics application number: eUEBS002252). All participants provided informed, electronic consent prior to their enrollment in the study. Data from this study, including the preregistration protocol, are available on the Open Science Framework website (see Trial Registration section).

Study Design

The current study was a randomized, waitlist control trial with assessments conducted at baseline, 2 weeks, postintervention (4 weeks), and follow-up (6 weeks). Participants randomly assigned to the intervention group took part in a 4-week well-being intervention (the BioBase program). Those assigned to the waitlist control group received the intervention after 8 weeks.

Recruitment

Participants were recruited using institutional participant pools at different UK universities as well as via social media, mailing lists, and flyers and through university staff. Recruitment took place between October and November 2019, and potential participants were screened for eligibility via a Qualtrics survey. Inclusion criteria comprised being a full-time university student attending a university in the United Kingdom and (1) being aged between 18 and 25 years, (2) having scored >14 points on the Depression, Anxiety and Stress Scale-21 items (DASS-21 [39]) stress subscale or >7 points on the DASS-21 anxiety subscale, (3) owning an iPhone 6 or above, (4) not having any previous psychiatric or neurological conditions, (5) not being pregnant at the time of testing, and (6) being able to read and understand English. Participants were also excluded if they were currently in therapy or were using counseling services. Individuals taking part in the initial screening survey were entered into a lottery to win a UK £50 (US $63.93) Amazon Voucher.

Randomization and Blinding

The original design was devised as a single-blind study; however, due to logistical reasons (ie, clarity of communications between the research team and participants) it was decided to unblind the design.

Eligible participants (n=262) were sent a reminder email prompting them to confirm their willingness to take part in the study. A total of 130 participants were randomly assigned to the intervention group and 132 participants were randomly assigned to the waitlist control group based on minimization factors: gender (2 categories: male and female), age (7 categories: 18, 19, 20, 21, 22, 23, 24, and 25 years), DASS-21 anxiety subscale (5 categories: normal, mild, moderate, severe, extremely severe), and DASS-21 stress subscale (5 categories: normal, mild, moderate, severe, extremely severe). DASS-21 categories were used for inclusion (ie, participants scoring within the normal range at screening were excluded from the study and those scoring normal at baseline were excluded from the analysis) and minimization purposes only. The first participant was allocated at random. Each subsequent participant's group membership was allocated such that their addition to that group would lead to a closer match between the groups according to the minimization factors at screening. The random number list used to create the 2 groups was generated using the R *Minirand* package. Following randomization, the intervention group received their BioBeams (which are not functional until paired with a registered BioBase account) via post at their selected address.

Intervention Group

After randomization, participants in the intervention group were emailed the first set of questionnaires to complete (Figure 1). At the end of the questionnaires, they were given details on how to download and register on the BioBase app.

The BioBase program is a multidimensional mobile app comprising psychoeducational content on mental health and well-being, mood tracking (via an ecological momentary assessment, EMA [40]), and in-the-moment exercises (eg, deep breathing and relaxation techniques). Furthermore, passive data on sleep, heart rate, and physical activity are collected via a wearable device (BioBeam) and presented to the users via a dashboard view.

The psychoeducational content is delivered via 42 five-min long modules, each tackling different aspects of psychological and emotional distress (see Multimedia Appendix 1 for a detailed description of the modules). The content is organized in three different courses, based on the job demands-resources model [41]. Each course relates to a different aspect of the model (ie, demands, control, and support) and it comprises 14 modules. Demands and control are widely recognized as relevant workplace stressors [42,43], while social support has been shown to positively impact perceived well-being [44]. Embedded in these modules are elements of CBT and self-compassion (see Multimedia Appendix 1). Digitally delivered CBT interventions have been proven efficacious in reducing the levels of anxiety and depression [45] and similarly, self-compassion has been shown to predict symptom severity in anxious and depressed individuals [46]. By incorporating these therapeutic elements, the courses aim to foster an individual's recognition of internal physiological and emotional processes as a trigger for stress and identify effective coping strategies (eg, setting achievable goals aligned with the individual's personal values).

The EMA tool allows an individual to report their mood in the moment by choosing a mood from a list of options, each with different valence (positive or negative) and arousal (high or low). Furthermore, individuals can specify any ecological

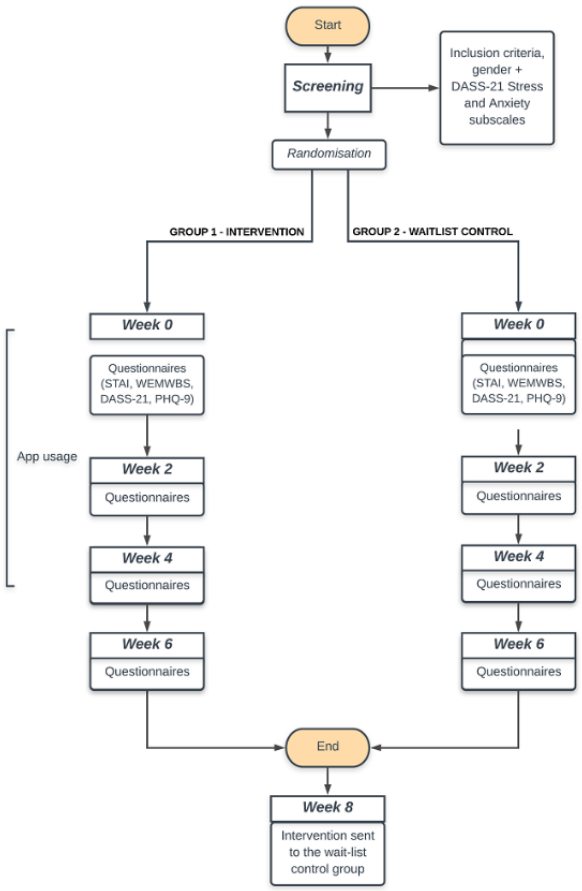
component surrounding the moment they chose to declare their mood (ie, where they were, whether they were alone or with somebody, and what activities they were engaged with). EMAs are a valuable mood-tracking tool in the context of digital interventions specifically aimed at reducing levels of anxiety and depression (see [47] for a review).

The deep-breathing tool is designed as a quick intervention aimed at reducing stress and increasing relaxation and consists of 10 guided deep diaphragmatic breaths. Respiration biofeedback has been shown to lead to a reduction of symptoms of depression and anxiety (see [48] for a review on results of biofeedback interventions). Similarly, the body scan has been devised as a standalone quick relaxation intervention due to its effectiveness in reducing anxiety and depression (for a detailed review, see [49]).

Finally, passive data collection on physical activity (ie, number of steps performed every 20 seconds), sleep duration and quality (via a triaxial accelerometer with a sample rate of 100 Hz), and heart rate (via a photoplethysmography sensors) was obtained via the BioBeam wearable. This information was made available to participants via an in-app dashboard. Increased sleep awareness and implementation of sleep hygiene techniques have been recognized as a mediating factor in anxiety [50], thus supporting the notion that insights into individuals' sleeping patterns may prove beneficial in stress reduction. Furthermore, as physical inactivity is associated with greater levels of anxiety [51], awareness of and insight into one's own activity patterns may foster improvements in individuals' well-being.

Participants were not prompted to use the app in any specific fashion and were left to freely engage with it for the whole intervention (ie, 4 weeks). Participants were, however, encouraged to continuously wear the BioBeam and engage with the therapeutic content (modules and tools) on a daily basis for at least 5 min. App usage was discontinued after the 4-week intervention ended.

Figure 1. Timeline of the study. Participants were screened for inclusion and exclusion criteria. Of the 805 participants screened, only 262 were eligible to take part. These selected participants were subsequently randomly assigned to one of two groups: intervention or wait-list control. Each group received the questionnaires at baseline, 2, 4 and 6 weeks. The intervention group received access to the program following the first questionnaire completion.



Waitlist Control Group

The waitlist control group received the baseline questionnaire at the same time as the intervention group, followed by an email stating that they would be provided with the app and the wearable device in 8 weeks. Throughout the 8 weeks during which the intervention group used BioBase, the waitlist control participants received the 2-, 4-, and 6-week questionnaires, preceded by a reminder email to complete them. After 8 weeks, participants received a BioBeam at their selected address as well as an email with instructions on how to download and register the app.

Measures and Incentives

Both groups completed four surveys via an online platform (Qualtrics). The surveys consisted of the following questionnaires: the State-Trait Anxiety Inventory (STAI-S-6 [52]), the Warwick-Edinburgh Mental Well-Being Scale

(WEMWBS [53]), the DASS-21**,** and the Patient Health Questionnaire (PHQ-9 [54]). The DASS-21 Stress and Anxiety Subscales were used as a screening tool for participants' inclusion in the study, whereas the depression subscale, together with the PHQ-9, was used as an outcome measure for depression. Demographic characteristics of the sample were collected at baseline. At the end of the study, each participant received a monetary incentive of £40 (£10 per each completed set of questionnaires at T0, T1, T2**,** and T3) plus an additional £5 if they decided to send back the wearable device received as part of the intervention.

Primary Outcome

The primary outcomes of the study were responses on the STAI [52]. The STAI-S-6 is a short version of the 10-item state subscale of the STAI. It is a 6-item scale, measuring state anxiety, with responses ranging from 1 (*Not at all*) to 4 (*Very much*). Scaled scores are obtained by multiplying the summed responses to each item by 20 and subsequently dividing the score by 6 (range 20-80).

Secondary Outcome

The secondary outcome of the current study was the WEMWBS [53], measuring perceived well-being. WEMWBS is a 14-item scale assessing subjective well-being and psychological functioning. Scoring is obtained by summing each response, ranging from 1 (*None of the time*) to 5 (*All of the time*) (range 14-70). WEMWBS has been validated for use in the United Kingdom with those aged 16 and above [53].

Additional Measures

Anxiety and stress were further measured via the DASS-21 subscales to ensure participants were still reporting elevated levels of stress or anxiety at baseline as well as during the screening procedure. Moreover, depression levels were investigated via the DASS-21 Depression subscale and the PHQ-9 questionnaire, a widely employed clinical tool. Although DASS-21 focuses on 1-week periods, PHQ-9 instructs individuals to report changes in the previous 2 weeks. Given that the focus on longer periods may mitigate the effects of random fluctuations in mood, both measures were collected.

Statistical Analysis

Power

A power analysis, based on a previous feasibility pilot study, was conducted to estimate the required sample size for the randomized controlled trial. Accounting for potential dropout, the estimated sample size was at least 200 participants (100 per group), providing .95 power to detect a large effect size of .96 with an alpha of .05 in a final sample of at least 55 participants per group.

Data Exclusion

Given that the inclusion criteria for the current study comprised indication of anxiety or stress (as indexed by DASS-21 Anxiety and Stress subscale scores), 15 participants from the intervention group and 16 from the waitlist control group who initially scored above the normal range at screening but who scored in the normal range at baseline (T0) were excluded from statistical analysis. Participants were further excluded from final analyses if they did not download or open the app during the 4-week intervention as well as if they did not complete all questionnaires.

Data Analysis

The current study employed a mixed design with a between-subjects variable (group) with 2 levels: intervention versus waitlist control and a within-subjects variable (time) with 4 different levels: baseline, 2 weeks, 4 weeks, and 6 weeks. Given the advantages of linear mixed models (LMMs) in dealing with lack of homogeneity of variance and incomplete data sets across time points [55], LMMs were used to analyze our primary and secondary outcomes. Specifically, group and time were the fixed effects and time/subjects were the nested random effects. Planned comparisons (paired-samples *t* tests) were conducted to explore the direction of significant interactions between group and time. Effect sizes for planned comparisons were calculated using Cohen *d* (pooled SD) to allow maximum comparability with previous research [56]. The *P* values reported later have not been corrected for multiple comparisons but remain significant if corrected. Data were analyzed and plotted using the *tydiverse*, *ggplot2* [57], *lmer4* [58], and *lmerTest* [59] packages for R.

Results

Participant Enrollment and Demographics

Figure 2 illustrates the flow of participants through the study and reasons for exclusion. Of 805 participants that were screened via an online questionnaire for inclusion and exclusion criteria, 262 participants were deemed eligible and were randomized into either the intervention (n=130) or waitlist control (n=132) groups. Of those, 59 participants from the intervention group and 64 from the control group completed the final questionnaire at T3 and were included in the analysis. Engagement data from the participants in the intervention group showed participants engaged with the app 21.9 of 29 days on average (median 26 days, IQR 13 days, range: 2-29 days). On average, participants engaged with the app 5.33 (SD=5.03) minutes per day (range: 2.13-28.68) over the 29 days of the intervention (see Figures 1 and 2, Multimedia Appendix 2). However, no correlation was found between the total amount of engagement with the app and differences, from baseline to T2 (4-weeks), in the main outcome measures (see Figures 3 and 4, Multimedia Appendix 3).

Participants in the two groups did not differ significantly with respect to age, gender, nor their levels of stress and anxiety at baseline (see Table 1). A total of 59 participants from the intervention group and 70 from the control group partook in the second questionnaire (intervention group: 38 females, age range: 18-25 years, mean 19.9 years, SD 1.9; waitlist control: 48 females, age range: 18-25 years, mean 19.9 years, SD 1.89; Figure 2), 55 and 61 (respectively) in the third (intervention group: 36 females, age range: 18-25 years, mean 19.9, SD 1.82; waitlist control: 43 females, age range: 18-25 years, mean 19.93 years, SD 1.95); and finally 59 and 64 participants completed the follow-up questionnaire (intervention group: 38 females,

age range: 18-25 years, mean 19.92 years, SD 1.86; waitlist control: 43 females, age range: 18-25 years, mean 20, SD 1.90).

Figure 2. Flowchart of enrollment and retention rates throughout the study in the intervention and wait-list control groups.

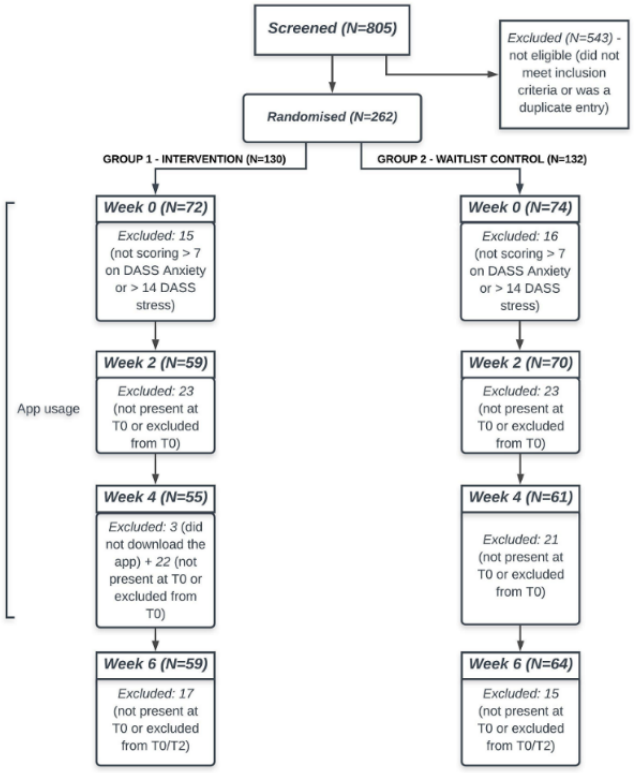


Figure 3. STAI scores at baseline, week 2, week 4 and week 6 follow-up from the start of the intervention in both intervention and wait-list control groups. Solid line=median; black dot=mean; whiskers: upper whisker=min(max(x), Q_3 + 1.5 x IQR); lower whisker=max(min(x), Q_1 - 1.5 x IQR). *P<.01.

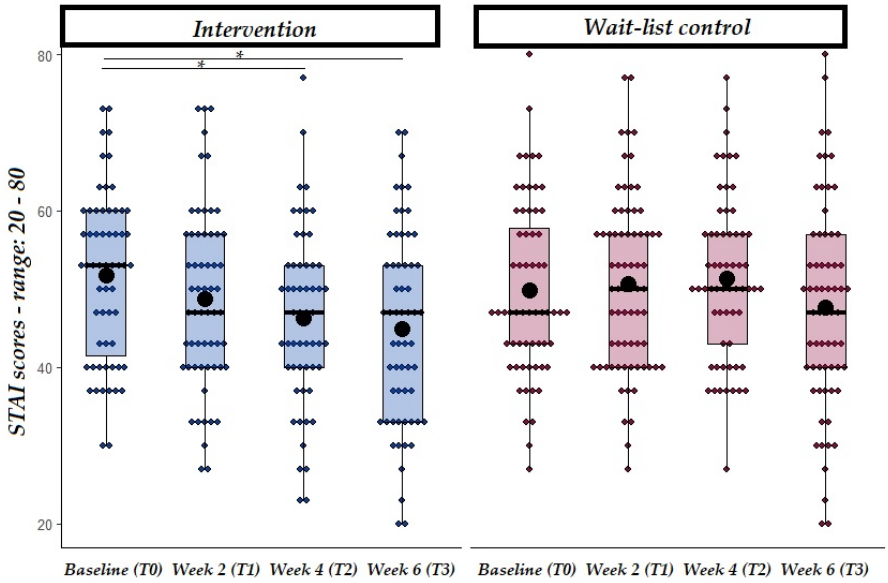


Figure 4. WEMWBS scores at baseline, week 2, week 4 and follow-up from the start of the intervention in both intervention and wait-list control groups. Solid line=median; black dot=mean; whiskers: upper whisker=min(max(x), Q_3 + 1.5 x IQR); lower whisker=max(min(x), Q_1 - 1.5 x IQR). *P<.05.

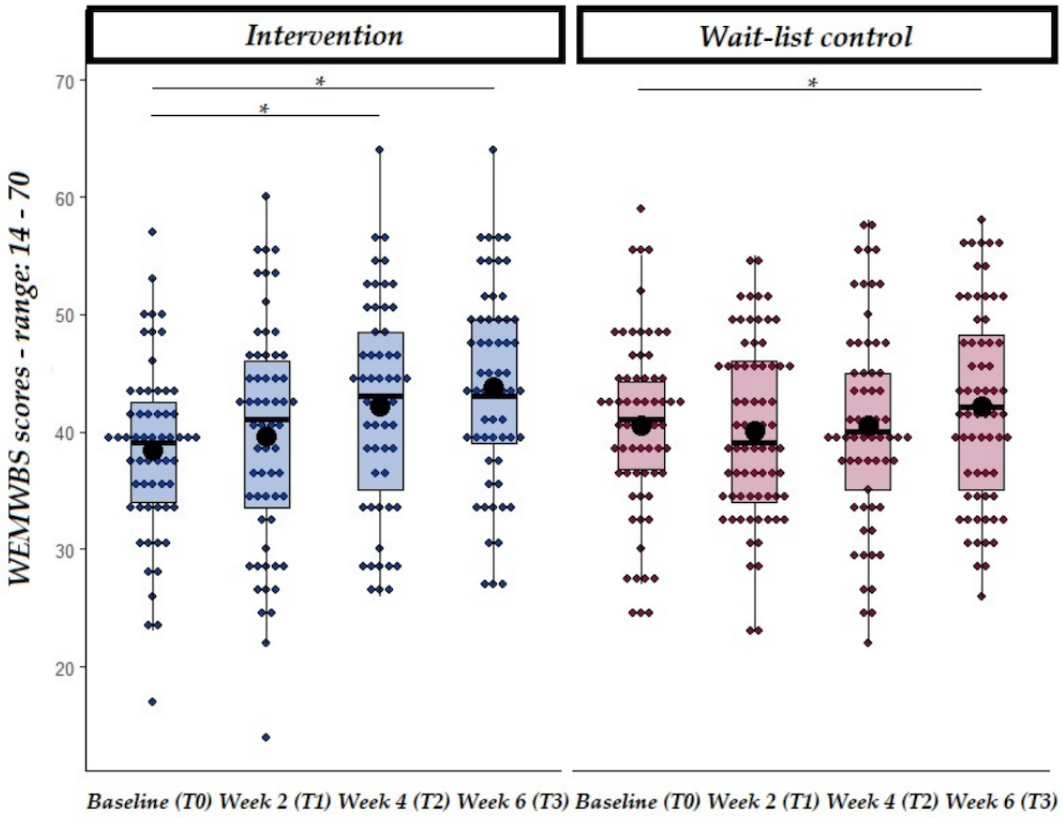


Table 1. Summary of participant characteristics at baseline (T0).

Characteristics	Intervention group (n=72)	Waitlist control (n=74)	*P* value
Gender (n)			
Females	45	47	N/A[a]
Males	27	27	N/A
Age (years), mean (SD)	19.9 (1.83)	19.84 (1.76)	.83
DASS-21[b] anxiety, mean (SD)	15.39 (6.68)	14.46 (7.23)	.42
DASS-21 stress, mean (SD)	21.08 (7.02)	19.86 (7.66)	.32

[a]N/A: not applicable.

[b]DASS-21: Depression, Anxiety, and Stress Scale-21 items.

Primary Outcomes

State-Trait Anxiety Inventory: Baseline to Follow-Up

The primary hypothesis was that, in comparison with the waitlist control group, the intervention group treated with BioBase would show a significant reduction in anxiety levels (measured via STAI-S-6) at the end of the intervention (ie, 4 weeks following baseline measures). Furthermore, it was hypothesized that such effects would be sustained at follow-up (ie, 2 weeks after the end of the intervention). An LMM with STAI-S-6 as the dependent variable and group and time (as well as their interaction) as independent variables was carried out. This analysis revealed a significant main effect of time (at both week 4 and week 6), with scores being lower in comparison with baseline. Furthermore, a significant interaction between group and time (at both week 4 and week 6) on perceived anxiety levels (see Table 2 for a summary of the LMMs) was observed.

To further explore the significant interaction between time and group, planned comparisons were conducted separately in the intervention and waitlist control groups comparing STAI-S-6 values at baseline with week 4 and follow-up (6 weeks), respectively. Findings revealed that STAI-S-6 at week 4 was significantly lower in the intervention group but not in the control group (see Table 3 for a summary of descriptive statistics and planned comparisons and Figure 3) and that such a reduction was still present at follow-up in the intervention group only.

One of our secondary hypotheses was that, in line with the results from Fitzpatrick and colleagues [32], the BioBase intervention would show efficacy in decreasing self-reported levels of anxiety after 2 weeks of treatment compared with the control group. However, no interaction between time and group was found at week 2, suggesting that changes in anxiety did not occur within the first 2 weeks of the intervention (Table 2).

Table 2. Summary of the linear-mixed model on State-Trait Anxiety Inventory scores over the 4 time points in the intervention and waitlist control groups.

Predictors	STAI-S-6[a]		
	Estimates	95% CI	*P* value
Intercept	53.61	47.07-60.15	<.001
Group	−1.90	−5.99 to 2.19	.36
T1—2 weeks	−6.35	−12.87 to 0.17	.06
T2—4 weeks	−12.25	−18.90 to −5.59	*<.001* [b]
T3—6 weeks	−11.32	−17.84 to −4.81	*.001*
Group: Time T1	3.44	−0.65 to 7.52	.10
Group: Time T2	7.09	2.94 to 11.24	*.001*
Group: Time T3	4.56	0.49 to 8.63	*.03*
Random effects			
σ^2	16.82	N/A[c]	N/A
τ00 Time: Participants ID	49.44	N/A	N/A
τ00 Participants ID	67.20	N/A	N/A
ICC[d]	0.87	N/A	N/A
N time	4	N/A	N/A
N ID	123	N/A	N/A
Observations	491	N/A	N/A
Marginal R^2/conditional R^2	0.037/0.879	N/A	N/A

[a]STAI: State-Trait Anxiety Inventory.

[b]Italicized values are significant.

[c]N/A: not applicable.

[d]ICC: intraclass correlation coefficient.

Table 3. Mean, SD, and planned comparisons on State-Trait Anxiety Inventory scores over the duration of the study (T0, T1, T2, and T3) in the intervention and waitlist control groups.

Time point	STAI-S-6[a]: planned comparisons							
	Intervention group				Waitlist control			
	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*
T0—baseline	51.71 (10.78)	N/A[b]	N/A	N/A	49.81 (10.96)	N/A	N/A	N/A
T1—week 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
T2—week 4	46.31 (11.32)	3.507 (54)	<.001	0.67	51.33 (10.35)	−1.449 (60)	.15	0.26
T3—week 6	44.95 (12.52)	4.35 (58)	<.001	0.81	47.61 (13.29)	1.542 (63)	.13	0.27

[a]STAI-S-6: State-Trait Anxiety Inventory.

[b]N/A: not applicable.

Secondary Outcomes

Warwick-Edinburgh Mental Well-Being Scale: Baseline (T0)—Follow-Up (T3)

One of the secondary hypotheses was that participants in the intervention group only would report higher levels of well-being (as measured by WEMWBS) at both the end of the intervention and at follow-up. An LMM with WEMWBS as the dependent variable and group and time (as well as their interaction) as the independent variables revealed a significant main effect of time (at both week 4 and week 6), suggesting that perceived well-being increased over time regardless of groups. Furthermore, a significant interaction between group and time (week 4 and week 6) was found (see Table 4), which was further analyzed with planned comparisons. *t* tests were conducted separately in the intervention and waitlist control group comparing WEMWBS values at baseline (T0) and following the 4-week intervention as well as at follow-up (week 6). Results showed that in the intervention group only, WEMWBS values significantly increased between baseline and week 4, suggesting a higher perceived well-being in the intervention group (see Table 5 and Figure 4). WEMWBS values significantly increased between baseline and follow-up in both groups, but with higher values on average in the intervention group, suggesting an increase in perceived well-being.

Table 4. Summary of the linear-mixed model on Warwick-Edinburgh Mental Well-Being Scale scores over the 4 time points in the intervention and waitlist control groups.

Predictors	WEMWBS[a]		
	Estimates	95% CI	*P* value
Intercept	36.40	31.65 to 41.16	*<.001* [b]
Group	2.07	−0.90 to 5.04	.17
T1—2 weeks	2.69	−1.25 to 6.64	.18
T2—4 weeks	7.41	3.38 to 11.44	*<.001*
T3—6 weeks	8.94	4.99 to 12.88	*<.001*
Group: Time T1	−1.57	−4.05 to 0.90	.21
Group: Time T2	−3.96	−6.47 to −1.45	*.002*
Group: Time T3	−3.65	−6.11 to −1.18	*.004*
Random effects			
σ^2	5.00	N/A[c]	N/A
τ00 Time: Participants ID	19.25	N/A	N/A
τ00 Participants ID	46.28	N/A	N/A
ICC[d]	0.93	N/A	N/A
N time	4	N/A	N/A
N ID	123	N/A	N/A
Observations	533	N/A	N/A
Marginal R^2/conditional R^2	0.029/0.794	N/A	N/A

[a]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.

[b]Values in italics are significant.

[c]N/A: not applicable.

[d]ICC: intraclass correlation coefficient.

Table 5. Mean, SD, and planned comparisons on Warwick-Edinburgh Mental Well-Being Scale scores over the duration of the study (T0, T1, T2, and T3) in the intervention and waitlist control groups.

Time point	WEMWBS[a]—planned comparisons							
	Intervention group				Waitlist control			
	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*
T0—Baseline	38.47 (7.54)	N/A[b]	N/A	N/A	40.55 (7.76)	N/A	N/A	N/A
T1—week 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
T2—week 4	42.15 (9.02)	−3.385 (54)	.001	0.65	40.51 (8.64)	0.814 (62)	.42	0.15
T3—week 6	47.76 (8.31)	−6.260 (58)	<.001	1.16	42.19 (8.37)	−2.127 (63)	.04	0.38

[a]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.

[b]N/A: not applicable.

Additional Measures

To explore the potential of the BioBase program to reduce depression over the 4-week period of use, depression was measured via the PHQ-9 questionnaire and a linear mixed model with depression scores as the dependent variable and group and time (as well as their interaction) as the independent variables was carried out. This analysis revealed that depressive symptoms decreased at 4-weeks from the start of the intervention, regardless of groups, but that in the intervention group this effect was more pronounced (as suggested by the significant interaction between group and time at week 4). This significant interaction (Table 6) was further analyzed via planned comparisons on depression scores in the intervention and waitlist control group at baseline and following the 4-weeks intervention. Findings revealed that in the intervention group only, PHQ-9 values significantly decreased between baseline and week 4, suggesting a lower perceived level of depression (see Table 7 and Figure 5). Changes in the Depression subscale of the DASS-21 were also explored. This analysis revealed a main effect of Time at both week 2 and week 4 (see Multimedia Appendix 4), with depression levels reducing over time irrespective of groups. Although the same pattern highlighted by the PHQ-9 scores was observed (intervention group: Baseline: mean 18.58, SD 10.87; week 4: mean 12.76, SD 8.77; waitlist control: baseline: mean 16.44, SD 9.67; week 4: mean 12.16, SD 8.90), there was no significant interaction between Group and Time. Such finding could be due to the intrinsic characteristics of the scales (ie, DASS-21 focuses on 1-week periods, while PHQ-9 asks individuals to report changes in the previous 2-weeks).

Table 6. Summary of the linear mixed model on Patient Health Questionnaire scores over the duration of the intervention (T0, T1, and T2) in the intervention and waitlist control groups.

Predictors	PHQ-9[a]		
	Estimates	95% CI	*P* value
Intercept	12.65	9.73-15.58	*<.001* [b]
Group	−0.87	−2.70 to 0.95	.35
T 1—2 weeks	−1.34	−3.53 to 0.84	.23
T 2—4 weeks	−4.91	−7.15 to −2.67	*<.001*
Group: Time T1	0.46	−0.90 to 1.83	.51
Group: Time T2	2.07	0.67 to 3.46	*.004*
Random effects			
σ^2	4.25	N/A[c]	N/A
τ00 Time:Participants ID	3.22	N/A	N/A
τ00 Participants ID	19.22	N/A	N/A
ICC[d]	0.84	N/A	N/A
N time	3	N/A	N/A
N ID	123	N/A	N/A
Observations	368	N/A	N/A
Marginal R^2/conditional R^2	0.026/0.845	N/A	N/A

[a]PHQ-9: Patient Health Questionnaire.

[b]Value in italics are significant.

[c]N/A: not applicable.

[d]ICC: intraclass correlation coefficient.

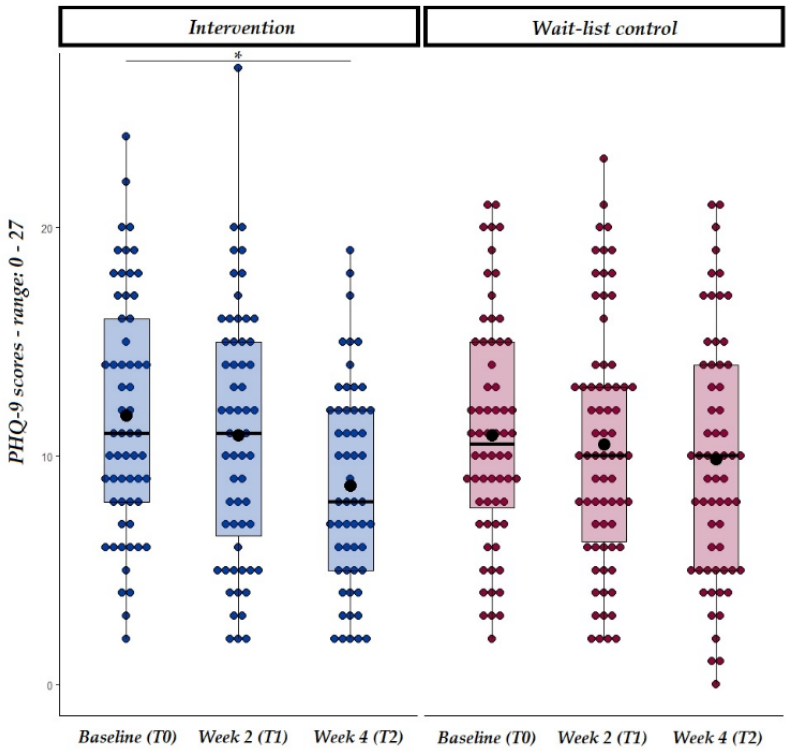
Table 7. Mean, SD, and planned comparisons on Patient Health Questionnaire scores over the duration of the intervention (T0, T1, and T2) in the intervention and waitlist control groups.

Time point	PHQ-9[a]—planned comparisons							
	Intervention group				Waitlist control			
	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*	Mean (SD)	*t* test (*df*)	*P* value	Effect size, *d*
T0—baseline	11.78 (5.2)	N/A[b]	N/A	N/A	10.91 (4.93)	N/A	N/A	N/A
T1—week 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
T2—week 4	8.71 (4.45)	5.139 (54)	<.001	0.99	9.85 (5.38)	1.392 (60)	.17	0.25

[a]PHQ-9: Patient Health Questionnaire.

[b]N/A: not applicable.

Figure 5. PHQ-9 scores at baseline and week 2 from the start of the intervention in both intervention and wait-list control groups. Solid line=median; black dot=mean; whiskers: upper whisker=min(max(x), Q_3 + 1.5 x IQR); lower whisker=max(min(x), Q_1 - 1.5 x IQR). *P<.01.



Discussion

Principal Findings

The aim of the current study was to investigate the efficacy of BioBase, a 4-week app-based intervention, in reducing anxiety and increasing well-being in university students with high self-reported levels of stress or anxiety. Results revealed that using the BioBase program for 4 weeks led to reduced self-reported anxiety and increased self-reported well-being. Such results were sustained at follow-up, with participants in the intervention group maintaining lower levels of self-reported anxiety and higher levels of well-being at 6 weeks from the study start date. Effect sizes ranged from moderate to large throughout the different outcomes.

Comparison With Prior Work

The primary hypotheses of the current study were that in the intervention group only, levels of anxiety would decrease following enrollment in the BioBase program and that this reduction would be sustained after 2 weeks from the end of the intervention. In line with our first primary hypothesis, we found that self-reported levels of anxiety were significantly reduced in the intervention group after 4 weeks of app usage. This finding is in line with results from previous studies using digital interventions in both student [23] and nonstudent [32] populations. As mentioned in the Introduction section, Huberty et al [23] found that the mobile app Calm, consisting of a guided mindfulness meditation program, was effective in reducing stress levels among university students. In contrast to the BioBase 4-week program of 5 min a day; however, the Calm intervention was an 8-week program, requiring participants to first complete a 1-week course and then actively engage with the therapeutic content for at least 10 min a day. The efficacy of the BioBase program despite the reduced *dosage* may be related to the nature of the BioBase program: the therapeutic content is only one aspect of the hypothesized factors at play in anxiety reduction. Interactions with the app dashboard (showing participants their levels of activity, sleep quality, mood declarations over time, and heart rate), as well as usage of the tools, are hypothesized to be causally efficacious in addition to the traditional therapeutic content. Future studies using BioBase could shed light on the individual contribution of each of these aspects in reducing anxiety levels.

These results are also in line with previous findings [38], suggesting a significant reduction in anxiety following a 4-week intervention with the BioBase program in a sample of full-time employees. However, in this previous study, the effect of the intervention was not assessed beyond the end of the program. In the current study, we showed that the effect of the intervention persisted for 2 weeks following the end of the program. This result, in line with previous research [23], highlights the efficacy of mobile apps to reduce stress and anxiety over time, and their potential to supplement existing therapeutic support [18,27,29,30]. Future studies should investigate the extent to which these effects persist over longer timeframes, with the aim of identifying optimal guidelines for engagement to maximize outcomes.

A secondary hypothesis was that reduction in anxiety would be present following 2 weeks of enrollment in the BioBase program in the intervention group (but not in the waitlist control group). However, we did not find evidence of efficacy at 2 weeks. This finding is in contrast with a previous study conducted in the young adult population [32], using a CBT-based intervention

to reduce anxiety and depression, which found significant results following 2-week long interactions with a Web-based conversational agent. Nevertheless, the current study significantly differed in both methods of delivery (app vs Web-based) as well as type of intervention. Although Fitzpatrick and colleagues employed a daily intervention, comprising specific time windows of interaction with the therapeutic content, the current study had a more ecological approach, with the BioBase program being available to participants at all times yet not being a daily commitment. Thus, the reason behind the lack of efficacy following a 2-week enrollment in the program may be due to differences in perceived benefit from the participants' perspective, that is, it may be easier to recognize the impact of a daily conversational intervention versus a natural, progressive engagement with a multidimensional program. Further research, comparing different kinds of interventions, would be needed to shed light on these findings.

In terms of secondary outcomes, it was hypothesized that perceived well-being would increase following a 4-week intervention with BioBase and that this effect would be sustained at follow-up (6 weeks). As predicted, we found that participants in the intervention group reported higher levels of perceived well-being after 4 weeks, which were still significant at 2 weeks from the end of the intervention. Nevertheless, we also found a main effect of time, with levels of perceived well-being being higher at T2 and T3, regardless of the grouping. Further studies with single- or double-blind designs could investigate the impact of being enrolled in a study on perceived well-being.

Finally, additional measures of depression were obtained via the PHQ-9 questionnaire and DASS-21 Depression subscale to assess the feasibility of the BioBase program in reducing depressive symptoms. Results showed that participants taking part in the current study reported lower depression levels after 4 weeks of BioBase usage and sustained effects at follow-up (as measured via the PHQ-9). Nevertheless, despite showing the same pattern of reduction, the same results were not significant for the DASS-21 Depression subscale. Such a discrepancy may be due to differences in sensitivity of the 2 measures, given the focus on periods of different length, and further research is needed to shed light on these findings. Furthermore, given that the trial was conducted in November 2019 through December 2019, it is possible that the reduction in DASS Depression scores observed in the waitlist control group could be due to changes in university work demands, such as coursework deadlines and exams, over this period.

This result is nonetheless particularly relevant when assessing the lack of engagement of individuals at risk of suicide with established pathways of support. Specifically, the possibility to access a digital mental health intervention, which could be efficacious in reducing depressive symptomatology could, represent a novel approach in students at risk of suicide [1,14]. Future studies should specifically investigate the efficacy of such intervention in a student population with individuals suffering from self-reported depressive symptoms.

Limitations

A limitation of the current study is the lack of a blinding procedure. As mentioned in the Methods section, the current study was an unblinded, randomized controlled trial, with participants in the control group being aware of the fact that they were not currently partaking in the intervention. This was a consequence of the type of control group employed. However, both groups received the same kind of communications and were prompted to respond to the questionnaires in the same way. A targeted standardized email was sent every week, with the timeline of the study and key dates as a reminder to participants. Although these measures reduced the possibility that unblinding could influence the results of the current study, future studies should investigate the extent to which being enrolled in an intervention leads to improvements in anxiety and well-being by employing a single-blind design, with an information-based control group.

Moreover, owing to lack of data on ethnicity, or information on the characteristics of the students underusing mental health services, it was not possible to assess the generalizability of our sample. Further studies should further investigate this, by replicating the current study while controlling for these variables.

Furthermore, in the current study, it was not possible to differentiate the effect of the different components of the BioBase program. Although this is a characteristic of digital interventions [60], future studies should explore what components of the BioBase program are most efficacious for which individuals.

In addition, the current study targeted subclinical levels of anxiety; therefore, participants with a psychiatric diagnosis of anxiety were excluded. This decision was made to explore symptom reduction and well-being increase without the confounding factors of being currently in treatment for anxiety. It could be the case, however, that effect sizes were underestimated if BioBase is more efficacious in participants with higher anxiety levels. Further research is needed to better understand the potential effects of BioBase in individuals with a clinical diagnosis of anxiety or stress.

In terms of the follow-up measure, the current study employed a 6-weeks follow-up, aimed at investigating the sustained effects of the intervention. However, it should be noted that further research is needed to explore long-lasting effects of the intervention (eg, 8 weeks).

Finally, in the current study, no specific criterion was used with regard to app usage. Given that we wanted to observe how participants would naturally engage and interact with the program, there was no strict indication nor control on participants' way to use the app. Nevertheless, the majority of the sample engaged with the intervention, with only 3 people not downloading or installing the app. Future research could explore whether a more controlled intervention, with specific engagement criteria, could lead to more efficacious results while still maintaining ecological validity.

Conclusions

In this study, we showed that a 4-week digital intervention was efficacious in reducing anxiety and increasing well-being in a student population with high levels of self-reported stress and anxiety. These effects were sustained after 2 weeks from the

end of the intervention, thus suggesting prolonged efficacy over time. To the best of our knowledge, this is the first study showing the efficacy of a multidimensional digital program, comprising therapeutic content, biofeedback, and mood-journaling, in reducing anxiety and increasing well-being in a student population. These findings are particularly relevant given the documented preference of students to self-help, rather than accessing on-site facilities, when facing mental health issues. Furthermore, the common use of mobile phones makes this type of intervention both accessible and scalable for higher education institutions who aim to extend the support provided to their students [27]. Future research should investigate the feasibility of including digital mental health interventions in the existing therapeutic pathways, thus encouraging preventative as well as intervention-driven approaches to mental health, tailored to the needs of the individuals.

Acknowledgments

The authors thank the students who took part in the study and Soukaina Amira, a research assistant who helped with the logistics of the study.

Conflicts of Interest

DP was the CEO of BioBeats, the provider of the BioBase program. DM was the CTO of BioBeats, the provider of the BioBase program. SP, NH, JK, and GB were employees of BioBeats, the provider of the BioBase program. The company owners were not involved in the analysis, which was conducted by SP and reviewed by JK.

Multimedia Appendix 1
Content and theoretical contextualization of the three psycho-educational courses contained in BioBase.
[DOCX File , 19 KB - mhealth_v8i4e17767_app1.docx]

Multimedia Appendix 2
Summary of the Linear Mixed Model on DASS-21 Depression scores over the duration of the intervention (T0, T1 and T2) in the intervention and wait-list control groups.
[DOCX File , 16 KB - mhealth_v8i4e17767_app2.docx]

Multimedia Appendix 3
App Engagement figures.
[DOCX File , 159 KB - mhealth_v8i4e17767_app3.docx]

Multimedia Appendix 4
Engagement time and differences in anxiety and wellbeing.
[DOCX File , 56 KB - mhealth_v8i4e17767_app4.docx]

Multimedia Appendix 5
CONSORT EHEALTH checklist (v 1.6.1).
[PDF File (Adobe PDF File), 1429 KB - mhealth_v8i4e17767_app5.pdf]

References

1. Thorley C. Institute for Public Policy Research. 2017 Sep. Not By Degrees: Improving Student Mental Health in the UK's Universities URL: https://www.ippr.org/files/2017-09/1504645674_not-by-degrees-170905.pdf [accessed 2020-03-09]
2. Andrews B, Wilding JM. The relation of depression and anxiety to life-stress and achievement in students. Br J Psychol 2004 Nov;95(Pt 4):509-521. [doi: 10.1348/0007126042369802] [Medline: 15527535]
3. Edwards DJ, Rainey E, Boukouvala V, Wells Y, Bennett P, Tree J, et al. Novel ACT-based eHealth psychoeducational intervention for students with mental distress: a study protocol for a mixed-methodology pilot trial. BMJ Open 2019 Jul 16;9(7):e029411 [FREE Full text] [doi: 10.1136/bmjopen-2019-029411] [Medline: 31315873]
4. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J Affect Disord 2015 Mar 1;173:90-96. [doi: 10.1016/j.jad.2014.10.054] [Medline: 25462401]
5. New J. Inside Higher Ed. 2017 Jan 13. Colleges Struggle to Provide Ongoing Treatment as Demands for Mental Health Services Increases URL: https://www.insidehighered.com/news/2017/01/13/colleges-struggle-provide-ongoing-treatment-demands-mental-health-services-increases [accessed 2019-10-14]
6. Novotney A. American Psychological Association (APA). 2014 Sep. Students Under Pressure URL: https://www.apa.org/monitor/2014/09/cover-pressure [accessed 2019-10-14]

7. Pierceall EA, Keim MC. Stress and coping strategies among community college students. Community Coll J Res Pract 2007;31(9):703-712. [doi: 10.1080/10668920600866579]
8. Bewick B, Koutsopoulou G, Miles J, Slaa E, Barkham M. Changes in undergraduate students' psychological well‐being as they progress through university. Stud High Educ 2010;35(6):633-645. [doi: 10.1080/03075070903216643]
9. Newbury-Birch D, Lowry RJ, Kamali F. The changing patterns of drinking, illicit drug use, stress, anxiety and depression in dental students in a UK dental school: a longitudinal study. Br Dent J 2002 Jun 15;192(11):646-649 [FREE Full text] [doi: 10.1038/sj.bdj.4801448] [Medline: 12108944]
10. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
11. Eisenberg D, Downs MF, Golberstein E, Zivin K. Stigma and help seeking for mental health among college students. Med Care Res Rev 2009 Oct;66(5):522-541. [doi: 10.1177/1077558709335173] [Medline: 19454625]
12. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu S, et al. Mental health of college students and their non-college-attending peers: results from the National Epidemiologic Study on Alcohol and Related Conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
13. Cardwell JM, Lewis EG, Smith KC, Holt ER, Baillie S, Allister R, et al. A cross-sectional study of mental health in UK veterinary undergraduates. Vet Rec 2013 Sep 21;173(11):266. [doi: 10.1136/vr.101390] [Medline: 23956162]
14. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. J Am Coll Health 2013;61(7):398-406 [FREE Full text] [doi: 10.1080/07448481.2013.820731] [Medline: 24010494]
15. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
16. Marsh C, Wilcoxon S. Underutilization of mental health services among college students: an examination of system-related barriers. J College Stud Psychother 2015;29(3):227-243. [doi: 10.1080/87568225.2015.1045783]
17. Randall E, Bewick B. Exploration of counsellors' perceptions of the redesigned service pathways: a qualitative study of a UK university student counselling service. Br J Guid Counc 2016;44(1):86-98. [doi: 10.1080/03069885.2015.1017801]
18. Lungu A, Sun M. Time for a change: college students' preference for technology-mediated versus face-to-face help for emotional distress. Telemed J E Health 2016 Dec;22(12):991-1000. [doi: 10.1089/tmj.2015.0214] [Medline: 27223266]
19. Prentice JL, Dobson KS. A review of the risks and benefits associated with mobile phone applications for psychological interventions. Can Psychol 2014;55(4):282-290. [doi: 10.1037/a0038113]
20. Kählke F, Berger T, Schulz A, Baumeister H, Berking M, Auerbach RP, et al. Efficacy of an unguided internet-based self-help intervention for social anxiety disorder in university students: A randomized controlled trial. Int J Methods Psychiatr Res 2019 Jun;28(2):e1766 [FREE Full text] [doi: 10.1002/mpr.1766] [Medline: 30687986]
21. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing Web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
22. Donker T, Batterham PJ, Warmerdam L, Bennett K, Bennett A, Cuijpers P, et al. Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression. J Affect Disord 2013 Oct;151(1):343-351. [doi: 10.1016/j.jad.2013.06.020] [Medline: 23953024]
23. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app 'Calm' to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
24. Mowbray C, Megivern D, Mandiberg JM, Strauss S, Stein CH, Collins K, et al. Campus mental health services: recommendations for change. Am J Orthopsychiatry 2006 Apr;76(2):226-237. [doi: 10.1037/0002-9432.76.2.226] [Medline: 16719642]
25. Herrman H, Saxena S, Moodie R. Promoting Mental Health: Concepts, Emerging Evidence, Practice. Geneva: World Health Organisation; 2004.
26. Brown P. The Higher Education Policy Institute. 2016 Sep 22. The Invisible Problem? Improving Students' Mental Health URL: https://www.hepi.ac.uk/2016/09/22/3592/ [accessed 2019-10-14]
27. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
28. He Y, Li Y. Physical activity recognition utilizing the built-in kinematic sensors of a smartphone. Int J Distrib Sens Netw 2013;9(4):481580. [doi: 10.1155/2013/481580]
29. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, 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]
30. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]

31. Waltz E. Pear approval signals FDA readiness for digital treatments. Nat Biotechnol 2018 Jun 6;36(6):481-482. [doi: 10.1038/nbt0618-481] [Medline: 29874220]
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Larsson A, Hooper N, Osborne LA, Bennett P, McHugh L. Using brief cognitive restructuring and cognitive defusion techniques to cope with negative thoughts. Behav Modif 2016 May;40(3):452-482. [doi: 10.1177/0145445515621488] [Medline: 26685210]
34. Marlow SL, Lacerenza CN, Paoletti J, Burke CS, Salas E. Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. Organ Behav Hum Decis Process 2018 Jan;144:145-170. [doi: 10.1016/j.obhdp.2017.08.001]
35. Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K. Oxford Guide to Low Intensity CBT Interventions. UK: OUP Oxford; 2010.
36. Barr T, Sessa V, Sumner K, Bragger J. Applying the job demands-resources model towards understanding student stress. Comput Hum Behav 2015;77:317-325. [doi: 10.13140/RG.2.1.4981.4885]
37. Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 2005;17(3):715-734 [FREE Full text] [doi: 10.1017/S0954579405050340] [Medline: 16262989]
38. Kawadler J, Hemmings N, Ponzo S, Morelli D, Bird G, Plans D. Effectiveness of a smart phone application BioBase for reducing anxiety and increasing mental wellbeing: a pilot feasibility and acceptability study. JMIR Preprints 2020 preprint(forthcoming) [FREE Full text] [doi: 10.2196/preprints.18067]
39. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Australia: Psychology Foundation of Australia; 1996.
40. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
41. Bakker AB, Demerouti E. The Job Demands ‐ Resources model: state of the art. J Manag Psych 2007;22(3):309-328. [doi: 10.1108/02683940710733115]
42. Karasek RA. Job demands, job decision latitude, and mental strain: implications for job redesign. Adm Sci Q 1979 Jun;24(2):285-308. [doi: 10.2307/2392498]
43. Häusser JA, Mojzisch A, Niesel M, Schulz-Hardt S. Ten years on: A review of recent research on the Job Demand–Control (-Support) model and psychological well-being. Work Stress 2010 Jan;24(1):1-35. [doi: 10.1080/02678371003683747]
44. Haslam C, Jetten J, Cruwys T, Dingle G, Haslam S. The New Psychology of Health: Unlocking the Social Cure. London: Routledge; 2018.
45. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:46-54. [doi: 10.1192/bjp.185.1.46] [Medline: 15231555]
46. van Dam NT, Sheppard SC, Forsyth JP, Earleywine M. Self-compassion is a better predictor than mindfulness of symptom severity and quality of life in mixed anxiety and depression. J Anxiety Disord 2011 Jan;25(1):123-130. [doi: 10.1016/j.janxdis.2010.08.011] [Medline: 20832990]
47. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
48. Schoenberg PL, David AS. Biofeedback for psychiatric disorders: a systematic review. Appl Psychophysiol Biofeedback 2014 Jun;39(2):109-135. [doi: 10.1007/s10484-014-9246-9] [Medline: 24806535]
49. Blanck P, Perleth S, Heidenreich T, Kröger P, Ditzen B, Bents H, et al. Effects of mindfulness exercises as stand-alone intervention on symptoms of anxiety and depression: Systematic review and meta-analysis. Behav Res Ther 2018 Mar;102:25-35. [doi: 10.1016/j.brat.2017.12.002] [Medline: 29291584]
50. Chou T, Chang L, Chung M. The mediating and moderating effects of sleep hygiene practice on anxiety and insomnia in hospital nurses. Int J Nurs Pract 2015 May;21(Suppl 2):9-18. [doi: 10.1111/ijn.12164] [Medline: 26125570]
51. Stubbs B, Koyanagi A, Hallgren M, Firth J, Richards J, Schuch F, et al. Physical activity and anxiety: A perspective from the World Health Survey. J Affect Disord 2017 Jan 15;208:545-552. [doi: 10.1016/j.jad.2016.10.028] [Medline: 27802893]
52. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
53. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
54. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

55. Hesselmann G. Applying linear mixed effects models (LMMs) in within-participant designs with subjective trial-based assessments of awareness-a caveat. Front Psychol 2018;9:788 [FREE Full text] [doi: 10.3389/fpsyg.2018.00788] [Medline: 29887820]
56. Field A. Discovering Statistics Using IBM SPSS Statistics. London, UK: Sage Publications Ltd; 2017.
57. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. Cham: Springer; 2016.
58. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
59. Kuznetsova A, Brockhoff P, Christensen R. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Soft 2017;82(13) [FREE Full text] [doi: 10.18637/jss.v082.i13]
60. Whitton AE, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]

Abbreviations

CBT: cognitive behavioral therapy
DASS-21: Depression, Anxiety and Stress Scale-21 items
EMA: ecological momentary assessment
LMM: linear mixed model
PHQ: Patient Health Questionnaire
STAI: State-Trait Anxiety Inventory
WEMWBS: Warwick-Edinburgh Mental Well-Being Scale

Edited by C Dias, G Eysenbach; submitted 11.01.20; peer-reviewed by R Brewer, K Van Mens, A Daros; comments to author 31.01.20; revised version received 20.02.20; accepted 21.02.20; published 06.04.20.

Please cite as:
Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D
Efficacy of the Digital Therapeutic Mobile App BioBase to Reduce Stress and Improve Mental Well-Being Among University Students: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e17767
URL: https://mhealth.jmir.org/2020/4/e17767
doi:10.2196/17767
PMID:31926063

Original Paper

Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App

Tyler M Yasaka[1], BS; Brandon M Lehrich[2], BS; Ronald Sahyouni[2,3], MS, PhD

[1]Department of Otolaryngology – Head and Neck Surgery, University of California, Irvine, Irvine, CA, United States

[2]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA, United States

[3]Medical Scientist Training Program, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Tyler M Yasaka, BS
Department of Otolaryngology – Head and Neck Surgery
University of California, Irvine
1001 Health Sciences Rd
Irvine, CA, 92617
United States
Phone: 1 949 824 5011
Email: yasaka.tyler@gmail.com

Related Article:

Comment in: https://mhealth.jmir.org/2020/4/e19359/

Abstract

Background: The novel coronavirus disease 2019 (COVID-19) pandemic is an urgent public health crisis, with epidemiologic models predicting severe consequences, including high death rates, if the virus is permitted to run its course without any intervention or response. Contact tracing using smartphone technology is a powerful tool that may be employed to limit disease transmission during an epidemic or pandemic; yet, contact tracing apps present significant privacy concerns regarding the collection of personal data such as location.

Objective: The aim of this study is to develop an effective contact tracing smartphone app that respects user privacy by not collecting location information or other personal data.

Methods: We propose the use of an anonymized graph of interpersonal interactions to conduct a novel form of contact tracing and have developed a proof-of-concept smartphone app that implements this approach. Additionally, we developed a computer simulation model that demonstrates the impact of our proposal on epidemic or pandemic outbreak trajectories across multiple rates of adoption.

Results: Our proof-of-concept smartphone app allows users to create "checkpoints" for contact tracing, check their risk level based on their past interactions, and anonymously self-report a positive status to their peer network. Our simulation results suggest that higher adoption rates of such an app may result in a better controlled epidemic or pandemic outbreak.

Conclusions: Our proposed smartphone-based contact tracing method presents a novel solution that preserves privacy while demonstrating the potential to suppress an epidemic or pandemic outbreak. This app could potentially be applied to the current COVID-19 pandemic as well as other epidemics or pandemics in the future to achieve a middle ground between drastic isolation measures and unmitigated disease spread.

(JMIR Mhealth Uhealth 2020;8(4):e18936) doi:10.2196/18936

KEYWORDS

COVID-19; smartphone; privacy; contact tracing; peer-to-peer; epidemic; personal data; mobile phone; coronavirus; pandemic

Introduction

The novel coronavirus disease 2019 (COVID-19) global pandemic represents an urgent public health crisis. According to epidemiologic modeling conducted by Ferguson et al [1], both the United States and the United Kingdom face a dilemma in terms of choosing a public health response. On one hand, the models predict severe consequences, including high death rates, if the virus is permitted to run its course without any intervention or response. However, the authors conclude that an optimal outcome following a strategy of disease suppression would likely require dramatic alterations to daily life, including social distancing for the entire population until a vaccine is available. Such an intervention may result in significant economic loss. Ferguson and colleagues [1] noted that technological solutions such as a contact tracing smartphone app may provide alternatives to the drastic measures they proposed.

Contact tracing is the process of tracing potential transmission routes of an infection through a population for the purposes of isolating those who may have been exposed and reducing further transmission. Contact tracing in varying forms has been used for several diseases including tuberculosis, HIV, and Ebola infection [2,3]. Smartphone-based contact tracing presents a viable solution to limiting disease transmission; however, such an app presents significant concerns regarding privacy.

Recent events such as the Equifax security breach and the Cambridge Analytica controversy regarding data collection through Facebook have highlighted the privacy concerns regarding the use of personal data. These concerns are especially pertinent in a health care setting [4-6]. Existing contact tracing apps typically rely on the collection of personal data such as timestamped locations to determine exposure risk [3]. Location data are highly personal, and the privacy concerns detailed above are especially salient for location data [7]. Furthermore, location is only a proxy for contact, and inferences about exposure based on location may not always be accurate due to noise in the data [8].

The balance between privacy and other objectives is certainly controversial, but there is an additional concern in the case of a contact tracing app. As we will demonstrate in this study, the efficacy of a contact tracing app depends on its adoption rate. Even those that may not prioritize privacy should be concerned about the efficacy of a contact tracing app that is believed by many to be an invasion of privacy. If a sufficiently large portion of a population does not participate due to privacy concerns, such an intervention may have limited impact on the outcome of a pandemic.

We propose a novel method for contact tracing using a smartphone app without the use of location data. The objective of this app is to provide an effective contact tracing mechanism without compromising user privacy.

Methods

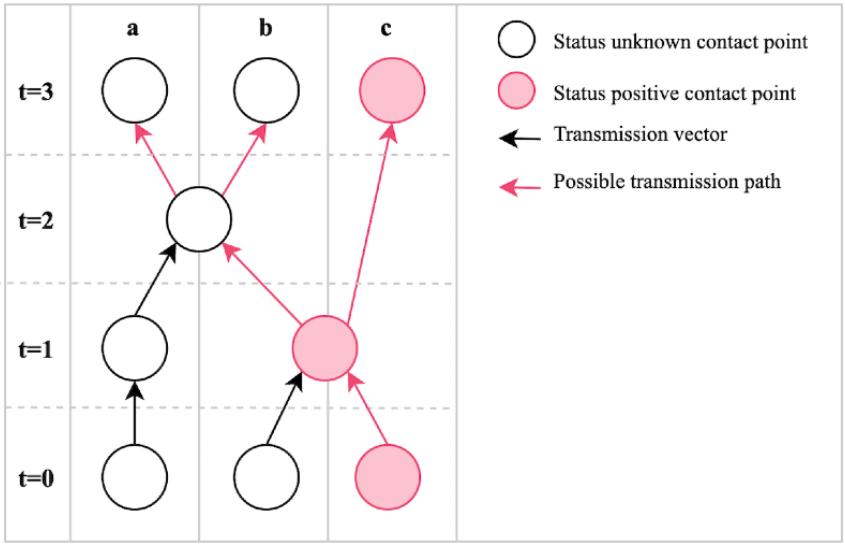
Novel Proposal for Tracing Possible Routes of Transmission

At the core of our approach is a data structure, which we will call the *transmission graph*. The transmission graph consists of nodes, which represent *contact points* between individuals, and directed edges, which represent *transmission vectors* between contact points. Whenever an individual participates in a contact point, a transmission vector is added to the transmission graph from the individual's prior contact point to the current contact point. The transmission graph, then, is a network of interactions between individuals. In this graph, each node (contact point) represents a physical interaction between two or more individuals at a specific time and place, during which microbial agents could potentially be transmitted from one individual to others. Of note, there are no entities in the graph that represent an individual, and location information is never encoded in this data structure. These properties of the transmission graph are fundamental to its privacy-preserving nature.

Each contact point (node) in the transmission graph can be in one of two states: *status positive* or *status unknown*. A status positive contact point is one that has been flagged as having one or more participating individuals that were positive for infection at the time of interaction. A status unknown contact point is simply one that has not been marked status positive; it is unknown whether any participating individuals were positive for disease at the time of the interaction. Status positive and status unknown are merely internal states of the graph. As we will demonstrate, they are distinct from the actual states that will be displayed to the end users.

Using the simple data structure of the transmission graph, *possible transmission paths* can be determined for any given target contact point. A possible transmission path is defined as a path from a status positive node to a given target node. In other words, a transmission path is a sequence of transmission vectors that could be carrying microbial agents from a reported point of exposure. For any given target node, there may be 0, 1, or multiple possible transmission paths. An illustration of a simple transmission graph is provided in Figure 1.

Figure 1. A network of interactions over time represented as a transmission graph. The rows represent units of time, and the columns represent individuals. By time t=3, all individuals have contact points with possible transmission paths. t: time point.



Smartphone App

We have developed a proof-of-concept smartphone app using the transmission graph data structure to implement peer-to-peer contact tracing. The app code is open source and publicly available on GitHub [9].

In the current prototype, users are assigned one of two risk levels: standard or elevated. The user is considered to have an elevated risk level if they have any recent contact points with possible transmission paths, and a standard risk level otherwise. “Recent” refers to some predetermined time, during which individuals who may have been exposed should take extra precautions, which we will call the *safety period*. Based on the protocol recommended by the World Health Organization (WHO) for quarantining individuals who have been in contact with a patient with confirmed COVID-19, we have set this time to 14 days in the current version of the app [10].

Possible transmission paths can also be assigned a *maximum length*, limiting the extent that a single reported diagnosis can raise the risk level of other users. A maximum length of 1, for example, would mean that only individuals that directly interacted with an infected individual would be assigned elevated risk levels. In the current app, the maximum length for transmission paths is set to 3.

When a user receives and reports a diagnosis, it is assumed that the infection was present for some time prior to the diagnosis. The app will determine the earliest contact point that occurred before this time and report it to the server accordingly. The amount of time that a user is assumed to have been infected is another parameter (*diagnosis delay*) that can be adjusted according to data and expert opinion. It is currently set to 2 days in accordance with the abovementioned WHO protocol.

Simulation Model

Rigorous mathematical models of contact tracing have been previously described in the literature, and we did not attempt to replicate them here [11,12]. Rather, we developed a low-fidelity computer simulation model that facilitates disease spread through interaction of individuals at contact points across time, allowing for the explicit modeling of the transmission graph structure we have proposed here. With this model of disease spread, we can compare outbreak trajectories both with and without peer-to-peer contact tracing as we have described. Such a model, while not intended to describe real-world trajectories, allows for the demonstration of the feasibility of our proposal and provides a rudimentary mechanism to compare various scenarios and app parameters, such as the adoption rate, the diagnosis delay estimated by the app, and the safety period used by the app. The model was written in the R (R Foundation for Statistical Computing) programming language, and the source code is publicly accessible on GitHub [9]. Additionally, a public, web-based interface for the model is provided [13].

The model is based on the susceptible, infected, and recovered epidemiological model, [14,15] where each individual is considered to be in one of the *susceptible*, *infected*, or *recovered* states. However, unlike many models, the nodes in our graph are *contact points*, not individuals. This format explicitly captures the spread of disease across time, and contact points can be arranged in layers to visualize spread across time, analogous to the illustration in Figure 1. Individuals may move to a new contact point at each unit of time, forming a directed edge from the prior to the current contact point. Individuals may also refrain from being at a contact point at any point in time, to model home isolation.

The time between infection and awareness of infection (*diagnosis delay*) is a parameter of the model. Similarly, individuals are in the *infected* state for a time period according to the *infectious period* model parameter. Individuals may transmit disease to each other at the same contact location in a point in time according to the *transmission rate* parameter.

When contact tracing is enabled, a separate graph is generated, containing only the information that would be available to individuals using the peer-to-peer contact tracing app. This graph is the transmission graph we have described. At the beginning of the simulation, each individual is designated as either adopting the app (*participating*) or not adopting the app

(*abstaining*), according to some probability specified by a model parameter (the *adoption rate*). Individuals that have adopted the app will perform three additional actions, in addition to the standard model behavior: (1) Participating individuals will log their diagnosis of infection onto the transmission graph, at their earliest contact point within the *estimated diagnosis delay* (a model parameter); (2) these individuals will log all their contact points and movements between contact points onto the graph; and (3) participating individuals will self-isolate if a path between a status positive contact point and a recent contact point exists, where "recent" is defined as those contact points within the time frame specified by the *safety period* parameter.

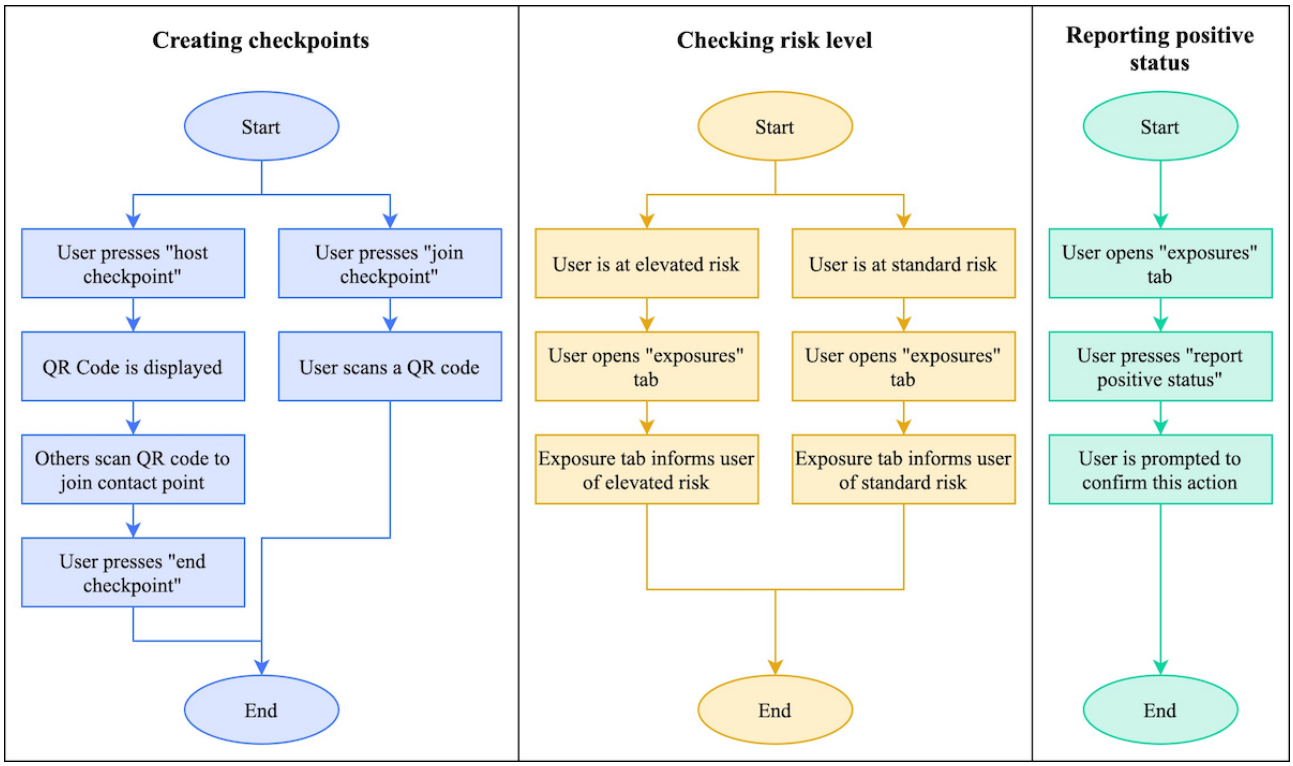
Thus, our model allows for the comparison of simulations both with and without use of a contact tracing app, as well as between different adoption rates in the population.

Results

Smartphone App

In the current app prototype, there are three primary user flows: creating contact points (termed "checkpoints" in the user interface), checking risk level, and reporting positive status. A diagram of these user flows is provided in Figure 2.

Figure 2. User flow diagrams for the three primary user flows in the peer-to-peer contact tracing app: creating checkpoints, checking risk, and reporting positive status. QR: Quick Response.



To create a checkpoint, a user may either host a new checkpoint or join an existing one. The host of a checkpoint is given a QR (Quick Response) code that is displayed on this user's screen; other users may join the checkpoint by scanning the QR code. Creating and joining checkpoints are intended to be frequent activities for users of the app; a checkpoint should ideally be created for any gathering or interaction where disease transmission could occur, and all who participate in the gathering or interaction should join the checkpoint. Thus, checkpoints would not only be created for interactions among friends, but also for public gatherings at places such as restaurants and grocery stores. At public gathering places, a device with the checkpoint QR code could be made available for users to scan at the entrance.

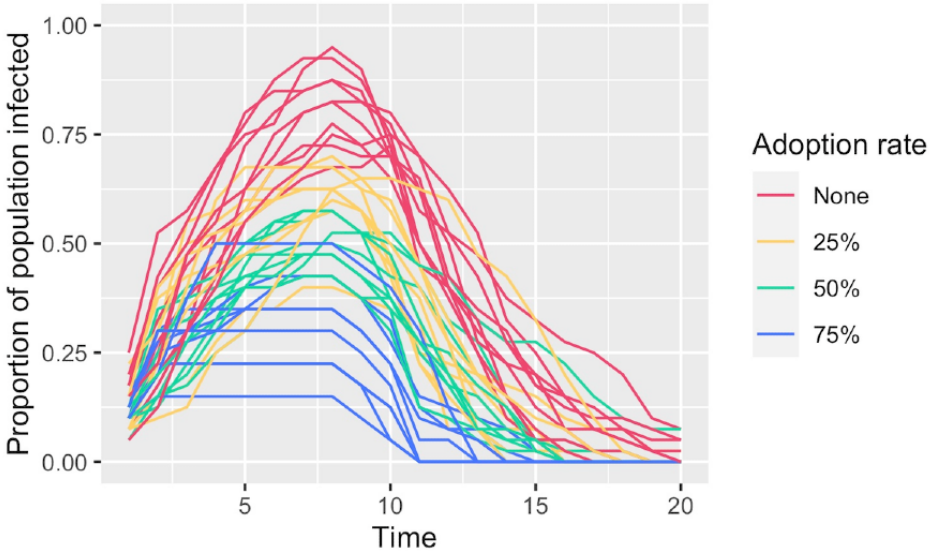
In the current app prototype, users may check their risk level by opening the app. The "exposures" tab displays the user's risk level, denoting whether any possible routes of transmission have been identified. Additionally, when the user has an elevated risk level, a banner is displayed at the top of the screen in the other tabs to make this information prominent. The app currently implements polling on a 30-second interval to retrieve the updated risk level from the server while the app is open. Future versions of the app may implement push notifications so that users are alerted when their risk level becomes elevated.

To report a diagnosis, users may press a button in the "exposures" tab. They are prompted to confirm this action before the server is notified. The app does not store any state changes associated with this action; this sacrifices good usability design, but it enhances the security and privacy of the user's health data by not retaining any record of the self-reported diagnosis.

Simulation Model

A comparison of simulations produced by our simulation model with varying levels of adoption is presented in Figure 3. Our results demonstrate that adoption rate is key to the impact that such an app could have on the extent of an outbreak. Visually, even a 25% adoption would provide some suppression of the infection curve compared to no adoption. However, more substantial improvements on the trajectory of the outbreak are observed at higher levels of adoption.

Figure 3. Comparison of infection curves from simulations at varying rates of peer-to-peer contact tracing application adoption. The proportion of the population with active infection is plotted across time for multiple adoption rates. Time is an arbitrary unit that represents the sequence of events in the simulation. The results of 10 random simulations per adoption rate are given.

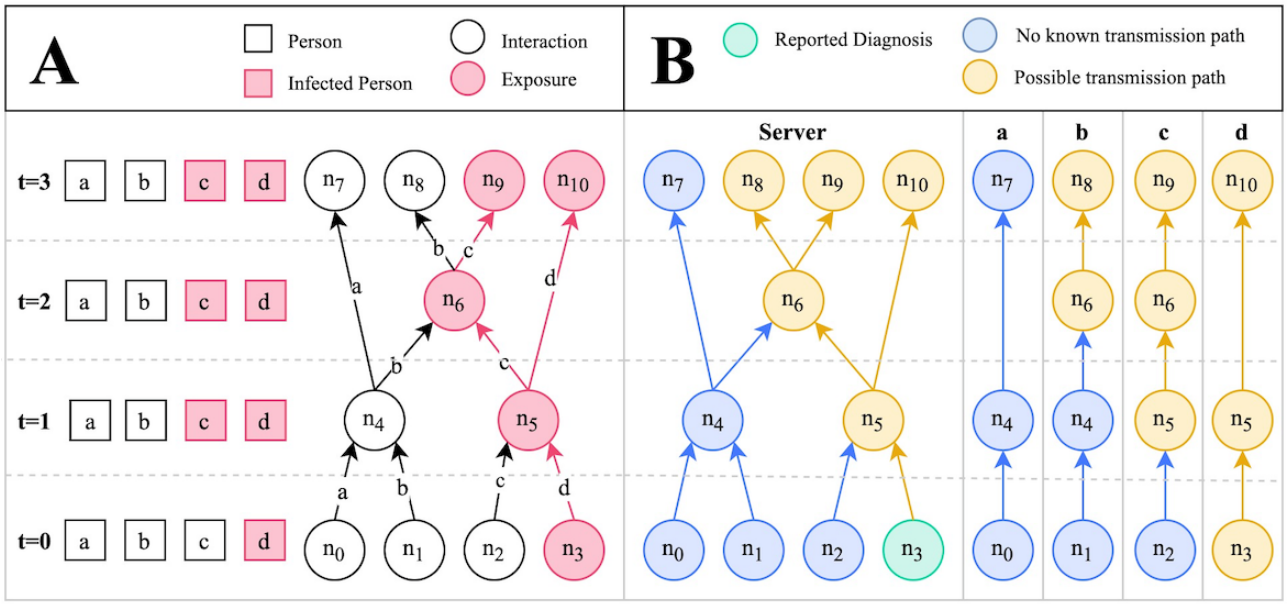


Discussion

Information Access

For the purposes of the smartphone app we are proposing, a key concept is the storage of and access to information. In the current prototype, all transmission graph data as described above is stored on a centrally managed server. Importantly, no user registration is required, and no personal information is collected; thus, the data on the server is anonymized. Users of the smartphone app will only be able to access their own checkpoints and whether there are any possible transmission paths to those checkpoints. An illustration of a hypothetical disease spread scenario, along with the corresponding transmission graph and the information available to the server and each user, is provided in Figure 4.

Figure 4. Disease spread scenario modeled as a transmission graph. (A) Graphical representation of a disease spread scenario across time. Contact points with infected individuals are denoted as exposures. Uninfected individuals may become infected at exposure points according to some probability (the transmission rate); hence, b does not become infected at n6. (B) Transmission graph corresponding to the scenario in A, depicting the information that is available to the server and each individual's smartphone app. Only one node, n3, is associated with a reported diagnosis. The infection risk level at the other contact points can be inferred by checking for possible transmission paths. n: node; t: time point.



Peer-to-Peer Design

Another important characteristic of the proposed app is its peer-to-peer nature. Traditional contact tracing apps often rely on a central entity to monitor individuals, their infection status, and their locations [3,16]. Our proposal does not require or use such entities. Instead, individuals rely on the joint participation of their peers, both in terms of creating contact points and

reporting diagnoses. For the purposes of restricting manipulation of the system through false reports, we have introduced a centralized system for administering confirmation codes to users with confirmed diagnoses; this system will be described shortly. A centralized system for this purpose was introduced because such a system provides the only mechanism by which it is possible to confirm diagnoses. Nonetheless, only the generation of confirmation codes is centralized. The network of user interactions remains peer-to-peer and anonymous.

Privacy, Usability, and User Adoption

User adoption is a critical factor in the success of any contact tracing app. Our simulation results suggest that if participation is not high, the infection will continue to spread through the population via those not using the app and the intervention will have limited effect. It is worth emphasizing that in certain countries (such as the United States), privacy concerns may present a significant barrier to adoption [17]. Location tracking is a type of data collection that may deter many from using an app, [18] especially when this data is being collected by or shared with government entities [17]. By not requesting access to location data, our app avoids this potential hurdle to adoption.

An additional barrier to entry that may often be overlooked is user registration. By user registration, we simply refer to the process of creating an account to use a web-based service, often by providing an email address and password. Some individuals will refrain from such a process due to concerns with disclosing their email address or other personal information such as their name or home address. For others, the process will become a barrier simply due to the inconvenience associated with the process [19]. Regardless of the motivating factor, user registration is likely to play a role in determining user adoption. Our app's lack of a user registration process can be expected to improve adoption rates.

Development Time and Complexity

Two aspects of developing smartphone apps that are often underappreciated are development time and complexity. Complex software not only takes longer to develop, but also is more prone to failures; a noteworthy example was the failed initial launch of the HealthCare.gov website [20]. The COVID-19 pandemic represents a time-sensitive public health crisis, and any technological solution applied to this crisis would need to be developed in a manner that is not only rapid, but also robust. The smartphone app we have proposed, due to its simplistic design, could be deployed swiftly without sacrificing robustness. Additionally, our proposed app would ideally be released under the endorsement or guidance of government entities, especially for the purposes of confirming diagnoses, but it would not require the type of intimate data sharing and coordination that are implied by alternative designs. Thus, government agency overhead would be minimal.

Manipulation and Fraud

In the first iteration of the app, there was no mechanism for verifying reports of positive diagnoses. Thus, the system was vulnerable to manipulation. We have addressed this issue by building a mechanism for reporting *confirmed* diagnoses. For this, we built a proof-of-concept administrative system that allows authorized users to generate *confirmation codes*. Authorized users could be health care or laboratory workers who process test results or deliver such results to patients. These confirmation codes, which are stored in a database on the server, can be printed by the authorized personnel in the form of QR codes, which are given to patients who receive positive diagnoses. The app allows users to opt for reporting a *confirmed* diagnosis, in which case the user is prompted to scan their printed QR code. The QR code data is sent to the server, where it is checked for validity against its database, and then marked as redeemed, so that it cannot be reused. The associated status positive contact point is marked as confirmed in the server's transmission graph.

In addition, users of the app are now presented with a setting that allows them to rely solely on confirmed diagnoses, in which case their risk level is calculated based only on *confirmed* status positive contact points. When this setting is enabled, users are protected from fraudulent reports. The app can therefore be used both with and without confirmation of diagnoses, depending on whether the app is used in a setting where generation of confirmation codes has been coordinated with health care and testing facilities.

It is important to note that while this feature requires some additional overhead in the form of coordinating with health care facilities that administer or deliver tests to patients, no patient data is ever used in generating or storing these confirmation codes. The codes are simply random sequences of characters, and authorized users are never prompted or allowed to enter patient data into the system. Thus, user privacy is not compromised with the addition of this feature. Because patient information is not processed through this system, health care or laboratory workers could be authorized liberally through a system of tiered privileges. Should a malicious user gain access to the system, this user would be able to do nothing more than manipulate the app by reporting false confirmed diagnoses.

Limitations

The smartphone app we propose is not without its limitations and concerns. The primary concern we have identified relates directly to the topic of user adoption. The use of location-based traffic detection algorithms may provide a more robust measure for estimating user location at points of contact, but this practice presents many potential privacy concerns, where some users simply may not be comfortable with an app that tracks their real-time location. Specifically with this app, users will be expected to create contact points by scanning QR codes whenever gathering with other people. This would ideally include not only private gatherings, but also public outings at places such as local businesses. To facilitate contact points for larger numbers of people, local businesses could allow customers to join a contact point by scanning a QR code upon entry. However, users may become fatigued from such behavior over time and choose to discontinue or may be dissuaded from participating at the onset. Under normal circumstances, these hurdles might deter most users; however, due to the tremendous impact of a pandemic, users may be motivated to overlook these inconveniences in light of alternative, more invasive location-tracking measures.

An additional concern is that there is no way to ensure that confirmed diagnoses will be reported through the app. Some users may feel uncomfortable reporting this information through the app, despite the anonymity provided. It should be expected that some fraction of users will not report their confirmed diagnoses, either due to negligence or privacy concerns. However, we have demonstrated through our simulations that a peer-to-peer contact tracing app such as ours can be effective without 100% participation. Thus, as long as a significant fraction of users with a positive diagnosis report through the app, we can expect the system to maintain its efficacy.

Finally, it should be noted that the transmission graph of interactions between individuals may not completely capture all possible routes of transmission, even with complete participation. For example, microbial agents may be transmitted via transportation of objects such as mailed envelopes and packages. Additionally, microbes may linger on surfaces and thus be transferred between individuals who are at the same location at different times. Measures can be taken to mitigate these transmission routes. First, objects and surfaces can be disinfected between interactions to limit transfer beyond the interaction. Second, contact points can be left open for longer periods of time to more accurately capture the possible transmission routes. For example, a grocery store may host a contact point by allowing customers to scan a QR code upon entry. This contact point could be left open from the opening time to the closing time so that all customers shopping that day would be registered into a single contact point, reflecting the fact that an infected shopper in the morning could transfer microbes across surfaces to a shopper that was present in the afternoon. To limit transmission of microbes from one day to the next, stores could thoroughly disinfect surfaces between closing time one day and opening time the next day.

Comparison With Related Work

The majority of contact tracing and similar surveillance apps rely heavily on the collection of intimate personal data and are not designed to preserve privacy. A popular smartphone app in the United States at the time of this writing is HEALTHLYNKED COVID-19 Tracker. This app differs from ours in two fundamental aspects. First, it is not a contact tracing app. The core feature of HEALTHLYNKED COVID-19 Tracker is a map that broadcasts locations of COVID-19 suspected cases, confirmed cases, and deaths based on a combination of self-reported and WHO-confirmed data. This data alone is insufficient for contact tracing, which requires identification with high specificity of individuals who have interacted with an infected person during some period of time prior to diagnosis. Second, HEALTHLYNKED COVID-19 Tracker requests access to user location, raising privacy concerns. Our peer-to-peer contact tracing app is able to perform detailed contact tracing while protecting user privacy by refraining from requesting personal data such as user location.

To our knowledge, only one prior attempt has been made to develop a privacy-preserving contact tracing app. PrivateKit: Safe Paths was developed in response to the COVID-19 pandemic as a privacy-preserving alternative to more invasive apps. PrivateKit: Safe Paths relies on location data, similar to other contact tracing apps; however, it offers unique protections for user data, such as encryption of location information, which can only be disclosed on a voluntary basis to government officials upon request.

Our app differs notably from PrivateKit: Safe Paths in two respects. First, PrivateKit: Safe Paths attempts to keep user location data private and secure, while our app does not collect this data in the first place, thus minimizing risk and maximizing user confidence. Second, the PrivateKit: Safe Paths app currently relies on intervention from a government entity once a positive diagnosis has been reported, while our app is peer-to-peer in this sense and can function entirely without intervention from any centralized entities, except to validate positive diagnoses. Both PrivateKit: Safe Paths and our peer-to-peer contact tracing app pursue similar objectives, but they differ considerably in their complexity and approach to protecting privacy. We offer that our app provides key advantages in terms of user privacy, simplicity, and minimal dependence on centralized entities, factors that can play a crucial role in responding to a public health emergency such as the COVID-19 pandemic.

An additional recent development is the release of Apple's official COVID-19 app. This app was developed in coordination with the Centers for Disease Control and Prevention and provides users with resources and guidance relating to the pandemic. In particular, it provides users with a questionnaire regarding symptoms and recent exposures and offers guidance on how to respond, including recommendations for measures such as self-isolation and social distancing, and whether or not testing is recommended. Such an app serves as a valuable resource for iOS users, and future iterations could potentially incorporate features of the app we have proposed here. The Apple COVID-19 app could additionally provide guidance on how to use a peer-to-peer contact tracing app.

Conclusions

We have proposed a novel peer-to-peer smartphone app for contact tracing that does not use personal data, such as location, and hence preserves user privacy. We have developed a prototype of this app, which is open source and publicly available as well as a computer simulation model that demonstrates the potential of our app to impact the course of a pandemic. Such an app could potentially be applied to the COVID-19 pandemic as well as others in the future to achieve a middle ground between drastic isolation measures and unmitigated disease spread. We hope that our proposal inspires future work in developing technology-based contact tracing solutions that preserve user privacy. Future work may involve testing our app on real devices in either a real or artificial pandemic setting as well as enhancements that build on our approach.

Conflicts of Interest

None declared.

References

1. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College 2020 Mar 16 [FREE Full text]
2. Armbruster B, Brandeau ML. Contact tracing to control infectious disease: when enough is enough. Health Care Manag Sci 2007 Dec;10(4):341-355 [FREE Full text] [doi: 10.1007/s10729-007-9027-6] [Medline: 18074967]
3. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, et al. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Glob Health Sci Pract 2015 Nov 12;3(4):646-659. [doi: 10.9745/ghsp-d-15-00207]
4. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer 2018 Aug;51(8):56-59. [doi: 10.1109/mc.2018.3191268]
5. Iyengar A, Kundu A, Pallis G. Healthcare informatics and privacy. IEEE Internet Comput 2018 Mar;22(2):29-31. [doi: 10.1109/mic.2018.022021660]
6. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. JAMA 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
7. Smith M, Szongott C, Henne B, Von Voigt G. Big data privacy issues in public social media. 2012 Jul 02 Presented at: IEEE International Conference on Digital Ecosystems and Technologies (DEST); 2012; Campione d'Italia, Italy p. 2012-2016. [doi: 10.1109/dest.2012.6227909]
8. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. PLoS One 2014 May 1;9(5):e95133 [FREE Full text] [doi: 10.1371/journal.pone.0095133] [Medline: 24787614]
9. GitHub. TrackCOVID: a peer-to-peer, privacy-preserving contact tracing app URL: https://github.com/tyleryasaka/covid-watch [accessed 2020-03-26]
10. World Health Organization. 2020 Mar 19. Considerations for quarantine of individuals in the context of containment for coronavirus disease (COVID-19) URL: https://tinyurl.com/rss5vzl [accessed 2020-03-26]
11. Kwok KO, Tang A, Wei VW, Park WH, Yeoh EK, Riley S. Epidemic models of contact tracing: systematic review of transmission studies of severe acute respiratory syndrome and Middle East respiratory syndrome. Comput Struct Biotechnol J 2019;17:186-194 [FREE Full text] [doi: 10.1016/j.csbj.2019.01.003] [Medline: 30809323]
12. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. J Theor Biol 2015 Nov 07;384:33-49. [doi: 10.1016/j.jtbi.2015.08.004] [Medline: 26297316]
13. Epidemic Simulation With and Without Peer-to-Peer Contact Tracing. URL: https://tyleryasaka.shinyapps.io/covidwatch [accessed 2020-03-26]
14. Osthus D, Hickmann KS, Caragea PC, Higdon D, Del Valle SY. Forecasting seasonal influenza with a state-space SIR model. Ann Appl Stat 2017 Mar;11(1):202-224. [doi: 10.1214/16-aoas1000]
15. Bjørnstad ON, Finkenstädt BF, Grenfell BT. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 2002 May;72(2):169-184. [doi: 10.1890/0012-9615(2002)072[0169:domees]2.0.co;2]
16. Prasad A, Kotz D. ENACT: encounter-based architecture for contact tracing. 2017 Jun Presented at: Proceedings of the 4th International on Workshop on Physical Analytics; 2017; Niagara Falls, New York p. E. [doi: 10.1145/3092305.3092310]
17. Dinev T, Hart P, Mullen MR. Internet privacy concerns and beliefs about government surveillance – an empirical investigation. The Journal of Strategic Information Systems 2008 Sep;17(3):214-233. [doi: 10.1016/j.jsis.2007.09.002]
18. Xu H, Gupta S. The effects of privacy concerns and personal innovativeness on potential and experienced customers' adoption of location-based services. Electron Markets 2009 Jun 30;19(2-3):137-149. [doi: 10.1007/s12525-009-0012-4]
19. Crumlish C, Malone E. Designing Social Interfaces: Principles, Patterns, and Practices for Improving the User Experience. Sebastopol, California: O'Reilly Media, Inc; 2009.
20. Anthopoulos L, Reddick CG, Giannakidou I, Mavridis N. Why e-government projects fail? an analysis of the Healthcare.gov website. Government Information Quarterly 2016 Jan;33(1):161-173. [doi: 10.1016/j.giq.2015.07.003]

Abbreviations

COVID-19: novel coronavirus disease 2019
QR: Quick Response
WHO: World Health Organization

Edited by T Sanchez; submitted 27.03.20; peer-reviewed by S Shahrestani, I Elsayed, E Da Silva, N Mostaghni; comments to author 31.03.20; revised version received 31.03.20; accepted 01.04.20; published 07.04.20.

Please cite as:
Yasaka TM, Lehrich BM, Sahyouni R
Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App
JMIR Mhealth Uhealth 2020;8(4):e18936
URL: https://mhealth.jmir.org/2020/4/e18936
doi:10.2196/18936
PMID:32240973

Original Paper

Lifestyle Intervention Enabled by Mobile Technology on Weight Loss in Patients With Nonalcoholic Fatty Liver Disease: Randomized Controlled Trial

Su Lin Lim[1], BSc, PhD; Jolyn Johal[1], BSc; Kai Wen Ong[1], BSc; Chad Yixian Han[1], BSc; Yiong Huak Chan[2], PhD; Yin Mei Lee[3], MBChB, MRCP, SAB; Wai Mun Loo[3], MBBS, MRCP

[1]Dietetics Department, National University Hospital, National University Health System, Singapore, Singapore

[2]Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]Department of Medicine, Gastroenterology and Hepatology, National University Hospital, National University Health System, Singapore, Singapore

Corresponding Author:
Su Lin Lim, BSc, PhD
Dietetics Department
National University Hospital
National University Health System
5 Lower Kent Ridge Road, Main Building, Level 1
Singapore, 119074
Singapore
Phone: 65 67724580
Email: su_lin_lim@nuhs.edu.sg

Abstract

Background: The prevalence of nonalcoholic fatty liver disease (NAFLD) reaches up to 30% in the Asian adult population, with a higher prevalence in obese patients. Weight reduction is typically recommended for patients at high risk or diagnosed with NAFLD, but is a challenge to achieve.

Objective: We aimed to evaluate the effect of a lifestyle intervention with a mobile app on weight loss in NAFLD patients.

Methods: This prospective randomized controlled trial included 108 adults with NAFLD confirmed by steatosis on ultrasound and a body mass index ≥23 kg/m^2 who were recruited from a fatty liver outpatient clinic. The patients were randomly allocated to either a control group (n=53) receiving standard care, consisting of dietary and lifestyle advice by a trained nurse, or an intervention group (n=55) utilizing the Nutritionist Buddy (nBuddy) mobile app in addition to receiving dietary and lifestyle advice by a dietitian. Body weight, alanine aminotransferase (ALT), aspartate aminotransferase (AST), waist circumference, and blood pressure were measured at baseline, and then at 3 and 6 months. Intention-to-treat and per-protocol analyses were used for statistical comparisons.

Results: The intervention group had a 5-fold higher likelihood (relative risk 5.2, *P*=.003, 95% CI 1.8-15.4) of achieving ≥5% weight loss compared to the control group at 6 months. The intervention group also showed greater reductions in weight (mean 3.2, SD 4.1 kg vs mean 0.5, SD 2.9 kg; *P*<.001), waist circumference (mean 2.9, SD 5.0 cm vs mean –0.7, SD 4.4 cm; *P*<.001), systolic blood pressure (mean 12.4, SD 14.8 mmHg vs mean 2.4, SD 12.4 mmHg; *P*=.003), diastolic blood pressure (mean 6.8, SD 8.9 mmHg vs mean –0.9, SD 10.0 mmHg; *P*=.001), ALT (mean 33.5, SD 40.4 IU/L vs mean 11.5, SD 35.2 IU/L; *P*=.004), and AST (mean 17.4, SD 27.5 U/L vs mean 7.4, SD 17.6 IU/L, *P*=.03) at 6 months.

Conclusions: Lifestyle intervention enabled by a mobile app can be effective in improving anthropometric indices and liver enzymes in patients with NAFLD. This treatment modality has the potential to be extended to a larger population scale.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12617001001381; https://tinyurl.com/w9xnfmp

(JMIR Mhealth Uhealth 2020;8(4):e14802) doi:10.2196/14802

KEYWORDS

diet; NAFLD; mHealth; mobile app, weight loss; liver enzymes; lifestyle intervention

Introduction

Nonalcoholic fatty liver disease (NAFLD) is characterized by excessive accumulation of fat in the liver that is not directly caused by alcohol consumption. NAFLD represents a clinicopathological spectrum that consists of hepatic steatosis and nonalcoholic steatohepatitis (NASH), and varying in degrees of inflammation and fibrosis [1,2]. Up to 20% of patients with NASH progress to cirrhosis and develop end-stage liver disease and associated complications [3,4].

With the epidemic surge in obesity and type 2 diabetes, the prevalence of NAFLD is on the rise, which is increasingly being recognized as a major cause of morbidity and mortality [5]. From 2002 to 2012, the number of patients undergoing liver transplantation for NASH-related hepatocellular carcinoma increased by 4-fold [6]. NASH is now poised to become the leading cause of hepatocellular carcinoma and indication for liver transplantation in the United States [6].

In Asia, an estimated 20%-30% of the adult population have been diagnosed with NAFLD, with a higher prevalence among patients with obesity [7]. The increase in prevalence can be attributed to a shift in dietary and lifestyle habits due to rapid globalization [8]. Notably, Asians are more susceptible to NAFLD at equivalent levels of overnutrition as compared to their Western counterparts, in part due to differences in the adiposity-muscle composition [1]. In view of the increased risk of morbidity in the Asian population, the World Health Organization (WHO) recommends 23 kg/m^2 as the body mass index (BMI) cut-off point for clinical action [9], while the International Diabetes Federation recommends waist circumference cut-off points of 90 cm and 80 cm for Asian men and women, respectively [10].

Despite pharmacological advances, lifestyle interventions remain a fundamental approach for the therapeutic management of NAFLD [11-13]. Suzuki et al [14] demonstrated the ability of a 5% weight loss in alleviating high serum alanine aminotransferase (ALT) and reducing hepatic steatosis in patients with NAFLD. However, the success of such weight loss interventions is dependent on the intensity of nutrition counseling and frequency of visits to dietitians and exercise therapists [15]. This renders the treatment modality resource-intensive and costly with high attrition rates, thereby limiting reach and scalability [16]. Mobile apps have recently gained increasing popularity in facilitating weight loss [17,18]. Self-monitoring of dietary intake and physical activity achieved through the convenience of mobile devices has been associated with greater reductions in energy intake and weight loss [19,20]. By mitigating the barriers associated with committing to repeated in-house nutrition and exercise therapy sessions, mobile apps increase the potential of expanding the reach and efficacy of lifestyle interventions.

To our knowledge, there is a paucity of research investigating the effects of integrating mobile apps into specialized weight loss programs for patients with NAFLD, with no full-sized randomized controlled trials published to date. Furthermore, a systematic review identified that none of the commercially available weight management mobile apps available included all major aspects of evidence-based strategies of self-monitoring, goal-setting, motivational strategies, healthy eating and physical activity support, social support, health or weight assessment, and personalized feedback, along with the involvement of health care professionals and formal scientific evaluation [21]. Accordingly, the objectives of the present study were to evaluate the effect of a lifestyle intervention consisting of diet and physical activity enabled by a mobile app that integrates a spectrum of evidence-based strategies along with health care professional involvement in facilitating weight loss and improving relevant health indicators in patients with NAFLD.

Methods

Study Participants

This parallel randomized controlled trial was conducted between July 2017 and November 2018 at National University Hospital (NUH), a tertiary university hospital in Singapore. Patients were recruited from an NAFLD clinic in NUH through referrals from clinicians after screening. NAFLD was defined as the presence of hepatic steatosis, either determined by imaging or histology, in the absence of secondary causes of hepatic fat accumulation such as significant alcohol consumption, long-term use of a steatogenic medication, or monogenic hereditary disorders, in accordance with the American Association for the Study of Liver Diseases guidelines [11]. Adults above 21 years of age who were diagnosed with NAFLD after excluding secondary causes of liver fat accumulation, and with a BMI≥23 kg/m^2, able to read and write in English, and who owned a smartphone with a data plan were included in the study. Exclusion criteria were consumption of more than one and a half times the limit of alcohol recommended for the population (15 g/day for women and 30 g/day for men) and those with infections of hepatitis B or C virus. Patients who were pregnant, receiving hepatotoxic medication, with cirrhosis, poorly controlled diabetes mellitus (HbA1c>10%), diabetes requiring insulin, cardiovascular event in the past 6 months, stage 4 and above kidney disease, concomitant liver disease, depression, untreated hypothyroidism, heart failure, and clinically or biochemically recognized systemic diseases were also excluded from the study. Written informed consent was obtained from each patient before enrolment. All participants were provided with a standardized model digital weighing scale (Omron HN-286, Japan) for self-monitoring of weight. This study was conducted in accordance with the Declaration of Helsinki, and received ethical approval from the National Healthcare Group Domain Specific Review Board in Singapore. The trial was registered at the Australian New Zealand Clinical Trials Registry (ACTRN12617001001381).

Randomization and Blinding

Stratified randomization of screened participants was carried out by gender, age (<40 years or ≥40 years), and BMI (<27.5 kg/m^2 or ≥27.5 kg/m^2). Within each stratum, participants were assigned to either the control or intervention group based on drawing from sealed opaque envelopes with an allocation ratio of 1:1, prepared by a third party not involved in the study and blinded to the study objectives according to the Consolidated

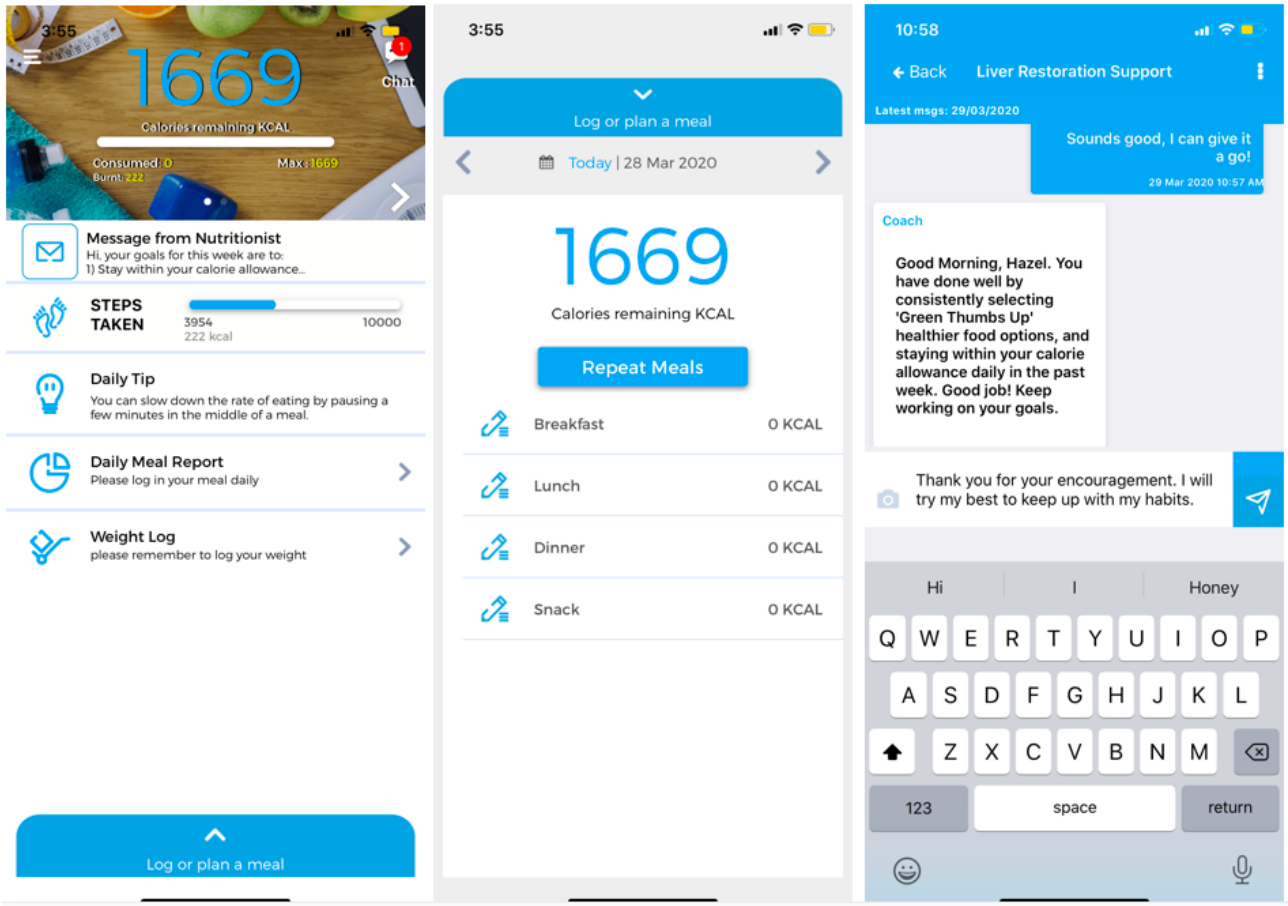
Standards of Reporting Trials (CONSORT) statement (see Multimedia Appendix 1) [22].

Mobile App Group (Intervention)

To evaluate the effect of lifestyle intervention enabled by a mobile app in facilitating weight loss, participants allocated to the intervention group were each provided with advice on dietary and physical activity modification by a dietitian via a single face-to-face session in the clinic followed by remote support through a mobile app for a 6-month period. They were taught how to utilize the Nutritionist Buddy (nBuddy) mobile app to track their diet and physical activity and induce behavioral changes to achieve optimal weight (Figure 1). The app was conceptualized by the principal investigator (SL) and developed by Verita Analytics in Singapore. nBuddy was developed using the Obesity-Related Behavioral Intervention Trials model for behavioral treatment as a framework for translating behavioral science discoveries into treatments, which is a flexible and robust process to design, conduct, and evaluate mobile app–based behavioral interventions [23]. The app is available commercially in app stores, with basic features accessible for free. Payment is required to unlock additional features such as videos, daily tips, and nutritionist support. Full features of the app were made available to the intervention participants as part of collaboration with Verita Analytics. The following set of in-built functions in nBuddy represent an amalgamation of evidence-based behavioral modification strategies to promote weight loss or maintenance.

The app includes a food diary logging system, coupled with individualized caloric goals based on the user's age, gender, and physical activity level, allowing for self-monitoring of intake [24]. Automatic recording of daily steps is achieved via syncing with the built-in pedometer of users' mobile devices to enable self-monitoring of physical activity [25]. The step goal increases automatically each week, from an initial goal of 3000 steps to 10,000 steps by the third week of usage. A range of physical activities can be logged manually if exercises were done in the absence of mobile devices. A weight logging function encourages the self-tracking of weight loss progression [26]. A dashboard enables dietitians to monitor users' input (ie, food intake and physical activity) and progress (ie, weight) to provide real-time feedback and encouragement [27]. A peer support chat channel allows users to connect with selected family members and peers to bolster user motivation [25,27]. A video viewing function delivers weekly educational clips [28]. An automated response system evaluates the suitability of food choices and provides instantaneous feedback, generating a list of healthier and culturally appropriate alternatives via an algorithm [29]. Daily, weekly, and monthly graph reports on weight, calorie intake, and steps facilitate the tracking of progress [24]. Finally, scripted daily tips and timed automated reminders prompt users to log in daily meal intake and twice weekly weight [30].

Figure 1. Screenshot of nBuddy homepage, food log page, and dietitian support chat channel from left to right, respectively.



Standard Care Group (Control)

Participants randomized to the control group were provided with usual standard care, which consisted of advice on dietary and physical activity modification as per American Heart Association guidelines [31] by a nurse trained in diet counseling via a single face-to-face session in the NAFLD clinic. The syllabus was similar to that provided in the face-to-face visit with the dietitian in the intervention group. This was part of the routine clinical service offered to overweight or obese NAFLD patients seen at the NAFLD clinic.

Outcome Measurements

All outcomes were part of routine measurements taken by trained nurses and blood tests conducted at the outpatient NAFLD clinic. Assessors were not blinded to the groups allocated to the study participants. Body weight was measured using a calibrated digital weighing machine (Seca 767, Germany) to the nearest 0.1 kg. Height was measured in meters to two decimal points using the stadiometer attached to the Seca scale and the corresponding BMI was calculated. Waist circumference was measured using a tape measure at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest as recommended by the WHO [32]. ALT and aspartate aminotransferase (AST) levels were determined by an automated kinetic method. Blood pressure (for participants with hypertension) was measured using an automatic blood pressure monitor (Carescape Dinamap V100; GE Healthcare, Chicago, IL, USA). Participants' characteristics such as age, gender, ethnicity, and existing relevant comorbidities were also collected at baseline. App utilization data were collected through the app developer.

Statistical Analysis

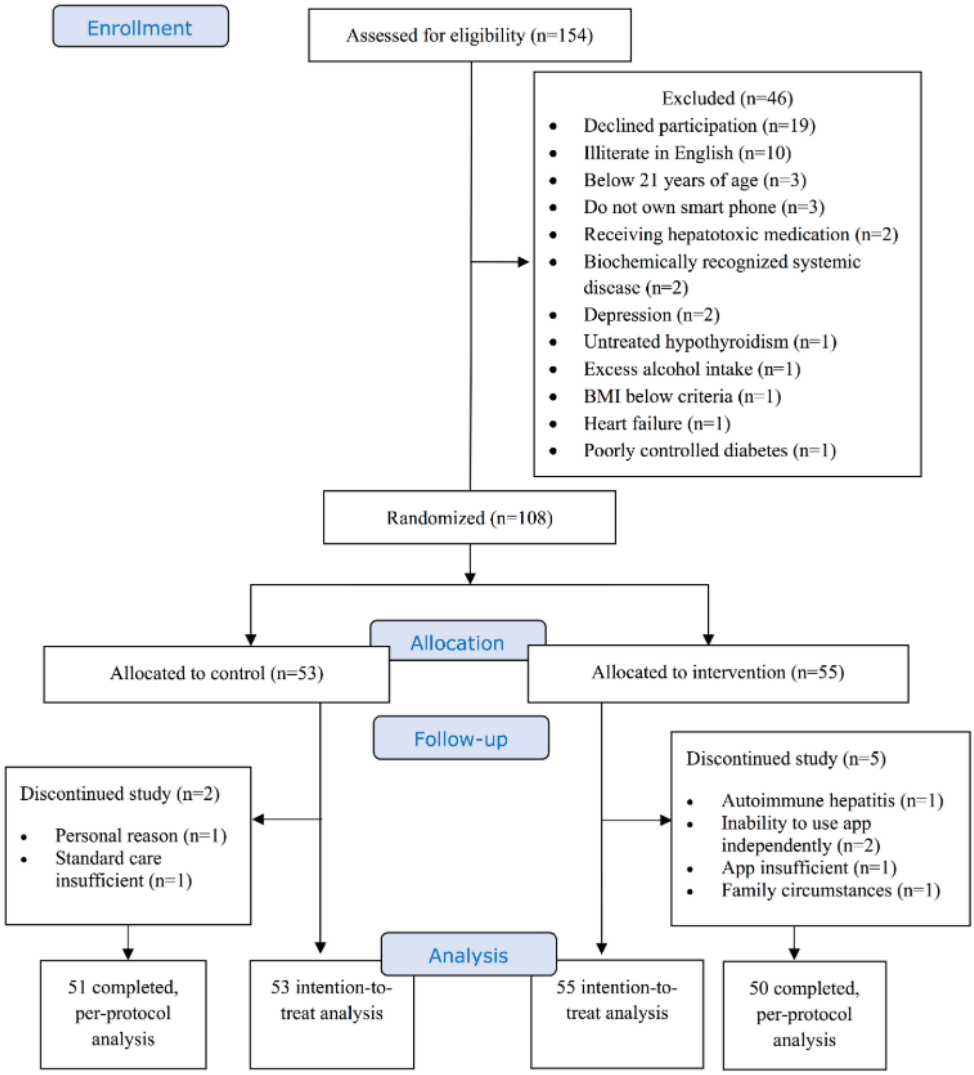
Studies that provided nutrition therapy interventions targeted at weight loss in patients with NAFLD were referenced for sample size calculation [14,24]. The primary unit of interest was weight loss of at least a 5% by 6 months. It was postulated that 10% of the control subjects will achieve this successful outcome and the intervention would increase this rate by 4-fold [33]. With 90% power at a 5% significance level, and allowing for a 10% dropout rate, a sample size of 100 (50 per group) would be required. Data analyses were performed using SPSS for Windows version 21.0 (SPSS Inc, Chicago, IL, USA). Results are expressed as mean and SD for normally distributed variables, and as the median and interquartile range for variables that did not satisfy normality criteria. Categorical data are expressed as frequencies and percentages. To compare baseline characteristics, and 3- and 6-month changes between groups, Chi square and independent samples *t* tests were used for categorical and continuous variables, respectively. Between-group differences in the numerical and binary outcomes were compared using a general linear model and Poisson regression model, respectively, adjusting for age, gender, and ethnicity. Between-group Cohen *d* effect sizes were calculated with the formula $(M_{intervention}–M_{control})/SD_{pooled}$, where $M_{intervention}$ and $M_{control}$ are the mean changes in outcomes from baseline in the intervention and control groups, respectively.

Results

Participants

A total of 154 referred subjects were screened for participation. Forty-six did not meet the eligibility criteria, primarily due to refusal to participate and a lack of English literacy. The remaining 108 subjects were enrolled and randomized (55 to the intervention group and 53 to the control group). Seven of the enrolled participants withdrew from the study, including 5 (4.6%) from the intervention group and 2 (1.9%) from the control group. A total of 101 patients completed the study, with 50 allocated to the intervention group (Figure 2).

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) study flow diagram. BMI: body mass index.



Baseline Characteristics

The baseline characteristics of the participants are summarized in Table 1. There were no significant differences between groups in terms of baseline weight and other characteristics such as presence of relevant comorbidities, age, weight, and clinical and biochemical outcomes of interest. Per-protocol analysis was carried out on the 101 patients who completed the study, and intention-to-treat analysis was carried out among all 108 patients assigned to the original groups, where data were available for each of the parameters.

Table 1. Baseline characteristics of study participants.

Variable	Intention-to-treat			Per-protocol		
	Control (n=53)	Intervention (n=55)	*P* value[a]	Control (n=51)	Intervention (n=50)	*P* value[a]
Gender, n (%)			.30			.37
Male	36 (68)	32 (58)		35 (69)	30 (60)	
Female	17 (32)	23 (42)		16 (31)	20 (40)	
Ethnicity, n (%)			.21			.15
Chinese	41 (77)	42 (76)		39 (77)	39 (78)	
Malay	9 (17)	6 (11)		9 (18)	5 (10)	
Indian	0 (0)	4 (7)		0 (0)	4 (8)	
Others	3 (6)	3 (6)		3 (6)	2 (4)	
Diabetes, n (%)	19 (36)	11 (20)	.66	18 (35)	10 (20)	.09
Hyperlipidemia, n (%)	40 (76)	39 (71)	.59	40 (78)	37 (74)	.60
Hypertension, n (%)	37 (70)	38 (69)	.94	37 (73)	36 (72)	.95
Age (years), mean (SD)	46.2 (10.1)	46.8 (11.1)	.76	46.1 (10.3)	46.2 (11.0)	.96
Weight (kg)	86.1 (19.4)	81.5 (15.2)	.17	86.6 (19.6)	82.0 (15.7)	.20
BMI[b] (kg/m^2)	30.8 (4.8)	30.1 (4.0)	.04	30.9 (4.8)	30.1 (4.1)	.37
Waist circumference (cm), mean (SD)	101.7 (11.4)	98.6 (9.0)	.14	101.9 (11.6)	98.9 (10.2)	.17
ALT[c] (IU/L), mean (SD)	78.1 (46.7)	75.6 (43.3)	.77	75.3 (40.9)	76.8 (41.8)	.85
AST[d] (IU/L), mean (SD)	50.1 (28.8)	49.1 (28.4)	.86	48.6 (26.2)	48.6 (25.7)	>.99
Systolic blood pressure (mmHg), mean (SD)	135 (12)	134 (14)	.72	140 (10)	141 (10)	.58
Diastolic blood pressure (mmHg), mean (SD)	76 (9)	75 (11)	.46	78 (10)	79 (10)	.67

[a]Chi square or independent samples *t* test as appropriate.

[b]BMI: body mass index.

[c]ALT: alanine aminotransferase.

[d]AST: aspartate aminotransferase.

Weight Loss

There was a significantly greater number of participants who achieved ≥5% weight loss in the intervention as compared to the control group at both 3 and 6 months (Table 2). After adjustment for age, gender, and ethnicity, the intention-to-treat and per-protocol analyses showed that use of the mobile-enabled lifestyle intervention program was independently associated with a higher likelihood of achieving a ≥5% weight loss at 3 and 6 months when compared with standard care.

Table 2. Proportion of patients with ≥5% weight loss in the control and intervention groups.

Analysis	Control[a], n (%)	Intervention[b], n (%)	Unadjusted RR[c] (95% CI)	*P* value	Adjusted[d] RR (95% CI)	*P* value
Intention-to-treat						
3 months	4 (8)	14 (25)	3.4 (1.1-10.2)	.03	3.5 (1.1-10.9)	.03
6 months	4 (8)	22 (40)	5.3 (1.8-15.3)	.002	5.2 (1.8-15.4)	.003
Per-protocol						
3 months	4 (8)	14 (28)	3.6 (1.2-10.8)	.03	3.8 (1.3-11.6)	.02
6 months	4 (8)	22 (44)	5.6 (1.9-16.3)	.002	5.8 (2.0-16.9)	.001

[a]N=53 for intention-to treat; N=51 for per-protocol.

[b]N=55 for intention-to-treat; N=50 for per-protocol.

[c]RR: relative risk; the control group is the reference (1.00).

[d]Adjusted for age, gender, ethnicity.

Table 3 and Table 4 show that participants in the intervention group lost significantly more weight than those in the control group at 3 and 6 months based on the intention-to-treat and per-protocol analysis, respectively. The significantly greater improvements in both absolute and percentage weight loss in the intervention group were more pronounced at 6 months as compared to 3 months, with large between-group effect sizes in percentage weight loss at both 3 and 6 months (Cohen d>0.8 for all). Reductions in BMI were also significantly greater in the intervention group at 3 and 6 months. Results of the univariate analyses remained significant (P<.001) after further adjustments for age, gender, and ethnicity in both analyses.

Table 3. Mean (SD) changes in anthropometric, biochemical, and clinical outcomes in patients with nonalcoholic fatty liver disease from baseline at 3 and 6 months after enrolment based on intention-to-treat analysis.

Variable	n	Change from baseline		Between-group differences			
		Control	Intervention	Mean difference (95% CI)	*P* value (unadjusted)	*P* value (adjusted[a])	Cohen *d*
Weight, kg							
3 months	108	–0.8 (2.1)	–3.2 (3.1)	2.3 (1.3-3.3)	<.001	<.001	0.91
6 months	108	–0.5 (2.9)	–3.2 (4.1)	2.6 (1.3-4.0)	<.001	<.001	0.76
Weight, %							
3 months	108	–0.9 (2.5)	–3.7 (3.7)	3.0 (1.8-4.2)	<.001	<.001	0.95
6 months	108	–0.6 (3.5)	–4.4 (5.6)	3.8 (2.0-5.5)	<.001	<.001	0.81
BMI[b], kg/m^2							
3 months	108	–0.4 (0.8)	–1.3 (1.1)	0.9 (0.5-1.2)	<.001	<.001	0.66
6 months	108	–0.3 (1.1)	–1.3 (1.4)	1.0 (0.5-1.5)	<.001	<.001	0.79
Waist circumference, cm							
3 months	105	0.4 (4.7)	–3.4 (5.1)	3.8 (1.9-5.7)	<.001	<.001	0.77
6 months	105	0.7 (4.4)	–2.9 (5.0)	3.6 (1.8-5.5)	<.001	<.001	0.76
ALT[c], IU/L							
3 months	75	–20.7 (32.2)	–37.2 (37.6)	16.5 (0.4-32.7)	.045	.04	0.47
6 months	103	–11.5 (35.2)	–33.5 (40.4)	22.0 (7.2-36.8)	.004	.006	0.58
AST[d], IU/L							
3 months	76	–11.1 (19.1)	–20.2 (26.9)	9.1 (–1.7-19.9)	.10	.07	0.39
6 months	104	–7.4 (17.6)	–17.4 (27.5)	10.0 (1.0-19.0)	.03	.03	0.43
Systolic blood pressure (mmHg)							
3 months	63	1.1 (12.6)	–13.7 (14.9)	14.8 (7.9-21.8)	<.001	<.001	1.07
6 months	72	–2.4 (12.4)	–12.4 (14.8)	10.1 (3.6-16.5)	.003	.008	0.74
Diastolic blood pressure (mmHg)							
3 months	63	2.3 (7.7)	–6.3 (9.8)	8.6 (4.1-13.1)	<.001	<.001	0.98
6 months	72	0.9 (10.0)	–6.8 (8.9)	7.7 (3.2-12.1)	.001	.003	0.81

[a]Adjusted for age, gender, and ethnicity.

[b]BMI: body mass index.

[c]ALT: alanine aminotransferase.

[d]AST: aspartate aminotransferase.

Table 4. Mean (SD) changes in anthropometric, biochemical, and clinical outcomes in patients with nonalcoholic fatty liver disease from baseline at 3 and 6 months after enrolment based on per-protocol analysis.

Variable	n	Change from baseline		Between-group differences			
		Control	Intervention	Mean difference (95% CI)	*P* value (unadjusted)	*P* value (adjusted[a])	Cohen *d*
Weight, kg							
3 months	101	–1.0 (2.0)	–3.2 (3.1)	2.3 (1.2-3.3)	<.001	<.001	0.88
6 months	101	–0.7 (2.8)	–4.1 (3.8)	3.5 (2.1-4.8)	<.001	<.001	1.05
Weight, %							
3 months	101	–1.1 (2.4)	–4.0 (3.7)	2.9 (1.7-4.1)	<.001	<.001	0.93
6 months	101	–0.7 (3.3)	–5.1 (4.5)	4.4 (2.9-6.0)	<.001	<.001	1.11
BMI[b], kg/m^2							
3 months	101	–0.4 (0.8)	–1.4 (1.1)	1.0 (0.6-1.3)	<.001	<.001	1.04
6 months	101	–0.3 (1.0)	–1.4 (1.5)	1.1 (0.6-1.6)	<.001	<.001	0.86
Waist circumference, cm							
3 months	99	0.24 (4.69)	–3.57 (5.14)	3.8 (1.8-5.8)	<.001	<.001	0.77
6 months	99	0.83 (4.4)	–3.24 (5.1)	4.1 (2.2-6.0)	<.001	<.001	0.86
ALT[c], IU/L							
3 months	70	–18.7 (26.7)	–37.2 (36.2)	18.5 (3.3-33.8)	.02	.02	0.58
6 months	98	–10.0 (32.7)	–32.7 (38.4)	22.7 (8.8-36.6)	.002	.002	0.64
AST[d], IU/L							
3 months	71	–9.18 (15.2)	–18.9 (23.8)	9.7 (0.2-19.3)	.046	.04	0.49
6 months	99	–6.35 (15.3)	–16.1 (24.5)	9.7 (1.6-17.9)	.02	.02	0.47
Systolic blood pressure (mmHg)							
3 months	61	1.1 (12.6)	–13.4 (15.2)	14.5 (7.4-21.6)	<.001	<.001	1.04
6 months	70	–2.4 (12.4)	–11.8 (15.0)	9.4 (2.9-16.0)	.005	.01	0.68
Diastolic blood pressure (mmHg)							
3 months	61	2.3 (7.7)	–6.1 (10.5)	8.4 (3.8-12.9)	.001	<.001	0.91
6 months	70	–0.9 (10.0)	–6.5 (9.0)	7.4 (2.8-11.9)	.002	.006	0.78

[a]Adjusted for age, gender, and ethnicity.

[b]BMI: body mass index.

[e]ALT: alanine aminotransferase.

[d]AST: aspartate aminotransferase.

Waist Circumference

Participants in the intervention group had greater reductions in waist circumference compared to those in the control group at 3 months. These greater reductions in waist circumference remained statistically significant at 6 months in both analyses, with a large effect size observed at 6 months in the per-protocol analysis (Table 4).

Liver Enzymes

Both ALT and AST levels decreased markedly at 3 and 6 months in both the intention-to-treat and per-protocol analyses, irrespective of treatment group. However, there were significantly greater reductions of ALT and AST in the intervention group compared to the control group at 6 months, which remained after further adjustments for age, gender, and ethnicity in both analyses. Between-group differences in AST showed a small effect size (Cohen d<0.5), whereas a moderate effect size in ALT was observed at 6 months in both the intention-to-treat and per-protocol analyses (Tables 3 and 4).

Blood Pressure

In both the intention-to-treat and per-protocol analyses, participants with hypertension in the intervention group had significantly greater reductions in systolic and diastolic blood pressure from baseline at 3 and 6 months as compared to those receiving standard care. After including all enrolled participants and adjusting for covariates, those assigned to the mobile app program were able to reduce their systolic and diastolic blood pressure to a markedly greater extent than those receiving standard care by an average of 15 mmHg and 9 mmHg, respectively, at 3 months. Using the same analysis methods,

this difference remained statistically significant at 6 months. The effect size for both systolic and diastolic pressure was large at 3 months (Cohen *d*>0.8 for all) and was moderate for systolic blood pressure at 6 months (Tables 3 and 4).

Mobile App Utilization

Data on app use of the 49 participants in the intervention group were obtained from the app developer. One account was deleted by the participant upon completing the program and was unavailable due to the privacy protection feature of the app. Overall, we observed a high percentage of active users in the intervention, with 76% (37/49) of participants showing a daily log-in of 137 days over a 182-day period (>75.3% of the total time). The average log-in days for the first 3 months, 4-6 months, and overall were 79.6 days (SD 17.9), 71.1 days (SD 25.6), and 151 days (SD 41.1), respectively. The mean percentage of log-in days was 87.6% (SD 19.6) in the first 3 months and decreased to 78.1% (SD 28.2) at 4-6 months. Furthermore, meal and weight logging were at 56.7% (SD 51.6) and 77.0% (SD 28.5) of the recommended utilization rate of daily and twice a week, respectively.

Discussion

Principal Findings

This study demonstrated that a 6-month mobile-enabled lifestyle intervention was able to produce clinically meaningful outcomes in patients with NAFLD. Patients enrolled in the 6-month nBuddy mobile app intervention program had a 5-fold higher likelihood of achieving ≥5% weight loss as compared to those receiving standard care. In addition, the mobile-enabled lifestyle intervention appeared to have a positive influence on components of surrogate markers of NAFLD such as waist circumference and BMI, along with improvements in liver enzymes (AST, ALT) and blood pressure. These positive results remained significant after an intention-to-treat approach, suggesting a notable effect among NAFLD patients. Our findings support the consensus that a modest weight loss of about 5% of baseline body weight within a 6-month period is associated with clinically meaningful reductions in liver enzymes [11,13,34].

Weight loss, along with control of metabolic risk factors, is the cornerstone of management of NAFLD in the absence of effective medical therapy. Current practice guidelines advocate a target weight loss of 7%-10% to achieve improvement in steatosis and inflammation [11,35]. In a Cuban study, a weight loss of ≥5% led to a resolution of steatohepatitis and reduction in the NAFLD activity score of at least 2 points, without worsening of fibrosis; however, regression of fibrosis occurred in those with ≥10% weight loss [36]. Hence, the greater the extent of weight loss with lifestyle changes, the more significant the improvement in NASH and fibrosis.

The use of mobile apps and online platforms as part of weight management in patients with noncommunicable diseases is increasingly growing in popularity. This stems partly from its ability to overcome issues faced by conventional weight management programs that require physical attendance. Health care providers observed a high nonattendance rate for patients with NAFLD requiring dietetics intervention and follow up [37,38]. Jiandani et al [39] reported a high attrition rate for in-house weight management programs whereby 34% of the participants discontinued after a single visit. The reasons cited included limited ability to pay for program services, limited time, and transportation challenges in attending appointments with health care professionals [40]. Other barriers included the lack of motivation and both patient and physician time constraints. In this study, the attrition rate of the intervention group using the mobile app (5%) was markedly lower than that of conventional face-to-face dietetics intervention cited in the literature [39]. This novel approach is also effective in addressing health behavior changes, thus leading to weight loss and beneficial outcomes [41]. The use of mobile apps similarly benefits health care practitioners through facilitating relationships with patients by enhancing the speed and ease of communication [42]. Use of a mobile app further provides practitioners with additional information in a timely manner that can improve patient care [42].

Previous studies have demonstrated the effectiveness of weight loss programs targeted at NAFLD patients in achieving a ≥5%-10% weight loss with a normalization of liver enzymes [43,44]. However, their scalability to a larger population poses a challenge. The use of mobile apps offers an alternative that might be scalable, more logistically friendly, and cost-effective with comparable or better results [45]. In fact, a recent meta-analysis on weight loss interventions in adults with chronic diseases found a markedly greater weight reduction of 2.5 kg among users in the mobile app group as compared with those in the nonmobile app group [45]. Another study examining lifestyle changes in patients with NAFLD reported an average weight reduction of 3.4 kg from baseline weight at 6 months via an online remote coaching method [34]. Similarly, our results demonstrated a mean weight loss of 4.1 kg over the 6-month period in participants who completed the mobile app program, which corresponded to a mean 5.1% reduction in weight. Our findings suggest that a mobile app with multifaceted automated algorithms and remote coaching is capable of helping patients with NAFLD lose a significant amount of weight.

Decreased aminotransferase levels have consistently been associated with histological improvement in NAFLD and are used as surrogate markers of inflammation. A study involving intensive nutrition counseling on patients with NAFLD found an average reduction of 33.1 IU/L and 15.4 IU/L in ALT and AST, respectively, at 12 months [15]. Although the current study did not have sufficient power to evaluate liver enzymes, our results suggest a similar trend, whereby participants enrolled in the mobile app intervention group had a mean reduction of 33.5 IU/L and 17.4 IU/L in ALT and AST, respectively, at 6 months. These improvements were also significantly greater when compared to those of the control group. In addition, Straznicky et al [46] found a concomitant 20% reduction in ALT along with 8% weight loss in participants receiving dietary intervention at 3 months [46]. Our results suggest a 2-fold (44%) reduction of ALT compared to the previous study, at approximately half (4%) the weight loss achieved at 3 and 6 months. This observation could be contributed by the differences in ethnicities between the two cohorts considering the higher

visceral adiposity at a lower BMI in the Asian population [47,48]. This also suggests that Asian patients with NAFLD may benefit more from weight loss, with an augmented response to reduction in liver enzymes. The effectiveness of a mobile app–enabled lifestyle intervention can also consequently be seen as comparable to conventional weight loss programs.

Our study showed a significant difference in reduction of waist circumference (3.6 cm) in intervention as compared to control participants at 6 months using an intention-to-treat analysis. Pooled analysis of existing literature showed a greater mean reduction in waist circumference of 2.5 cm among mobile app users [45]. Our results also concur with another study showing an average greater reduction of 3.9 cm among users in a mobile phone weight loss program as compared to those in the control group at 12 months [49].

It is well-established that weight loss is correlated to reductions in blood pressure. A meta-analysis of randomized controlled trials on the influence of weight changes on blood pressure found an average of 1 mmHg reduction in systolic and diastolic blood pressure per kilogram of weight loss [50]. In a recent pilot study, systolic and diastolic blood pressure among mobile app users with a mean of 9.4% weight loss were 6 mmHg and 4 mmHg, respectively, at 3 months [49]. Our results corroborate with these findings and substantiate the benefits of blood pressure control in patients with NAFLD involved in a mobile app–enabled weight loss program.

Strengths and Limitations

This study has several strengths. The use of randomized stratification ensured that the baseline characteristics of the study participants were similar between the control and intervention groups. Use of an adjusted intention-to-treat analysis model provided a more realistic indicator of the potential for program success when extrapolated to a clinical setting. This study is one of the first randomized controlled trials to investigate the effects of a mobile app utilizing a spectrum of evidence-based strategies in facilitating weight loss and improvement of relevant physiological and biochemical markers among patients with NAFLD. This intervention employed a mobile app that includes local dishes in the database, as well as culturally specific healthier alternatives. Having been developed with professional input, it fills the gap for research-tested apps developed with health care professional involvement, particularly in an Asian context [21].

This modality of intervention delivered through a well-designed mobile app was successful in achieving significant weight loss in NAFLD patients. Therefore, findings from this study can guide the successful incorporation of a mobile app into conventional lifestyle interventions for patients with NAFLD. Notably, the use of a mobile app in the NAFLD dietetics clinic has augmented service delivery in NUH since its inception. This program also has the potential to be applied in the management of other chronic diseases such as diabetes and cardiovascular diseases, as well as being extended to a wider population. Furthermore, our data suggest that a 3-month nBuddy app liver restoration program has the potential of yielding clinically meaningful weight loss in patients with NAFLD should a shorter time frame be desired.

The study also has a few limitations. First, this was a single-center trial with an intervention that precluded nonsmartphone users and those illiterate in English. Although smartphone ownership is rising, it continues to be dependent on education attainment and household income [51]. This trial might have thus included a smaller proportion of people of lower socioeconomic status. Along with the exclusion of nonEnglish users, this restricts the generalizability of findings to the greater population. Owing to the lack of research-tested culturally specific mobile apps for nonEnglish users, it would be worthwhile to explore different language versions of the mobile app should they be made available [52].

Second, the intervention continued to require the expertise of dietitians in coaching participants via the dashboard. However, the enablement of one dietitian in supporting multiple patients remotely may render the intervention more cost-effective as compared to face-to-face consultations. There is potential to scale up the solution considering the automated features such as provision of alternative healthier food options in the app, which can facilitate behavior change without the costs associated with intensive face-to-face counseling. Moreover, the delivery of advice by different health care practitioners in both groups also limits the ability to attribute the results solely to the mobile app. Nevertheless, this trial offers insight into the feasibility of introducing a combined dietitian and mobile app model in outpatient care settings where dietitians may not yet have a clear role. Finally, it would have been preferable for all patients enrolled in the study to have biopsy-proven NASH and an end-of-treatment biopsy. However, this was limited by funding and time constraints.

Conclusions

Lifestyle intervention enabled by a mobile app can be effective in improving weight loss and anthropometric and clinical indices in patients with NAFLD. This study provides new insights on the feasibility and effectiveness of mobile apps in facilitating lifestyle interventions in NAFLD, along with further guidance for the application of this treatment modality to other chronic diseases. Health care professionals seeking to provide weight loss interventions to patients with NAFLD can now consider employing the use of a mobile app in effecting behavioral changes, including self-monitoring. Future studies can investigate the effectiveness of this treatment modality on a wider population, as well as evaluate important histological outcomes such as liver fibrosis.

Acknowledgments

The study was funded by National University Health System Allied Health and Nursing Grant (NUHSRO/2016/061/AHN/08) and the National Medical Research Council Grant (HSRG/0063/2016). The authors would like to thank Dr Tan Poh Seng, Dr

Ong Li Zhen, Dr Anita Lim, Dr Junita Lee, Ms Genevieve Yeo, and the nurses from the University Digestive Clinic, National University Hospital, for supporting this study.

Authors' Contributions

SL, YL, and WL conceived the idea, developed the study design, and supervised the project. JJ, KO, and CH carried out the project. YC provided statistical expertise and contributed to the interpretation of the results. SL took the lead in writing the manuscript. All authors provided critical feedback and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH V 1.6.1 checklist.
[PDF File (Adobe PDF File), 1318 KB - mhealth_v8i4e14802_app1.pdf]

References

1. Farrell GC, Wong VW, Chitturi S. NAFLD in Asia--as common and important as in the West. Nat Rev Gastroenterol Hepatol 2013 May;10(5):307-318. [doi: 10.1038/nrgastro.2013.34] [Medline: 23458891]
2. Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013 Dec;10(11):686-690. [doi: 10.1038/nrgastro.2013.171] [Medline: 24042449]
3. Wong MCS, Huang J. The growing burden of liver cirrhosis: implications for preventive measures. Hepatol Int 2018 May;12(3):201-203. [doi: 10.1007/s12072-018-9865-y] [Medline: 29679258]
4. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther 2011 Aug;34(3):274-285 [FREE Full text] [doi: 10.1111/j.1365-2036.2011.04724.x] [Medline: 21623852]
5. Preiss D, Sattar N. Non-alcoholic fatty liver disease: an overview of prevalence, diagnosis, pathogenesis and treatment considerations. Clin Sci (Lond) 2008 Sep;115(5):141-150. [doi: 10.1042/CS20070402] [Medline: 18662168]
6. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. Hepatology 2014 Jun;59(6):2188-2195. [doi: 10.1002/hep.26986] [Medline: 24375711]
7. Wah-Kheong C, Khean-Lee G. Epidemiology of a fast emerging disease in the Asia-Pacific region: non-alcoholic fatty liver disease. Hepatol Int 2013 Mar;7(1):65-71. [doi: 10.1007/s12072-012-9384-1] [Medline: 26201622]
8. Wong MCS, Huang JLW, George J, Huang J, Leung C, Eslam M, et al. The changing epidemiology of liver diseases in the Asia-Pacific region. Nat Rev Gastroenterol Hepatol 2019 Jan;16(1):57-73. [doi: 10.1038/s41575-018-0055-0] [Medline: 30158570]
9. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
10. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med 2006 May;23(5):469-480. [doi: 10.1111/j.1464-5491.2006.01858.x] [Medline: 16681555]
11. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018 Jan;67(1):328-357. [doi: 10.1002/hep.29367] [Medline: 28714183]
12. Marchesini G, Petta S, Dalle Grave R. Diet, weight loss, and liver health in nonalcoholic fatty liver disease: Pathophysiology, evidence, and practice. Hepatology 2016 Jun;63(6):2032-2043. [doi: 10.1002/hep.28392] [Medline: 26663351]
13. Méndez-Sánchez N, Arrese M, Zamora-Valdés D, Uribe M. Treating nonalcoholic fatty liver disease. Liver Int 2007 Nov;27(9):1157-1165. [doi: 10.1111/j.1478-3231.2007.01567.x] [Medline: 17919226]
14. Suzuki A, Lindor K, St Saver J, Lymp J, Mendes F, Muto A, et al. Effect of changes on body weight and lifestyle in nonalcoholic fatty liver disease. J Hepatol 2005 Dec;43(6):1060-1066. [doi: 10.1016/j.jhep.2005.06.008] [Medline: 16140415]
15. Huang MA, Greenson JK, Chao C, Anderson L, Peterman D, Jacobson J, et al. One-year intense nutritional counseling results in histological improvement in patients with non-alcoholic steatohepatitis: a pilot study. Am J Gastroenterol 2005 May;100(5):1072-1081. [doi: 10.1111/j.1572-0241.2005.41334.x] [Medline: 15842581]
16. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]
17. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117. [doi: 10.1007/s11606-014-2987-6] [Medline: 25134692]
18. Lyzwinski L. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014 Jun 30;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]

19. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43. [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
20. 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]
21. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile Apps for Weight Management: A Scoping Review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
22. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
23. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
24. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
25. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
26. Rao G, Burke LE, Spring BJ, Ewing LJ, Turk M, Lichtenstein AH, American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity and Metabolism, Council on Clinical Cardiology, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, Stroke Council. New and emerging weight management strategies for busy ambulatory settings: a scientific statement from the American Heart Association endorsed by the Society of Behavioral Medicine. Circulation 2011 Sep 06;124(10):1182-1203. [doi: 10.1161/CIR.0b013e31822b9543] [Medline: 21824925]
27. Centis E, Moscatiello S, Bugianesi E, Bellentani S, Fracanzani AL, Calugi S, et al. Stage of change and motivation to healthier lifestyle in non-alcoholic fatty liver disease. J Hepatol 2013 Apr;58(4):771-777. [doi: 10.1016/j.jhep.2012.11.031] [Medline: 23201248]
28. Tuong W, Larsen ER, Armstrong AW. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. J Behav Med 2014 Apr;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]
29. Chance Z, Gorlin M, Dhar R. Why Choosing Healthy Foods is Hard, and How to Help: Presenting the 4Ps Framework for Behavior Change. Cust. Need. and Solut 2014 Sep 12;1(4):253-262 [FREE Full text] [doi: 10.1007/s40547-014-0025-9]
30. 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]
31. American Heart Association. Dallas, Texas, United States: American Heart Association; 2018 Apr 18. American Heart Association Recommendations for Physical Activity in Adults and Kids URL: https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults [accessed 2020-03-02]
32. World Health Organization. Waist circumference and waist-hip ratio: Report of a WHO expert consultation, Geneva, 8-11 December 2008. In: Waist Circumference and Waist-hip Ratio. Geneva, Switzerland: World Health Organization; 2011.
33. Spring B, Duncan JM, Janke EA, Kozak AT, McFadden HG, DeMott A, et al. Integrating technology into standard weight loss treatment: a randomized controlled trial. JAMA Intern Med 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1221] [Medline: 23229890]
34. Mazzotti A, Caletti MT, Brodosi L, Di Domizio S, Forchielli ML, Petta S, et al. An internet-based approach for lifestyle changes in patients with NAFLD: Two-year effects on weight loss and surrogate markers. J Hepatol 2018 Nov;69(5):1155-1163 [FREE Full text] [doi: 10.1016/j.jhep.2018.07.013] [Medline: 30290973]
35. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease. Obes Facts 2016 Apr 8;9(2):65-90 [FREE Full text] [doi: 10.1159/000443344] [Medline: 27055256]
36. Vilar-Gomez E, Martinez-Perez Y, Calzadilla-Bertot L, Torres-Gonzalez A, Gra-Oramas B, Gonzalez-Fabian L, et al. Weight Loss Through Lifestyle Modification Significantly Reduces Features of Nonalcoholic Steatohepatitis. Gastroenterology 2015 Aug;149(2):367-378. [doi: 10.1053/j.gastro.2015.04.005] [Medline: 25865049]
37. Lazo M, Solga SF, Horska A, Bonekamp S, Diehl AM, Brancati FL, Fatty Liver Subgroup of the Look AHEAD Research Group. Effect of a 12-month intensive lifestyle intervention on hepatic steatosis in adults with type 2 diabetes. Diabetes Care 2010 Oct;33(10):2156-2163 [FREE Full text] [doi: 10.2337/dc10-0856] [Medline: 20664019]
38. Reis TO, Ferolla SM, Lima MLP, Fausto MA, Albricker ACL, Armiliato GNDA, et al. Nonalcoholic fatty liver disease: a cohort study focusing on treatment response to nutritional counseling. Medical Express 2015;2(2). [doi: 10.5935/MedicalExpress.2015.02.04]

39. Jiandani D, Wharton S, Rotondi MA, Ardern CI, Kuk JL. Predictors of early attrition and successful weight loss in patients attending an obesity management program. BMC Obes 2016;3:14 [FREE Full text] [doi: 10.1186/s40608-016-0098-0] [Medline: 26966544]
40. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]
41. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
42. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
43. Oza N, Eguchi Y, Mizuta T, Ishibashi E, Kitajima Y, Horie H, et al. A pilot trial of body weight reduction for nonalcoholic fatty liver disease with a home-based lifestyle modification intervention delivered in collaboration with interdisciplinary medical staff. J Gastroenterol 2009;44(12):1203-1208. [doi: 10.1007/s00535-009-0115-x] [Medline: 19728009]
44. Romero-Gómez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. J Hepatol 2017 Oct;67(4):829-846. [doi: 10.1016/j.jhep.2017.05.016] [Medline: 28545937]
45. Fakih El Khoury C, Karavetian M, Halfens RJG, Crutzen R, Khoja L, Schols JMGA. The Effects of Dietary Mobile Apps on Nutritional Outcomes in Adults with Chronic Diseases: A Systematic Review and Meta-Analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651. [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
46. Straznicky N, Lambert E, Grima M, Eikelis N, Nestel P, Dawood T, et al. The effects of dietary weight loss with or without exercise training on liver enzymes in obese metabolic syndrome subjects. Diabetes Obes Metab 2012 Feb;14(2):139-148. [doi: 10.1111/j.1463-1326.2011.01497.x] [Medline: 21923735]
47. Lim U, Ernst T, Buchthal SD, Latch M, Albright CL, Wilkens LR, et al. Asian women have greater abdominal and visceral adiposity than Caucasian women with similar body mass index. Nutr Diabetes 2011 May 09;1:e6 [FREE Full text] [doi: 10.1038/nutd.2011.2] [Medline: 23449381]
48. Park YW, Allison DB, Heymsfield SB, Gallagher D. Larger amounts of visceral adipose tissue in Asian Americans. Obes Res 2001 Jul;9(7):381-387 [FREE Full text] [doi: 10.1038/oby.2001.49] [Medline: 11445659]
49. Martin CK, Miller AC, Thomas DM, Champagne CM, Han H, Church T. Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial. Obesity (Silver Spring) 2015 May 28;23(5):935-942. [doi: 10.1002/oby.21063] [Medline: 25919921]
50. Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 2003 Nov;42(5):878-884. [doi: 10.1161/01.HYP.0000094221.86888.AE] [Medline: 12975389]
51. Smith A. Pew Research Center. Washington DC; 2013. Smartphone Ownership - 2013 Update URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_Smartphone_adoption_2013_PDF.pdf [accessed 2020-02-24]
52. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Hardy D, Smith SA. Smartphone Applications for Promoting Healthy Diet and Nutrition: A Literature Review. Jacobs J Food Nutr 2015;2(3):021 [FREE Full text] [Medline: 26819969]

Abbreviations

ALT: alanine aminotransferase
AST: aspartate aminotransferase
BMI: body mass index
CONSORT: Consolidated Standards of Reporting Trials
NAFLD: non-alcoholic fatty liver disease
NASH: non-alcoholic steatohepatitis
nBuddy: Nutritionist Buddy
NUH: National University Hospital
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.12.19; peer-reviewed by C Fakih El Khoury, M Wong, J Zhang; comments to author 16.01.20; revised version received 09.03.20; accepted 20.03.20; published 13.04.20.

Please cite as:
Lim SL, Johal J, Ong KW, Han CY, Chan YH, Lee YM, Loo WM
Lifestyle Intervention Enabled by Mobile Technology on Weight Loss in Patients With Nonalcoholic Fatty Liver Disease: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e14802
URL: http://mhealth.jmir.org/2020/4/e14802/
doi:10.2196/14802
PMID:32281943

Original Paper

Alcohol Avoidance Training as a Mobile App for Problem Drinkers: Longitudinal Feasibility Study

Melissa C Laurens[1,2], MSc; Marcel E Pieterse[1], PhD; Marjolein Brusse-Keizer[3], PhD; Elske Salemink[4,5], PhD; Somaya Ben Allouch[2,6], PhD; Ernst T Bohlmeijer[1], PhD; Marloes G Postel[1,7], PhD

[1]Centre for eHealth and Well-being Research, University of Twente, Enschede, Netherlands
[2]Research Group Technology, Health & Care, Saxion University of Applied Sciences, Enschede, Netherlands
[3]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands
[4]Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
[5]Department of Clinical Psychology, University of Utrecht, Utrecht, Netherlands
[6]Digital Life, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[7]Tactus Addiction Treatment, Enschede, Netherlands

Corresponding Author:
Melissa C Laurens, MSc
Centre for eHealth and Well-being Research
University of Twente
PO Box 217
Enschede, 7500 AE
Netherlands
Phone: 31 626990064
Email: m.c.laurens@utwente.nl

Abstract

Background: Alcohol use is associated with an automatic tendency to approach alcohol, and the retraining of this tendency (cognitive bias modification [CBM]) shows therapeutic promise in clinical settings. To improve access to training and to enhance participant engagement, a mobile version of alcohol avoidance training was developed.

Objective: The aims of this pilot study were to assess (1) adherence to a mobile health (mHealth) app; (2) changes in weekly alcohol use from before to after training; and (3) user experience with regard to the mHealth app.

Methods: A self-selected nonclinical sample of 1082 participants, who were experiencing problems associated with alcohol, signed up to use the alcohol avoidance training app Breindebaas for 3 weeks with at least two training sessions per week. In each training session, 100 pictures (50 of alcoholic beverages and 50 of nonalcoholic beverages) were presented consecutively in a random order at the center of a touchscreen. Alcoholic beverages were swiped upward (away from the body), whereas nonalcoholic beverages were swiped downward (toward the body). During approach responses, the picture size increased to mimic an approach movement, and conversely, during avoidance responses, the picture size decreased to mimic avoidance. At baseline, we assessed sociodemographic characteristics, alcohol consumption, alcohol-related problems, use of other substances, self-efficacy, and craving. After 3 weeks, 37.89% (410/1082) of the participants (posttest responders) completed an online questionnaire evaluating adherence, alcohol consumption, and user satisfaction. Three months later, 19.03% (206/1082) of the participants (follow-up responders) filled in a follow-up questionnaire examining adherence and alcohol consumption.

Results: The 410 posttest responders were older, were more commonly female, and had a higher education as compared with posttest dropouts. Among those who completed the study, 79.0% (324/410) were considered adherent as they completed four or more sessions, whereas 58.0% (238/410) performed the advised six or more training sessions. The study identified a significant reduction in alcohol consumption of 7.8 units per week after 3 weeks (95% CI 6.2-9.4, P<.001; n=410) and another reduction of 6.2 units at 3 months for follow-up responders (95% CI 3.7-8.7, P<.001; n=206). Posttest responders provided positive feedback regarding the fast-working, simple, and user-friendly design of the app. Almost half of the posttest responders reported gaining more control over their alcohol use. The repetitious and nonpersonalized nature of the intervention was suggested as a point for improvement.

Conclusions: This is one of the first studies to employ alcohol avoidance training in a mobile app for problem drinkers. Preliminary findings suggest that a mobile CBM app fulfils a need for problem drinkers and may contribute to a reduction in alcohol use. Replicating these findings in a controlled study is warranted.

(JMIR Mhealth Uhealth 2020;8(4):e16217) doi:10.2196/16217

KEYWORDS

mobile alcohol avoidance training; approach bias; cognitive bias modification; alcohol

Introduction

Problematic alcohol use is one of the most prevalent health problems in modern life. It has several negative personal, social, and economic consequences [1-4]. When not addressed properly and timely, problematic alcohol use can result in alcohol use disorder (AUD). Regular treatment of AUD and support for reducing problematic alcohol use, such as cognitive behavioral therapy and motivational interviewing, primarily focus on influencing controlled cognitive mechanisms. Although these treatments have proven to be effective [5,6], long-term outcomes remain modest [7]. To achieve progress in the effectiveness of treatments, research should further investigate the role played by relatively automatic processes. The dual process model [8,9] integrates both relatively slow reflective processes and fast impulsive processes.

Cognitive bias modification (CBM) programs have been developed to influence these impulsive processes by, for example, changing biases in action tendencies [10]. Research demonstrates that problem drinkers have an approach bias for alcohol-related stimuli [11]. Different CBM programs have been developed to directly influence the approach bias, for example, the stimulus response compatibility task [11] where participants are required to make a symbolic approach/avoidance movement to pictures and alcohol avoidance training, which is an adapted version of the alcohol approach avoidance task (A-AAT) [12]. In alcohol avoidance training, participants respond to either alcoholic or nonalcoholic pictures of beverages on a screen by pulling toward or pushing away the pictures using a joystick or keyboard. An important feature of alcohol avoidance training is the zooming function, which follows the pushing or pulling movement, creating the sensation of the beverage moving either away or toward the user. The use of alcohol avoidance training has shown positive results in a clinical setting [13], where receiving four sessions of alcohol avoidance training displayed a long-term clinical effect in alcohol-dependent patients (n=214) when added to their regular treatment. This study and a large replication study (n=509) [14] illustrated significant reductions in relapse a year after treatment (13%, *P*=.05 and 10%, *P*=.04, respectively) in the CBM condition as compared with a placebo condition. This effect was found to be mediated by a change in approach tendencies in the latter study [15]. Additionally, a recent study comparing different combinations of approach bias and attention bias retraining to “sham” or no training with 1405 alcohol-dependent patients obtained a somewhat smaller but significant result (*P*=.04), showing on average a 8.4% higher success rate 1 year after treatment; however, it did not confirm the mediating effect of the change in approach tendencies on the outcome [16].

Nowadays, most bias modification training programs are offered in a laboratory setting, clinical setting, or online via a computer. Although transferring treatment from a face-to-face setting to a mobile setting could be accompanied by lower patient engagement and higher dropout rates [17], online training programs have the advantage that participants can use the intervention independent of time and place [18], thus making it particularly suitable for outpatient treatment. For example, Wiers et al [19] conducted a Web-based CBM study on self-selected problem drinkers (n=136). Participants in the different conditions (including the control condition) of the study reduced their alcohol intake by 2.31 to 9.94 units per week [19]. However, having to log onto a computer or laptop for every training session may hinder motivation to train [20]. As most adults use a smartphone or tablet daily [21] and other forms of CBM training are operated by a joystick or keyboard, offering CBM training on a mobile device is an intuitive next step. Delivering CBM training this way facilitates more frequent training, as it allows participants to perform each session anywhere and anytime and may therefore promote engagement. A small study by Boendermaker et al [22] found support for this assumption, as participants (young and regular drinkers, not specifically selected on the basis of their motivation to reduce their drinking behavior) appeared to be more involved with CBM training when using a smartphone version of CBM training than when training on a computer. Until now, however, little is known about the use and evaluation of mobile CBM training in people who are willing to change their drinking behavior.

In the present study, a smartphone/tablet version of alcohol avoidance training was tested among a self-selected sample of Dutch problem drinkers from the general population. The aims of this study were to (1) measure adherence to mobile alcohol avoidance training; (2) determine the change in weekly alcohol use from before to after training; and (3) assess user experience.

Methods

Design

This pilot study consisted of a single group design with the following three measurements: baseline measurement, postintervention assessment at 3 weeks, and follow-up assessment at 3 months. The study was approved by the ethics committee within the faculty of Behavioral Management & Social Sciences of the University of Twente (approval number: BCE16395).

Participants

Participants were recruited between November 10 and November 23, 2016, via free publicity in national and regional

newspapers and on radio stations and television. A total of 1214 participants signed up for the study. To be included, participants had to (1) be willing to reduce/stop their drinking habit or be concerned about their drinking habit; (2) be aged 18 years or older; (3) have access to and ability to use the internet via a smartphone or tablet; (4) have the ability to read and write in Dutch; and (5) provide (online) informed consent.

Intervention

The Breindebaas app (Figure 1) is a mobile version of alcohol avoidance training [13,14], which is an adapted version of the AAT [23]. The mobile version distinguishes itself from the original (joystick operated) and online (keyboard operated) versions of alcohol avoidance training by (1) using swiping movements on the screen (directly touching the picture and swiping it away with a finger) and (2) asking the participant to react to the actual content of the picture (relevant feature) instead of the orientation of the picture (irrelevant feature). Every session contained 100 pictures (drinks only, without context) from the Amsterdam Beverage Picture Set [24], and of these, half depicted alcoholic beverages and the other half depicted nonalcoholic beverages. Participants were instructed to respond to these pictures by swiping the alcoholic beverages away from them and swiping the nonalcoholic beverages toward them. Participants were encouraged to swipe as quickly and accurately as possible. If a mistake was made, such as reacting too slowly and not completing the "swipe movement" correctly, participants received a short error notification (text and sound) with instructions. When swiping correctly, a sound notified participants of their correct response. After every 20, 50, and 80 pictures, participants received an encouraging message on the screen of their device, such as "you're well on your way" or "almost there!." These messages were included to motivate participants to complete their training session. Between every two pictures, there was a 1-second interval. The time interval between an encouraging message and the next picture was 2 seconds. After every session, the participants received an overview of their score regarding their average time and percentage of correct responses.

Measures

Table 1 presents an overview of the characteristics and measures assessed at baseline, postassessment, and follow-up. Internal consistency of scales was assessed with Cronbach alpha, with values ≥.7 being considered as acceptable [25].

Figure 1. Example of the Breindebaas app.

Table 1. Characteristics and measures at baseline, postassessment, and follow-up.

	Baseline	Postassessment	Follow-up
Sociodemographics	√		
TLFB[a]	√	√	√
AUDIT[b]	√		
Other substances	√		
OCDS[c]	√		
DRSEQ[d]	√		
CSQ[e]		√	
User friendliness		√	
Treatment history		√	√
Use of the app + reasons		√	√

[a]TLFB: time-line follow-back.

[b]AUDIT: Alcohol Use Identification Test.

[c]OCDS: Obsessive Compulsive Drinking Scale.

[d]DRSEQ: Drinking Refusal Self-Efficacy Questionnaire.

[e]CSQ: Client Satisfaction Questionnaire.

Sociodemographic Characteristics

Participants reported their gender, birth date, source of income, daily occupation, educational level, and smartphone/tablet information (type, brand, and model).

Alcohol Consumption

The Dutch adaptation [26] of the self-reported time-line follow-back (TLFB) procedure [27] was used to assess alcohol consumption. Participants indicated the number of standard units of alcohol consumed throughout each day over the past week. The total score of the scale was calculated by the total sum of all 7 days [27]. TLFB is a highly used retrospective estimation measure, and it has been used with similar target groups [28], with adequate validity [29].

Alcohol-Related Problems

The 10-item Dutch version [30] of the Alcohol Use Disorders Identification Test (AUDIT) [31] was used to assess drinking behaviors and alcohol-related problems. Internal consistency in the current sample was acceptable with a Cronbach alpha of .8.

Use of Other Substances

Participants were asked about their prior use over the past year and current use of other substances (tobacco, cannabis, cocaine, lysergic acid diethylamide, amphetamines, XTC, GHB, opiates, benzodiazepines, and others).

Drinking Refusal Self-Efficacy

Items of the Drinking Refusal Self-Efficacy Questionnaire (DRSEQ) [32] were used to assess the following three dimensions of self-efficacy in relation to refusal of alcohol, as identified by Young et al [33]: social pressure, emotional relief, and opportunistic drinking. The original DRSEQ contains a total of 31 items for these three dimensions, and it has shown good psychometric quality, both for the subscales and for the total scale (Cronbach alpha >.8) [34]. For this study, a short measure of self-efficacy was constructed using three items from both the dimensions of social pressure and emotional relief and two items from the dimension of opportunistic drinking, representing the items that were most relevant for our study. Cronbach alpha for these eight items was .9 in the current sample.

Craving

Using the five-item scale [35] of the original 14-item Obsessive Compulsive Drinking Scale [36], participants were asked to rate their thoughts, feelings, and actions concerning alcohol. Cronbach alpha in the current sample was .74.

Adherence

Adherence was measured by recoding the self-reported number of sessions. Participants were advised to complete at least six training sessions; these were established as the mean optimum in a study by Eberl et al [37]. In our study, completing four or more of the advised six training sessions was considered as adherent, as research by Wiers et al previously showed this to have a significant effect (P=.05) [13].

User Satisfaction

User satisfaction regarding the CBM alcohol avoidance training was assessed using the Dutch version of the eight-item Client Satisfaction Questionnaire (CSQ) [38]. Cronbach alpha was .91.

User Experience

Participants were invited to answer several questions about their experience with the Breindebaas app. Questions concerned its overall impression, benefits and drawbacks in using the app, suggestions for future development, main reason behind using the app, technical problems, use of other alcohol intervention

treatments during the intervention period, place of using the app (eg, at home, at work, and in the pub), and intention to continue using the app in the future. Lastly, participants were questioned on concentration, which was measured by simply inquiring about their general concentration during a training session, on a 4-point scale (not concentrated at all to very much concentrated), which was converted to a dichotomous variable (1-2: not concentrated; 3-4: concentrated) for analyses. At follow-up, participants were asked whether they had kept on using the app and their reasoning behind this decision. They were also asked whether they used other forms of help associated with their alcohol use during the research period.

Procedure

Participants were referred to a website [39], where information about the study and the app was provided. Participants who demonstrated an interest were then asked to fill out a digital informed consent form and an online baseline questionnaire. Upon completion, instructions for downloading the app and an access code needed for using the app were provided digitally. Participants were requested to complete at least two training sessions every week for 3 weeks, leaving at least 24 hours between two sessions. Three days after completing a session, participants received an alert that a new training session was available. If participants did not finish a training session within 5 days, a reminder was sent via push message. Optionally, participants could choose to receive these messages via short message service text messaging. The link to the postassessment questionnaire was sent via email 3 weeks after the start of the training. Participants were not given specific instructions to keep on using the app after the postassessment. Three months after completing the postassessment questionnaire, participants were asked to fill out a follow-up questionnaire. A reminder was sent by email or short message service text messaging 1 week later. By completing all three questionnaires, participants had a chance of winning one of five available gift vouchers, each worth 100 euros.

Statistical Analysis

Descriptive statistics were used to describe the baseline characteristics of the participants and the characteristics of those who completed the training at posttest. Means and standard deviations (SDs) or medians and interquartile ranges (IQRs) are provided depending on the normality distribution for continuous variables. Categorical variables are presented as numbers with corresponding percentages. Independent samples *t*-tests or Wilcoxon rank-sum tests (continuous variables) and chi-square tests or Fisher exact tests (categorical variables) were used to compare baseline characteristics between posttest responders and posttest dropouts, as well as between follow-up responders and follow-up dropouts. A paired samples *t*-test was performed to compare alcohol consumption at baseline and posttest. Linear regression analysis was performed to identify any predictors correlated with a change in alcohol consumption between baseline and posttest. Variables associated ($P \leq .15$) in univariate analysis were all entered in the multivariate model, and subsequently, they were eliminated step by step based on significance (backward elimination method).

Changes in alcohol consumption from baseline to postassessment at 3 weeks and to follow-up assessment at 3 months were analyzed using a mixed-model analysis. In case of significant changes over time, Sidak post-hoc analyses were performed to test which measurements were statistically significantly different. All tests were performed using SPSS version 24.0 (IBM Corp, Armonk, New York).

Results

Baseline Characteristics

In total, 1238 participants completed the baseline questionnaire. Of these, 156 participants were excluded owing to age <18 years (n=2), not signing the informed consent form (n=22), duplicate records (n=3), or having a nonalcohol-related reason to participate (n=129), eg, professional interest in the app. Thus, 1082 participants were included for analysis at baseline. Table 2 demonstrates the baseline characteristics of the 1082 participants. The sample contained slightly more male participants (58.4%), with an overall mean age of 49.9 (SD 11.3) years. The mean weekly alcohol consumption was 36.6 (SD 24.5) standard units. Among the participants, 93.53% (1012/1082) reported an AUDIT score ≥8, indicating problematic alcohol use throughout the vast majority of the sample.

Posttest Responders and Adherence

Among the original 1082 participants, 410 (37.89%) completed the postintervention assessment (referred to as posttest responders), with 206 participants (19.0%) also completing the follow-up assessment after 3 months (referred to as follow-up responders). Posttest responders (n=410) and posttest dropouts (n=672) were compared regarding baseline characteristics (Table 2). Posttest responders were significantly older ($P<.001$), were less often male ($P=.01$), had a higher education ($P<.001$), and consumed less alcohol ($P<.001$) as compared with posttest dropouts. This was mainly caused by lower consumption of alcohol by females in the completer group. Furthermore, posttest responders had lower AUDIT scores among both males and females and lower DRSEQ scores mainly among males. Finally, posttest responders used fewer other substances.

Table 2. Baseline characteristics and differences in baseline characteristics between posttest responders and posttest dropouts.

Variable	Total (n=1082)	Posttest responders (n=410)	Posttest dropouts (n=672)	Analysis *t* value	χ^2*P* value
Age (years), mean (SD)	49.89 (11.32)	52.4 (10.2)	48.3 (11.7)	5.80	<.001[a]
Gender, n (%)				6.14	.01[a]
Male	632 (58.4)	220 (53.7)	412 (61.3)	—[h]	—
Female	450 (41.6)	190 (46.3)	260 (38.7)	—	—
Employed, n (%)	726 (70.9)	271 (70.6)	455 (71.1)	0.03	.86
Education, n (%)				26.60	<.001[a]
High[b]	583 (57.0)	257 (66.8)	326 (51.1)	—	—
Middle[c]	286 (28.0)	91 (23.6)	195 (30.6)	—	—
Low[d]	154 (15.1)	37 (9.6)	117 (18.3)	—	—
Weekly alcohol consumption, mean (SD)	36.6 (24.5)	33.3 (21.8)	38.7 (25.8)	−3.69	<.001[a]
Male	42.4 (26.5)	40.0 (24.7)	43.7 (27.3)	−1.71	.09
Female	28.5 (18.7)	25.5 (14.4)	30.7 (21.0)	−3.08	<.001[a]
AUDIT [e], mean (SD)	17.2 (6.7)	15.8 (6.1)	18.0 (6.9)	−5.27	<.001[a]
Male	18.2 (6.3)	17.1 (6.0)	18.8 (6.5)	−3.13	<.001[a]
Female	15.7 (6.8)	14.4 (5.9)	16.7 (7.3)	−3.71	<.001[a]
DRSEQ [f], mean (SD)	25.4 (7.4)	24.8 (7.4)	25.8 (7.3)	−2.17	.03[a]
Male	25.0 (7.3)	24.1 (7.3)	25.4 (7.3)	−2.21	.03[a]
Female	26.0 (7.4)	25.6 (7.5)	26.4 (7.3)	−1.06	.29
OCDS[g], mean (SD)	5.3 (3.2)	5.2 (3.0)	5.4 (3.3)	−1.31	.19
Other substances, n (%)	452 (41.8)	141 (34.4)	311 (46.3)	14.8	<.001[a]
Tobacco	338 (31.2)	100 (24.4)	238 (35.4)	14.41	<.001[a]
Benzodiazepines	142 (13.1)	45 (11.0)	97 (14.4)	2.67	.10
Cannabis	108 (10.0)	24 (5.9)	84 (12.5)	12.52	<.001[a]
Others	209 (19.3)	—	—	—	—

[a]*P*<.05 (two-tailed).

[b]University of research or university of professional education.

[c]Higher general secondary education or intermediate vocational education.

[d]Primary school or lower vocational education.

[e]AUDIT: Alcohol Use Identification Test.

[f]DRSEQ: Drinking Refusal Self-Efficacy Questionnaire.

[g]OCDS: Obsessive Compulsive Drinking Scale.

[h]Not applicable.

In a similar analysis, among 410 participants remaining in the study at posttest, baseline characteristics were compared between follow-up responders (n=206) and follow-up dropouts (n=204). A significant difference in baseline characteristics was found regarding participant age, with follow-up responders tending to be older (mean 55.2 vs 52.5 years, *P*=.01), and their use of tobacco, with follow-up responders smoking significantly less (30.6% [63/206] vs 38.2% [78/204], *P*=.01). No other significant differences were found.

Participants reported completing from 1 to over 10 sessions in the questionnaire (median 6, IQR 4-7). Of the 410 posttest responders, 323 (78.8%) completed four or more sessions, which was considered adherent in this study. Furthermore, 239 (58.3%) participants performed the recommended six sessions (n=123, 30.0%) or more (n=116, 28.4%). The main reasons for not completing the recommended number of sessions were “it does not seem to help me” (21.0%, n=86) and “not having enough time” (19.0%, n=78).

Concentration while performing a session was recoded as a dichotomous variable (concentrated/not concentrated). Of the 410 posttest responders, 375 (91.5%) reported to be concentrated throughout their training sessions.

Changes in Alcohol Consumption Over Time and Predictors

The average alcohol consumption of the posttest responders (n=410) decreased significantly by 7.8 units per week (95% CI 6.2-9.4, *P*<.001) from baseline (mean 33.3 [SD 21.8]) to postassessment (mean 25.5 [SD 20.4]).

Table 3 illustrates the results of the regression analyses, which evaluated the potential predictors of changes in alcohol use. The following variables were found to be univariately associated with a stronger decrease in alcohol consumption (*P*<.15): male gender, unemployment, high level of baseline craving, high baseline AUDIT score, high level of self-reported concentration during sessions, and high adherence. When these factors were entered in a multivariate regression model, only gender, adherence, and craving remained significant predictors of a change in alcohol consumption.

Changes in Alcohol Use at Follow-up

The subsample of 206 participants that completed the follow-up assessment displayed a reduction in alcohol use over time. Their mean weekly alcohol consumption decreased from 31.6 (SD 23.2) units at baseline to 24.4 (SD 19.2) units at 3 weeks and to 18.2 (SD 17.3) units at follow-up 3 months later, resulting in a total decrease of 13.4 units a week. Pairwise comparisons in mixed-model analysis demonstrated the reductions for this subsample as significant both from baseline to postassessment (mean difference 7.2, CI 4.9-9.6, *P*<.001) and from postassessment to follow-up (mean difference 6.2, CI 3.7-8.7, *P*<.001).

A mixed model for repeated measurements, in which all 410 participants were taken into account, produced similar results (Table 4). Pairwise comparisons in mixed-model analysis also displayed the reductions as significant both from baseline to postassessment (mean difference 7.4, CI 5.7-9.6, *P*<.001) and from postassessment to follow-up (mean difference 6.6, CI 4.2-9.0, *P*<.001).

Table 3. Univariate and multivariate linear regression coefficients, confidence intervals, and significance levels of baseline characteristics with regard to alcohol consumption.

	Univariate coefficient	95% CI	*P* value	Multivariate coefficient	95% CI	*P* value
Gender						
Male	9.92	0.75-7.09	.02	4.44	1.36-7.52	.01
Female	Reference			Reference		
Concentration				N/A	N/A	N/A
No	Reference					
Yes	6.74	1.08-12.40	.02			
Adherence						
No	Reference			Reference		
Yes	5.69	1.83-9.55	<.001	6.09	2.33-9.85	<.001
Work situation				N/A	N/A	N/A
Unknown	Reference					
Paid	1.54	−2.06 to 5.14	.04			
Unpaid	6.67	0.07-13.27	.01			
OCDS[a]	1.15	0.63-1.68	<.001	1.15	0.64-1.67	<.001
AUDIT[b]	0.05	0.23-0.75	<.001	N/A	N/A	N/A

[a]OCDS: Obsessive Compulsive Drinking Scale.

[b]AUDIT: Alcohol Use Identification Test.

Table 4. Mean alcohol consumption at baseline, postassessment, and follow-up using mixed models.

	Mean alcohol consumption (SD)	
Measurement	Subsample of participants who completed the follow-up assessment (n=206)	Total participants (n=410)
Baseline	31.6 (1.6)	32.6 (0.9)
Postassessment	24.4 (1.3)	25.2 (0.9)
Follow-up	18.2 (1.2)	18.6 (1.0)

User Experiences

When posttest responders (n=410) were asked about what they gained from using the app over a 3-week period, almost half of the participants stated having the feeling of more control over their drinking (eg, gained more control over alcohol use, decided more frequently not to drink, and chose to drink alcohol less automatically), with many participants also reporting being more aware of their alcohol use (36.1%, 148/410). However, 47.3% (194/410) of participants reported that they gained nothing from using the app.

Posttest responders had an overall CSQ score of 20.9 (SD 4.4) with an average score of 2.6 on a scale from 1-4 (item variances: 0.5), indicating moderate satisfaction. Participants were particularly positive about the simple, fast-working, user-friendly design of the app. Criticism and subsequent suggestions about the app mostly targeted elements concerning monotony and lack of personalization. Participants deemed app use as boring and monotonous owing to the repetition of the task and pictures. They suggested introducing motivational elements, such as levels or game options, as well as a shorter interval between swiping movements and subsequent pictures. Thus, the introduction of a greater variation in pictures and the possibility of choosing pictures was suggested.

Of the 410 posttest responders, 318 (77.6%) had never sought help or used an intervention to reduce alcohol use previously. Additionally, 46 (11.2%) participants reported receiving extra help in reducing alcohol use during the Breindebaas training period in the form of a self-help program, help from a general practitioner, help from a professional in (addiction) care, or peer support. Of these 46 participants, 20 had never sought help before.

Of the 206 participants that completed the follow-up questionnaire, 85 continued to use the app. The main reasons behind this decision were that using the app helped them to be more conscious of their alcohol use (n=51) and it assisted in maintaining their reduced drinking habit (n=15). Of the 121 participants who stopped using the app, the main reasons were doubts regarding the app's effectiveness (n=40) and simply forgetting to use the app (n=33).

Discussion

To our knowledge, this study is the first to evaluate a mobile version of AAT training in a sample of problem drinkers among the general population. Given the debate on the effectiveness of CBM [40,41], it is essential to differentiate between experimental studies with students, which are set up to show that biases can be influenced (but do not always show a change in behavior), clinical trials with alcohol dependent patients who are motivated to change [42], and studies with nonclinical problem drinkers from the general public. Gaining more insight into the feasibility and outcomes of CBM for the participants in this study, who were not clinically diagnosed with AUD but were willing to change their drinking behavior, is therefore especially relevant.

The baseline characteristics, adherence to the intervention, change in alcohol consumption, and user experiences were studied. Participants in this study were comparable to groups analyzed in previous research via Web-based self-help interventions regarding the level of problematic alcohol use [42-44] and no active search for professional help to aid the reduction of their drinking behavior [45,46]. We were pleasantly surprised by the large group of problem drinkers interested in using the Breindebaas app, considering the short timeframe of the study. It can be considered a strength that this low-threshold application seems to appeal to this hard-to-reach group, as it may reduce the stigma associated with directly meeting a professional [47].

The intervention adherence among posttest responders was high. Most of the posttest responders (78.8%, 323/410) used the app four or more times, doing better than an online CBM trial by Wiers et al on alcohol, where 43.3% (136/314) of the participants completed the prescribed four sessions [19]. The fact that the Breindebaas app is a mobile version of an AAT and therefore available to participants at any moment could be a particular contributing factor. For example, a pilot study using a mobile CBM app on obesity found a training session completion rate of 86% (17/20) [48].

In this study, we observed a significant reduction in alcohol use among posttest responders immediately after using the mobile intervention and 3 months later. Although a reduction of 13 units per week is substantial for such a brief intervention, the results need to be seen in the context of a pilot study without a control condition for comparing the findings, especially given earlier observed main effects across CBM and control conditions [19]. Further research should be implemented, in which the app training should be compared with sham training in a controlled design. The same caution should be exercised with the impact of the predictors (gender, adherence, and craving) that were established for changing an individual's alcohol consumption.

Participants were mostly positive about the Breindebaas app. The simple, fast-working, user-friendly design resulted in participants reporting more awareness and control over their alcohol use. Nevertheless, a considerable portion of participants also reported gaining nothing from using the app. Reportedly, this was because of the repetitious and monotonous characteristics of the AAT and its lack of personalization. Regarding the lack of personalization, participants mainly indicated that some of the pictures contained beverages (both alcoholic and nonalcoholic) that were not appealing to them at all. Wiers et al already indicated that personalizing alcohol-related stimuli is a potential way forward [41]. Studies on CBM related to eating habits indicated that personalizing CBM tasks may increase attention, motivation, and interest and therefore may increase adherence [49,50]. An additional reason for withdrawing from using the Breindebaas app was the participant's questions and doubts pertaining to the working mechanism behind the training. Other studies support this finding [20], which may mean that for the future development of similar tasks, explaining the reasoning and the importance of repeated training is crucial.

A number of limitations should be addressed. First, as this study was set up as a pilot study with the aim to assess feasibility and adherence, no control group was allocated. Consequently, the

change in self-reported alcohol consumption found in this study may well be the result of a placebo effect of the app, a nonspecific effect of engaging in any intervention, or even an effect of participating in a study. As already mentioned, just reporting alcohol use alone can have an effect on the reduction of drinking [51]. Nonetheless, the change in alcohol consumption demonstrated by participants in this study seems large enough to justify future studies on the effectiveness of the Breindebaas app. Second, participants were invited to partake in this study via free publicity channels and were only asked to provide basic information about themselves. As none of the participants of this study had face-to-face interactions, we needed to rely on their self-reporting. This, of course, could decrease the reliability of our results, although the reliability of the measures was described as acceptable to good in this study. Self-reported alcohol use reduction among subjects participating in treatment is likely to be positively biased, overestimating outcomes. In addition to self-reporting, the fact that more than half of the participants who had originally signed up dropped out during the training period and only 19% completed the follow-up questionnaire may have decreased the generalizability of the results. The dropout attrition rate in this study was comparable to that in other online CBM studies on alcohol or smoking [19,52,53]. Given the design of the study, the most likely factors influencing the dropout rate (between baseline and follow-up) were (1) ease of enrolment; (2) ease of dropout; (3) no personal contact via face-to-face interviews or telephone; and (4) fully paid for intervention [54]. Finally, no approach bias measurements were made before and after using the app. Therefore, it is unknown whether the approach bias of participants changed over time or whether this mediated the effect of training on alcohol use. Previous studies have indicated that relevant treatment effects of CBM on clinical outcomes almost always are accompanied by a decrease in cognitive bias [55,56]. A study by Eberl et al showed that patients with a strong approach bias at baseline elicited the best results in decreasing their bias [14], whereas no overall approach bias was established in the sample. This might be associated with the ambivalent stance that many patients with AUD hold with respect to alcohol [57]. Future developments in mobile CBM applications and research should therefore consider incorporating bias measurements, providing more insight into the working mechanism of CBM in the subclinical population.

In summary, several suggestions from users and researchers provide the following insights for the future development of the Breindebaas app: (1) using personalized stimuli in the app; (2) adding more information about the working mechanism and effects of CBM, which can increase motivation; (3) including bias measurement in the app, so participant progress in bias scores can be tracked; and (4) adding motivational/gamification elements (eg, levels and rewards) to improve user adherence to the app. The addition of gamification elements was mentioned by users of the Breindebaas app and has shown promising results in other forms of cognitive training [22,58]. In addition to app development, more research on the effects of using the Breindebaas app in controlled trials is advised. One suggestion is a three-armed study, in which participants are assigned to either training with the Breindebaas app, a mobile intervention with self-monitoring and goal setting features, or a waiting-list condition. As the Breindebaas app contains relevant feature approach/avoidance training, developing a credible placebo version is very challenging. Using a different mobile intervention to rule out nonspecific effects seems like a pragmatic choice. Following up on this research, approaching the same target group (problem drinkers from the general population) would be consistent. This is the first study in which alcohol avoidance training was adapted to a mobile app for problem drinkers. User evaluation suggests that this CBM app fulfils a need for problem drinkers reluctant to seek clinical treatment, as the majority of the sample never sought help prior to the study and had been mostly positive about using the app. Participants in this study reduced their alcohol intake by a total of 13 units per week. Although the results should be interpreted cautiously owing to the absence of a control group, adoption of the CBM app may contribute to reducing alcohol use among those who experience problems associated with drinking.

Acknowledgments

The Breindebaas app has been developed in a collaboration among Tactus Addiction Treatment Institute, University of Twente, University of Amsterdam, Saxion University of Applied Sciences, and a creative industry partner (theFactor.e). Part of this research is supported by funding from Nederlandse Organisatie voor Wetenschappelijk Onderzoek Kennis Innovatie Mapping under contract number 314-98-048.

Conflicts of Interest

None declared.

References

1. Beaglehole R, Bonita R. Alcohol: a global health priority. Lancet 2009 Jun 27;373(9682):2173-2174. [doi: 10.1016/S0140-6736(09)61168-5] [Medline: 19560583]
2. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009 Jun 27;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7] [Medline: 19560604]
3. Anderson P, Baumberg B. Alcohol in Europe – Public Health Perspective: Report summary. Drugs: Education, Prevention and Policy 2006;13(6):483-488. [doi: 10.1080/09687630600902477]

4. Cargiulo T. Understanding the health impact of alcohol dependence. Am J Health Syst Pharm 2007 Mar 01;64(5 Suppl 3):S5-11. [doi: 10.2146/ajhp060647] [Medline: 17322182]
5. Vasilaki EI, Hosier SG, Cox WM. The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. Alcohol Alcohol 2006;41(3):328-335 [FREE Full text] [doi: 10.1093/alcalc/agl016] [Medline: 16547122]
6. Magill M, Ray LA. Cognitive-behavioral treatment with adult alcohol and illicit drug users: a meta-analysis of randomized controlled trials. J Stud Alcohol Drugs 2009 Jul;70(4):516-527 [FREE Full text] [Medline: 19515291]
7. Riper H, Andersson G, Hunter SB, de Wit J, Berking M, Cuijpers P. Treatment of comorbid alcohol use disorders and depression with cognitive-behavioural therapy and motivational interviewing: a meta-analysis. Addiction 2014 Mar;109(3):394-406. [doi: 10.1111/add.12441] [Medline: 24304463]
8. Stacy AW, Wiers RW. Implicit cognition and addiction: a tool for explaining paradoxical behavior. Annu Rev Clin Psychol 2010;6:551-575 [FREE Full text] [doi: 10.1146/annurev.clinpsy.121208.131444] [Medline: 20192786]
9. Wiers RW, Teachman BA, De Houwer J. Implicit cognitive processes in psychopathology: an introduction. J Behav Ther Exp Psychiatry 2007 Jun;38(2):95-104. [doi: 10.1016/j.jbtep.2006.10.002] [Medline: 17112462]
10. Wiers RW, Gladwin TE, Hofmann W, Salemink E, Ridderinkhof KR. Cognitive Bias Modification and Cognitive Control Training in Addiction and Related Psychopathology. Clinical Psychological Science 2013 Jan 09;1(2):192-212. [doi: 10.1177/2167702612466547]
11. Field M, Kiernan A, Eastwood B, Child R. Rapid approach responses to alcohol cues in heavy drinkers. J Behav Ther Exp Psychiatry 2008 Sep;39(3):209-218. [doi: 10.1016/j.jbtep.2007.06.001] [Medline: 17640615]
12. Wiers RW, Rinck M, Kordts R, Houben K, Strack F. Retraining automatic action-tendencies to approach alcohol in hazardous drinkers. Addiction 2010 Feb;105(2):279-287. [doi: 10.1111/j.1360-0443.2009.02775.x] [Medline: 20078486]
13. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011 Apr;22(4):490-497. [doi: 10.1177/0956797611400615] [Medline: 21389338]
14. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
15. Gladwin TE, Rinck M, Eberl C, Becker ES, Lindenmeyer J, Wiers RW. Mediation of cognitive bias modification for alcohol addiction via stimulus-specific alcohol avoidance association. Alcohol Clin Exp Res 2015 Jan;39(1):101-107. [doi: 10.1111/acer.12602] [Medline: 25623410]
16. Rinck M, Wiers RW, Becker ES, Lindenmeyer J. Relapse prevention in abstinent alcoholics by cognitive bias modification: Clinical effects of combining approach bias modification and attention bias modification. J Consult Clin Psychol 2018 Dec;86(12):1005-1016. [doi: 10.1037/ccp0000321] [Medline: 30507226]
17. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter A. The Advantages and Disadvantages of Online and Blended Therapy: Survey Study Amongst Licensed Psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]
18. Wykes T, Haro JM, Belli SR, Obradors-Tarragó C, Arango C, Ayuso-Mateos JL, ROAMER Consortium. Mental health research priorities for Europe. Lancet Psychiatry 2015 Nov;2(11):1036-1042. [doi: 10.1016/S2215-0366(15)00332-6] [Medline: 26404415]
19. Wiers RW, Houben K, Fadardi JS, van Beek P, Rhemtulla M, Cox WM. Alcohol cognitive bias modification training for problem drinkers over the web. Addict Behav 2015 Jan;40:21-26. [doi: 10.1016/j.addbeh.2014.08.010] [Medline: 25218067]
20. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
21. Deloitte. Global Mobile Consumer Survey 2018. 2018. URL: https://www2.deloitte.com/ [accessed 2020-02-05]
22. Boendermaker WJ, Boffo M, Wiers RW. Exploring Elements of Fun to Motivate Youth to Do Cognitive Bias Modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
23. Wiers RW, Rinck M, Dictus M, van den Wildenberg E. Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. Genes Brain Behav 2009 Feb;8(1):101-106 [FREE Full text] [doi: 10.1111/j.1601-183X.2008.00454.x] [Medline: 19016889]
24. Pronk T, van Deursen D, Beraha EM, Larsen H, Wiers RW. Validation of the Amsterdam Beverage Picture Set: A Controlled Picture Set for Cognitive Bias Measurement and Modification Paradigms. Alcohol Clin Exp Res 2015 Oct;39(10):2047-2055. [doi: 10.1111/acer.12853] [Medline: 26431117]
25. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
26. Wiers RW, Hoogeveen KJ, Sergeant JA, Boudewijn Gunning W. High- and low-dose alcohol-related expectancies and the differential associations with drinking in male and female adolescents and young adults. Addiction 1997 Jul;92(7):871-888. [Medline: 9293046]

27. Sobell LC, Sobell MB. Timeline Follow-Back. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption. Totowa, NJ: Humana Press; 1992.
28. van Deursen DS, Salemink E, Boendermaker WJ, Pronk T, Hofmann W, Wiers RW. Executive Functions and Motivation as Moderators of the Relationship Between Automatic Associations and Alcohol Use in Problem Drinkers Seeking Online Help. Alcohol Clin Exp Res 2015 Sep;39(9):1788-1796. [doi: 10.1111/acer.12822] [Medline: 26247799]
29. Dulin PL, Alvarado CE, Fitterling JM, Gonzalez VM. Comparisons of Alcohol Consumption by Time-Line Follow Back vs. Smartphone-Based Daily Interviews. Addict Res Theory 2017;25(3):195-200 [FREE Full text] [doi: 10.1080/16066359.2016.1239081] [Medline: 29170622]
30. Schippers GM, Broekman TG. MATE: Measurements in the Addictions for Triage and Evaluation. 2010. De AUDIT: Nederlandse vertaling van de Alcohol Use Disorders Identification Test. URL: https://www.mateinfo.nl/audit/audit-nl.pdf
31. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. World Health Organization. 2001. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care, Second Edition URL: https://tinyurl.com/vbbf9ls [accessed 2020-02-05]
32. Oei TP, Hasking PA, Young RM. Drinking refusal self-efficacy questionnaire-revised (DRSEQ-R): a new factor structure with confirmatory factor analysis. Drug Alcohol Depend 2005 Jun 01;78(3):297-307. [doi: 10.1016/j.drugalcdep.2004.11.010] [Medline: 15893161]
33. Young RM, Oei TP, Crook GM. Development of a drinking self-efficacy questionnaire. J Psychopathol Behav Assess 1991 Mar;13(1):1-15. [doi: 10.1007/bf00960735]
34. Young RM, Oei TP. Drinking expectancy profile: a test manual. Brisbane, Australia: Behaviour Research & Therapy Centre (BRTC) The University of Queensland; 1996.
35. de Wildt WA, Lehert P, Schippers GM, Nakovics H, Mann K, van den Brink W. Investigating the structure of craving using structural equation modeling in analysis of the obsessive-compulsive drinking scale: a multinational study. Alcohol Clin Exp Res 2005 Apr;29(4):509-516. [doi: 10.1097/01.alc.0000158844.35608.48] [Medline: 15834215]
36. Anton RF, Moak DH, Latham P. The Obsessive Compulsive Drinking Scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. Alcohol Clin Exp Res 1995 Feb;19(1):92-99. [doi: 10.1111/j.1530-0277.1995.tb01475.x] [Medline: 7771669]
37. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Implementation of approach bias re-training in alcoholism-how many sessions are needed? Alcohol Clin Exp Res 2014 Feb;38(2):587-594. [doi: 10.1111/acer.12281] [Medline: 24164417]
38. De Wilde EF, Hendriks VM. The Client Satisfaction Questionnaire: psychometric properties in a Dutch addict population. Eur Addict Res 2005;11(4):157-162. [doi: 10.1159/000086396] [Medline: 16110221]
39. Tactus. URL: http://www.breindebaasapp.nl [accessed 2016-01-11]
40. Cristea IA, Kok RN, Cuijpers P. The Effectiveness of Cognitive Bias Modification Interventions for Substance Addictions: A Meta-Analysis. PLoS One 2016;11(9):e0162226 [FREE Full text] [doi: 10.1371/journal.pone.0162226] [Medline: 27611692]
41. Wiers RW, Boffo M, Field M. What's in a Trial? On the Importance of Distinguishing Between Experimental Lab Studies and Randomized Controlled Trials: The Case of Cognitive Bias Modification and Alcohol Use Disorders. J Stud Alcohol Drugs 2018 May;79(3):333-343. [Medline: 29885138]
42. Saitz R, Helmuth ED, Aromaa SE, Guard A, Belanger M, Rosenbloom DL. Web-based screening and brief intervention for the spectrum of alcohol problems. Prev Med 2004 Nov;39(5):969-975. [doi: 10.1016/j.ypmed.2004.04.011] [Medline: 15475031]
43. Koski-Jännes A, Cunningham JA, Tolonen K, Bothas H. Internet-based self-assessment of drinking-3-month follow-up data. Addict Behav 2007 Mar;32(3):533-542. [doi: 10.1016/j.addbeh.2006.05.020] [Medline: 16822619]
44. Riper H, Kramer J, Conijn B, Smit F, Schippers G, Cuijpers P. Translating effective web-based self-help for problem drinking into the real world. Alcohol Clin Exp Res 2009 Aug;33(8):1401-1408. [doi: 10.1111/j.1530-0277.2009.00970.x] [Medline: 19413646]
45. Hester RK, Lenberg KL, Campbell W, Delaney HD. Overcoming Addictions, a Web-based application, and SMART Recovery, an online and in-person mutual help group for problem drinkers, part 1: three-month outcomes of a randomized controlled trial. J Med Internet Res 2013 Jul 11;15(7):e134 [FREE Full text] [doi: 10.2196/jmir.2565] [Medline: 23846588]
46. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009 Dec;104(12):2023-2032 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
47. West R, Michie S. Development and evaluation of digital interventions in healthcare. In: A Guide To Development And Evaluation Of Digital Behaviour Change Interventions In Healthcare. London: Silverback Publishing; 2016.
48. Kakoschke N, Hawker C, Castine B, de Courten B, Verdejo-Garcia A. Smartphone-based cognitive bias modification training improves healthy food choice in obesity: A pilot study. Eur Eat Disord Rev 2018 Sep;26(5):526-532. [doi: 10.1002/erv.2622] [Medline: 30003634]

49. Forman EM, Goldstein SP, Flack D, Evans BC, Manasse SM, Dochat C. Promising technological innovations in cognitive training to treat eating-related behavior. Appetite 2018 May 01;124:68-77 [FREE Full text] [doi: 10.1016/j.appet.2017.04.011] [Medline: 28414042]
50. Stice E, Lawrence NS, Kemps E, Veling H. Training motor responses to food: A novel treatment for obesity targeting implicit processes. Clin Psychol Rev 2016 Nov;49:16-27. [doi: 10.1016/j.cpr.2016.06.005] [Medline: 27498406]
51. Miller WR, Brown JM, Simpson TL, Handmaker NS, Bien TH, Luckie LF, et al. What works? A methodological analysis of the alcohol treatment outcome literature. In: Handbook of Alcoholism Treatment Approaches: Effective Alternatives. Boston: Allyn & Bacon; 1995:12-44.
52. Elfeddali I, de Vries H, Bolman C, Pronk T, Wiers RW. A randomized controlled trial of Web-based Attentional Bias Modification to help smokers quit. Health Psychol 2016 Aug;35(8):870-880. [doi: 10.1037/hea0000346] [Medline: 27505210]
53. Wittekind CE, Lüdecke D, Cludius B. Web-based Approach Bias Modification in smokers: A randomized-controlled study. Behav Res Ther 2019 May;116:52-60. [doi: 10.1016/j.brat.2018.12.003] [Medline: 30782522]
54. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
55. Gladwin TE, Wiers CE, Wiers RW. Interventions aimed at automatic processes in addiction: considering necessary conditions for efficacy. Current Opinion in Behavioral Sciences 2017 Feb;13:19-24. [doi: 10.1016/j.cobeha.2016.08.001]
56. Grafton B, MacLeod C, Rudaizky D, Holmes EA, Salemink E, Fox E, et al. Confusing procedures with process when appraising the impact of cognitive bias modification on emotional vulnerability. Br J Psychiatry 2017 Nov 2;211(5):266-271. [doi: 10.1192/bjp.bp.115.176123] [Medline: 29092835]
57. Barkby H, Dickson JM, Roper L, Field M. To approach or avoid alcohol? Automatic and self-reported motivational tendencies in alcohol dependence. Alcohol Clin Exp Res 2012 Feb;36(2):361-368 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01620.x] [Medline: 21895719]
58. Prins PJ, Brink ET, Dovis S, Ponsioen A, Geurts HM, de Vries M, et al. "Braingame Brian": Toward an Executive Function Training Program with Game Elements for Children with ADHD and Cognitive Control Problems. Games Health J 2013 Feb;2(1):44-49. [doi: 10.1089/g4h.2013.0004] [Medline: 26196554]

Abbreviations

AAT: approach avoidance task
AUD: alcohol use disorder
AUDIT: Alcohol Use Disorders Identification Test
CBM: cognitive bias modification
CSQ: Client Satisfaction Questionnaire
DRSEQ: Drinking Refusal Self-Efficacy Questionnaire
IQR: interquartile range
SD: standard deviation
TLFB: time-line follow-back

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by M Rinck, B Bock; comments to author 22.10.19; revised version received 02.12.19; accepted 15.12.19; published 14.04.20.

Please cite as:
Laurens MC, Pieterse ME, Brusse-Keizer M, Salemink E, Ben Allouch S, Bohlmeijer ET, Postel MG
Alcohol Avoidance Training as a Mobile App for Problem Drinkers: Longitudinal Feasibility Study
JMIR Mhealth Uhealth 2020;8(4):e16217
URL: https://mhealth.jmir.org/2020/4/e16217
doi:10.2196/16217
PMID:32286235

Original Paper

Automated Mobile Delivery of Financial Incentives for Smoking Cessation Among Socioeconomically Disadvantaged Adults: Feasibility Study

Darla E Kendzor[1], PhD; Michael S Businelle[1], PhD; Joseph J C Waring[1], BPH; Ashley J Mathews[1], MS; Daryl W Geller[1], BS; Jocelyn M Barton[1], PhD; Adam C Alexander[1], PhD; Emily T Hébert[1], DrPH; Chaelin K Ra[1], PhD; Damon J Vidrine[2], DrPH

[1]Oklahoma Tobacco Research Center, Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States
[2]Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:
Darla E Kendzor, PhD
Oklahoma Tobacco Research Center
Stephenson Cancer Center
University of Oklahoma Health Sciences Center
655 Research Parkway, Suite 400
Oklahoma City, OK, 73104
United States
Phone: 1 4052718001 ext 50478
Email: Darla-Kendzor@ouhsc.edu

Abstract

Background: Socioeconomic disadvantage is associated with a reduced likelihood of smoking cessation. Smartphone ownership is increasing rapidly, including among low-income adults, and smartphone interventions for smoking cessation may increase access to smoking cessation treatment among socioeconomically disadvantaged adults.

Objective: This study aimed to evaluate the feasibility of an automated smartphone-based approach to delivering financial incentives for smoking cessation.

Methods: Socioeconomically disadvantaged adults initiating tobacco cessation treatment were followed from 1 week before a scheduled quit attempt through 26 weeks after the quit date. Participants received telephone counseling and nicotine replacement therapy. Smoking cessation was verified 5 times per week via smartphone prompts to self-report smoking status and submit a breath sample via a portable carbon monoxide (CO) monitor that was connected with participants' smartphones. Identity was verified during smoking status assessments using smartphone-based facial recognition software. When smoking abstinence and identity were verified, an automated credit card payment was triggered. Participants were incentivized for abstinence on the quit date and up to five days per week during the first 4 weeks after the scheduled quit date, with additional incentives offered during postquit weeks 8 and 12. In total, participants had the opportunity to earn up to US $250 in abstinence-contingent incentives over the first 12 weeks of the quit attempt.

Results: Participants (N=16) were predominantly female (12/16, 75%) and non-Hispanic white (11/16, 69%), black (4/16, 25%), or Hispanic of any race (1/16, 6%). Most participants (9/16, 56%) reported an annual household income of <US $11,000. During the first 4 weeks after the scheduled quit date, participants completed a median of 16 (out of 21; range 1-21) mobile smoking status assessments, and they earned a median of US $28 in abstinence-contingent incentives (out of a possible US $150; range US $0-US $135). Median earnings did not change during the 8- and 12-week incentivized follow-up periods (total median earnings over 12 weeks=US $28; range US $0-US $167). During the first 4 weeks after the scheduled quit date, participants abstained from smoking on a median of 5 (out of 21) assessment days (range 0-20). At the in-person follow-up visits, the expired CO-confirmed 7-day point prevalence abstinence rates were 19% (3/16) and 13% (2/16) at 12 and 26 weeks postquit, respectively. Overall, most participants reported that the system was easy to use and that they would recommend this treatment to their friends and family.

Conclusions: Preliminary data suggest that this smartphone-based approach to verifying identity and smoking status and automating the delivery of abstinence-contingent incentives to a credit card is feasible for use among socioeconomically disadvantaged adults. However, continued refinement is warranted.

(JMIR Mhealth Uhealth 2020;8(4):e15960) doi:10.2196/15960

KEYWORDS

socioeconomic status; smoking cessation; incentives; mobile health; mobile phone

Introduction

Background

Smoking prevalence rates are disproportionately high among socioeconomically disadvantaged adults [1], and socioeconomic disadvantage is associated with a reduced likelihood of smoking cessation [2-7]. Although individuals of lower socioeconomic status (SES) are just as likely to initiate quit attempts, they are less likely to succeed than those of higher SES [8]. Similarly, smoking cessation interventions for low SES populations have produced very low abstinence rates at follow-up [9-12]. Treatment approaches are needed to target socioeconomically disadvantaged populations and to reduce barriers to access. Smartphone interventions potentially offer a means of increasing access to treatment across settings. Notably, smartphone ownership is increasing rapidly, even among low-income adults. According to the Pew Research Center, 81% of US adults overall and 71% of US adults with an annual household income of <US $30,000 reported that they owned a smartphone in 2019 [13]. Importantly, adults with an annual household income of <US $30,000 are more than 4 times as likely to rely solely on their smartphones to access the internet than those earning >US $75,000 annually [13]. Smartphone-based treatments may be used in conjunction with traditional empirically supported approaches (counseling and pharmacotherapy), while also incorporating innovative components such as real-time assessment and intervention.

Contingency management (CM), the tangible reinforcement of abstinence and other related outcomes, is highly effective for promoting drug and alcohol abstinence among individuals with substance use disorders [14-16]. In addition, there is accumulating evidence that CM is an effective approach for promoting smoking cessation in a variety of populations [17-33], including adults of lower SES [34-39]. Research suggests that financial incentives for smoking cessation may be particularly appealing among socioeconomically disadvantaged individuals [40,41]. The findings of two meta-analyses have indicated that financial incentives are associated with greater odds of behavior change for a variety of behaviors, particularly among lower SES individuals [42,43].

To date, financial incentive interventions for smoking cessation have primarily relied on in-person visits to verify smoking abstinence. However, internet [44-48] and mobile phone-based [49-51] CM approaches have been developed to reduce or eliminate the need for in-person visits. An internet-based approach has been evaluated in several studies [44-48], where participants access a study website and record themselves via Webcam as they provide breath samples using a loaned carbon monoxide (CO) monitor. Participants upload the recordings for staff review, and a monetary credit is applied to their study credit cards when abstinence and identity are verified. A similar approach has been employed in several studies, where mobile phones equipped with video cameras allowed participants to record themselves as they provided a CO breath sample and, then, uploaded the videos for staff review via a study website [49-51]. Previous mobile CM approaches have lacked automation and have required substantial effort from participants and staff to record and upload videos, verify abstinence and identity, and administer payments. Technologies such as SCRAM continuous alcohol monitoring [52] and Soberlink [53] have combined facial recognition software with alcohol breath tests to automate the process of identity verification, and this approach could also be applied to CO breath tests among individuals who are attempting to quit smoking.

Objectives

The purpose of this project was to develop and test the feasibility of an automated smartphone-based CM approach that has the potential to allow socioeconomically disadvantaged adults to remotely benefit from smoking cessation treatments that offer financial incentives for evidence of smoking cessation. Feasibility was evaluated among socioeconomically disadvantaged males and females seeking smoking cessation treatment. This remote and automated intervention approach has the potential to increase the availability of financial incentive interventions for smoking cessation among individuals who are unable to attend in-person visits while reducing the need for staff monitoring and manual payment disbursement. This study represents a first step toward establishing the feasibility of a mobile CM intervention approach that is designed to be highly scalable and to ultimately facilitate the widespread adoption of CM treatments for smoking cessation.

Methods

Mobile Contingency Management

The INSIGHT mobile health (mHealth) platform is a versatile interface that empowers researchers to build, test, and launch smartphone-based assessments and interventions. With this platform, investigators have the ability to select how they would like their assessments or interventions deployed (eg, assessment type and interval, algorithms for delivering treatment content) [54]. Researchers may also enter their preferred assessment and intervention content (eg, surveys, treatment messages, videos). The following technologies were newly integrated into the INSIGHT platform for this study: (1) the Bedfont iCO Smokerlyzer CO monitor that connects with mobile phones [55] to remotely verify smoking abstinence, (2) Microsoft facial recognition software (Azure Face application programming

interface [API]) [56] to compare the participant's face at the time of the breath sample submission with a photo taken at baseline for identity verification, and (3) reloadable Greenphire ClinCards (ie, credit card) [57] for the remote delivery of abstinence-contingent incentives that are automatically triggered by biochemical confirmation of self-reported smoking abstinence (via smartphone-based ecological momentary assessment [EMA]). This newly developed intervention approach was tested for feasibility among socioeconomically disadvantaged adults.

Participants

Individuals were screened for eligibility following referral to the Tobacco Treatment Research Program (TTRP), which is located on the University of Oklahoma Health Sciences Center campus in Oklahoma City. The TTRP offers free tobacco cessation counseling and pharmacotherapy to the public, while also facilitating the recruitment, screening, and enrollment of participants into research studies [58]. TTRP referrals are primarily received through the electronic medical record, and they are also received via phone, internet, fax, and word-of-mouth. Following referral, participants are screened for ongoing studies at the TTRP over the phone and scheduled for an initial in-person treatment appointment. Most of the screening criteria for this study were assessed over the phone and verified in person, with the exception that CO level and literacy were only assessed in person. Interested participants were eligible for the study if they (1) were uninsured or received Medicaid benefits, (2) earned a score ≥4 on the Rapid Estimate of Adult Literacy in Medicine [59] indicating >6th grade English literacy level, (3) were willing to quit smoking 7 days from their first visit, (4) were 18 to 65 years of age, (5) had a CO level of ≥8 parts per million (ppm) suggestive of current smoking, and (6) were smoking ≥5 cigarettes per day. Participants were excluded from the study if they reported that they were (1) pregnant or breastfeeding, (2) had uncontrolled hypertension, (3) had a myocardial infarction within the past 2 weeks, (4) had an allergy to adhesive tape, or (5) were unwilling to use nicotine replacement therapy (NRT). Informed consent was obtained from eligible participants.

Participants referred to the TTRP who met eligibility criteria over the phone were scheduled for an in-person screening and enrollment visit (n=188). Of those, 91 attended their scheduled screening appointment, and 59 were eligible for this study. Of the 59 eligible patients, 36 were enrolled in another ongoing, higher priority study, 20 were enrolled in this study, and 3 declined to participate in either study. Of the 32 who were not eligible to participate, the reasons for exclusion were (in order of frequency): CO <8 ppm (n=10), reading level <6th grade (n=8), currently insured (besides Medicaid, n=6), smoking <5 cigarettes per day (n=3), uncontrolled hypertension (self-reported, n=3), unwilling to use NRT (n=2), or they were not ready to quit smoking in the next 7 days (n=1). Note that more than one reason for exclusion was possible. Those who were not eligible for the study were offered standard TTRP treatment. Participants (n=20) were enrolled in the study between August 2018 and February 2019, and final follow-up visits were completed in August 2019. Participants were intentionally enrolled slowly to ensure that the technology was working properly and to resolve technical problems as they arose before enrolling additional participants.

Measures

Smoking status was assessed on the scheduled quit day and on 5 days each week for 4 weeks following the scheduled quit date via self-report and CO levels with the portable Bedfont iCO Smokerlyzer (connected with a smartphone). Self-reported abstinence from smoking since 10 PM the previous evening (reported via smartphone-based EMAs) was verified with a CO level of <10 ppm on the quit day (given the recency of quitting). Consistent with current recommendations [60], a more stringent CO level of <6 ppm was required to corroborate smoking abstinence on all subsequent days. Key feasibility metrics included the number of smoking status assessments completed, number of facial recognition assessments initiated and working as expected, number of days of biochemically verified abstinence, abstinence-contingent incentives earned, counseling sessions completed, weeks of NRT requested/received, participant perceptions of the interventions, and attendance at follow-up visits. Participants were scheduled for in-person follow-up visits at 12 and 26 weeks post the quit date, where they were asked if they had smoked in the past 7 days. Participants' CO was measured at in-person visits with a Vitalograph BreathCO monitor, and a CO level of <6 ppm provided biochemical confirmation of abstinence. Note that participants' CO values and the CO thresholds used to verify abstinence were not shared with participants.

Procedure

Treatment

All participants were provided with the recommended components of an intensive tobacco treatment intervention [61]. Specifically, standard treatment included: (1) an in-person counseling session 1 week before the scheduled quit date with a certified tobacco treatment specialist (CTTS), (2) 5 weekly telephone counseling sessions with a CTTS starting on the scheduled quit date, (3) a 2-week supply of NRT (patches and gum or lozenges) during the first session, and (4) additional mailed patches and gum as needed for 12 weeks. The counseling sessions covered the following topics: (1) the health benefits of quitting, (2) mood/stress management strategies, (3) making positive lifestyle changes, (4) coping skills, and (5) relapse prevention. The CTTS checked in with participants each week about the difficulties and successes they experienced related to their quit attempt, and they planned for any challenging situations that were anticipated.

Paid Assessments

Participants earned US $30 to complete the in-person baseline visit that included Web-based questionnaires delivered via the research electronic data capture software platform (REDCap) [62,63], US $30 per week from the quit date through 4 weeks postquit for completing Web-based assessments via REDCap, US $40 for completing a Web-based REDCap assessment at 8 weeks postquit, US $40 for completing a 12-week postquit in-person visit (including Web-based questionnaires via REDCap and smoking status assessment), and US $20 for completing a 26-week in-person follow-up visit (including

Web-based REDCap questionnaires and smoking status assessment). Participants also earned up to US $150 depending on the percentage of prompted smartphone-based EMAs they had completed (including smartphone-based smoking status assessments). Note that questionnaire and EMA data not directly related to the primary purpose of the study are not presented or discussed further here. Overall, participants had the possibility of earning up to US $430 for the completion of questionnaires and EMAs (that were not contingent on abstinence), and all payments were credited to participants' ClinCards.

Smartphone-Based Smoking Status Assessment

Participants were provided with a Smokerlyzer iCO monitor, a Samsung Galaxy S7 smartphone preloaded with the INSIGHT app, and a remotely reloadable Greenphire ClinCard (credit card). Participants were offered unlimited calling and 2 gigabytes of data for personal use along with the phones. Participants were prompted via ring/vibration and a pop-up window (delivered through the INSIGHT app), indicating that it was time to complete a smoking status assessment (including a CO breath sample). Smoking status assessments were conducted at the end of the day (2 hours before typical bedtime) on the scheduled quit day and 5 days out of each week during the first 4 weeks after the quit day (days were selected to appear random). Participants received up to 2 additional reminder prompts (5 min apart) if the first and second assessment prompts were snoozed. Thus, they had a total of 10 min to complete the CO sample submission. If they missed the first assessment entirely (ie, did not snooze), they were prompted 30 min later, and if they missed the second assessment entirely, then 30 min later they were prompted a third time. After the third missed prompt, participants were no longer able to complete the smoking status assessment. In addition, participants were prompted to complete 5 smoking status assessments during weeks 8 and 12.

While participants provided a breath sample, 2 photos were taken at random times during the 20-second exhalation period using the front-facing smartphone camera. It was not apparent to participants exactly when the phone was taking pictures; rather, participants were only aware that during the process of providing the breath sample, they would be photographed for identity confirmation. The smartphone app compared the photos in vivo with a photo taken at each participant's baseline appointment to verify identity. Facial recognition assessments were considered to have worked as expected when identity was accurately processed as either a match or a nonmatch with the baseline photo. Study staff reviewed all of the stored photos, and nonmatches in which participant identity was confirmed by research staff were considered to be failures. There were no instances of erroneous matches.

Abstinence-Contingent Incentives

Abstinence-contingent incentives were automatically delivered to the credit card based on daily self-reports of smoking abstinence and CO levels. This was accomplished through an API that allowed the INSIGHT app to communicate with Greenphire. A spreadsheet was automatically generated and sent to Greenphire, which contained the payment amounts owed to participants' ClinCard accounts. These payments were automatically drawn and disbursed from an account with Greenphire that was set up in advance of the study. On the scheduled quit day, participants who were biochemically confirmed abstinent from smoking since 10 PM the previous evening received a US $20 credit on their ClinCard (incentive schedule is detailed in Table 1). After the quit day, a payment was earned following a smartphone-based self-report of abstinence during the past 24 hours combined with a breath CO sample of <6 ppm and a facial recognition match. For each abstinent day during the first week postquit, a US $4 credit was earned. The incentive amount per abstinent day increased by US $1 with each week of consecutive abstinence until 4 weeks postquit, when continuously abstinent participants earned US $7 per abstinent day. Participants who provided 5 negative breath samples within a week additionally received a US $5 bonus through 4 weeks postquit. Participants who were nonabstinent (or who did not provide a sample) did not earn incentives that day but could begin earning incentives for abstinence again on their next abstinent day, although the amount was reset to the starting level of US $4 per abstinent day. Participants earned US $8 per abstinent day during weeks 8 and 12 postquit, with a US $10 bonus for 5 negative samples each week. Payments did not reset or escalate during the 8- and 12-week follow-up periods. Altogether, participants could earn up to a possible US $250 for biochemically verified abstinence. The incentive schedule (Table 1) was adapted from the schedules utilized with socioeconomically disadvantaged adults in previous research by the investigators [34,36].

Table 1. Incentive schedule.

Weeks postquitting	Abstinence-contingent incentives[a] (US $)	Total[b] (US $)
Quit day	20 for negative CO[c] sample	20
Week 1	4 per negative CO sample (up to 20 + 5 bonus)	25
Week 2	5 per negative CO sample (up to 25 + 5 bonus)	30
Week 3	6 per negative CO sample (up to 30 + 5 bonus)	35
Week 4	7 per negative CO sample (up to 35 + 5 bonus)	40
Week 8	8 per negative CO sample (up to 40 + 10 bonus)	50
Week 12	8 per negative CO sample (up to 40 + 10 bonus)	50

[a]Participants earned a bonus incentive of US $5 when they achieved biochemically verified abstinence on all 5 smoking status assessments during the first 4 weeks after the scheduled quit date. The bonus payment increased to US $10 during postquit weeks 8 and 12.

[b]Partcipants could earn up to US $250 in abstinence-contingent incentives over the 12-week intervention period.

[c]CO: carbon monoxide.

Problem Reporting and Resolution

Participants were able to report *bugs* or contact staff with questions about using the smartphone by clicking *call study staff* on the main menu of the app. In addition, participants were asked if they had experienced any problems with the app as part of the end-of-day assessments during the first 4 weeks postquit. If they said yes, they were prompted to offer detailed information in an open text box. Participants also had weekly scheduled counseling calls where they could ask questions and report problems. Over the course of the study, instructions to participants about how to successfully use the facial recognition component of the app were improved. We began offering more detailed information (training and paper handout) about optimal facial placement on the phone screen when providing a CO breath sample, and we emphasized the importance of maintaining consistency in appearance related to eyeglasses, hairstyles, background, and facial position. Proper facial placement was important because the app did not provide feedback to the participants during facial recognition assessments (this will be addressed in future versions of the app). In addition, the mHealth programming staff worked through several technical problems that arose during the study, which interfered with the proper functioning of the facial recognition component of the app. Owing to these early improvements, outcomes were characterized separately for the first and second halves of participants enrolled to demonstrate how these improvements might have impacted study outcomes (see the Analytic Plan section).

Analytic Plan

Descriptive statistics were generated for key feasibility metrics (noted above in the measures section), including medians and ranges for continuous variables and frequencies for categorical variables. The sample was also divided into the first and second halves enrolled, and sample descriptives were generated for each half to illustrate the impact of improvements to the protocol and smartphone app. Nonparametric Mann-Whitney *U* tests and chi-square analyses were conducted to compare the first and second halves of participants enrolled on key metrics. Statistical comparisons may provide information about trends toward improvement, despite the small sample size.

Results

Participant Characteristics

Of those enrolled, 4 participants did not complete any smoking status assessments on their smartphones. Because these participants did not initiate the financial incentives component of the intervention, they were excluded from the primary analyses. The remaining 16 enrolled participants were predominantly female (12/16, 75%), with a median age of 47 years (range 18-63 years). Participants were 69% (11/16) non-Hispanic white, 25% (4/16) non-Hispanic black, and 6% (1/16) Latino/Hispanic. Participants reported a median of 12.5 years of education (range 9-14 years), with 25% (4/16) completing less than 12 years of education. Before quitting, participants reported smoking a median of 19.0 (range 5-40) cigarettes per day for 30.5 years (range 6-49). Most participants were not employed (12/16, 75%), and most reported an annual household income of <US $11,000 (9/16, 56%). Of the 16 individuals who participated in the intervention, only 1 phone was not returned (1/16, 6%). However, 3 of the 4 individuals who did not participate in the smartphone-based financial incentives component of the intervention did not return their phones. Thus, 4 phones were not returned among the 20 enrolled participants overall (4/20, 20% phone loss).

Although statistical comparisons between enrolled participants who initiated (n=16) or did not initiate (n=4) the intervention were not feasible given the small sample size, it is worth noting that those who did not initiate the intervention appeared to differ from those who initiated the intervention on a variety of characteristics. Those who did not initiate the intervention were less likely to be female (2/4, 50% vs 12/16, 75%), and were more likely to report being black (2/4, 50% vs 4/16, 25%) or American Indian (1/4, 25% vs 0/16, 0%). In addition, those who did not initiate the intervention were younger (median 41.5; range 36-45 years vs 47.0 years, range 18-63 years), smoked fewer cigarettes per day (median 12.5; range 5-20 vs 19.0; range 5-40), smoked for fewer years (median 20.0; range 13-30 years vs 30.5; range 6-49 years), had less education (median 10.5; range 9-12 years vs 12.5; range 9-14 years), and reported less

income (4/4, 100% vs 9/16, 56% with <US $11,000 in annual household income).

Completion of Mobile Smoking Status Assessments

Participants completed a median of 16 (range 1-21) of the 21 possible iCO/smoking status assessments during the first 4 weeks postquit (ie, 1 on the quit date, and 5 per week thereafter). Notably, the final 8 participants enrolled had higher median completion rates relative to the first 8 enrolled (19.5 completed assessments, range 14-21 vs 10 completed assessments, range 1-20; *P*=.02). Unfortunately, smoking status assessment completion rates declined during week 8 (median 3.0 completed assessments out of a possible 5) and declined even further during week 12 (median 1.5 assessments completed out of a possible 5; Table 2).

Table 2. Improvements in treatment-related variables for the first to second half of the participants enrolled (N=16).

Postquit date and treatment-related variables	All participants	First half enrolled	Last half enrolled	*P* value[a,b]
Weeks 1-4, median (range)				
Days carbon monoxide-confirmed abstinent[c] (out of 21)	5.0 (0.0-20.0)	3.0 (0.0-11.0)	8.5 (1.0-20.0)	.07
Completed smoking status assessments[c] (out of 21)	16.0 (1.0-21.0)	10.0 (1.0-20.0)	19.5 (14.0-21.0)	.02
Telephone counseling sessions completed (out of 5)	5.0 (1.0-5.0)	4.5 (1.0-5.0)	5.0 (4.0-5.0)	.20
Abstinence-contingent incentives earned (US $; up to US $150)	28.0 (0.0-135.0)	20.0 (0.0-72.0)	60.0 (16.0-167.0)	.05
Week 8, median (range)				
Days carbon monoxide-confirmed abstinent[d] (out of 5)	0.0 (0.0-4.0)	0.0 (0.0-1.0)	1.0 (0.0-4.0)	.13
Completed smoking status assessments[d] (out of 5)	3.0 (0.0-5.0)	2.0 (0.0-5.0)	4.0 (0.0-5.0)	.16
Abstinence-contingent incentives earned (US $; up to US $50)	0.0 (0.0-24.0)	0.0 (0.0-8.0)	4.0 (0.0-24.0)	.28
Week 12				
Days carbon monoxide-confirmed abstinent[d] (out of 5), median (range)	0.0 (0.0-3.0)	0.0 (0.0-1.0)	0.0 (0.0-3.0)	.65
Completed smoking status assessments[d] (out of 5), median (range)	1.50 (0.0-4.0)	1.50 (0.0-3.0)	1.50 (0.0-4.0)	.57
Weeks of NRT[e] (out of 12), median (range)	10.0 (2.0-12.0)	10.0 (2.0-12.0)	10.0 (2.0-12.0)	.86
Abstinence-contingent incentives earned (US $; up to US $50), median (range)	0.0 (0.0-16.0)	0.0 (0.0-16.0)	0.0 (0.0-16.0)	.51
Attended in-person follow-up visit, n (%)	13.0 (81.0)	6.0 (75.0)	7.0 (88.0)	.52
Carbon monoxide-confirmed 7-day point prevalence abstinence, n (%)	3.0 (19.0)	1.0 (13.0)	2.0 (25.0)	.52
Week 26				
Attended in-person follow-up visit, n (%)	11.0 (69.0)	5.0 (63.0)	6.0 (75.0)	.59
Carbon monoxide-confirmed 7-day point prevalence abstinence, n (%)	2.0 (13.0)	0.0 (0.0)	2.0 (25.0)	.13

[a]Mann-Whitney *U* tests were conducted for continuous variables, and chi-square analyses were conducted for dichotomous variables.

[b]*P* values reflect the difference between the first 8 participants and the last 8 participants enrolled in the study.

[c]Self-reported and carbon monoxide–confirmed abstinence were assessed on the day after quitting and on 5 days per week for 4 weeks after the scheduled quit attempt (21 total assessments).

[d]Smoking status was additionally assessed 5 times per week during weeks 8 and 12 postquit date. Assessments where participants self-reported smoking but did not complete the iCO assessment were not considered missing.

[e]NRT: nicotine replacement therapy.

Facial Recognition

Over the 12 week study period, there were 31 possible facial recognition assessments per participant (ie, 496 total possible for all 16 participants). Overall, the 16 study participants collectively initiated 282 facial recognition assessments as part of the smoking status assessments, and 48.6% (137/282) of those assessments worked as expected. Among the first 8 participants, 127 facial recognition assessments were initiated, and 30.7% (39/127) of those assessments *worked as expected*. Among the last 8 participants, 155 facial recognition assessments were initiated, of which 98 (63.2%) worked as expected. No instances were noted where someone other than the participant was photographed during a smoking status assessment. Although we do not have detailed documentation of the reasons for every instance of facial recognition failure, the failures seemed to fall into 3 broad categories: (1) improper facial placement in the frame during smoking status assessments (eg, face partially in the photo), (2) inconsistent appearance between the baseline photo and the photos taken during smoking status assessments (eg, inconsistencies in wear of hairstyles, glasses, hats), and (3) technical problems or *bugs*. The former categories were addressed by providing participants with additional guidance about proper facial placement and the importance of consistency

in appearance across assessments. The latter technical problems were addressed by the programming staff as they arose.

Treatment Engagement

All participants completed the in-person, prequit counseling session. After the initial visit, participants completed a median of 5 out of 5 possible weekly telephone counseling sessions (Table 2). A total of 69% (11/16) of participants completed all 5 counseling calls, whereas 19% (3/16) completed 4 counseling calls, 6% (1/16) completed 2 calls, and 6% (1/16) completed 1 call. A 2-week supply of NRT was offered at the in-person visit, and participants could request an additional 2-seek supply every 2 weeks for up to 12 weeks. Additional requested NRT was mailed to the participants. A 2-week supply included 14 patches and a box of 2 mg or 4 mg of nicotine gum or lozenges, depending on participant preference and smoking level (each box contained 100 pieces of gum or 108 lozenges). Participants requested and received a median of 10 weeks of NRT (range 2-12 weeks), with 13% (2/16) receiving a 2 week supply, 13% (2/16) receiving 4 weeks, 19% (3/16) receiving 3 weeks, 25% (4/16) receiving 10 weeks, and 31% (5/16) receiving a 12-week supply.

Perceptions of the Intervention

Of the 16 study participants, 14 (88%) completed the Web-based intervention perceptions questionnaire via REDCap 4 weeks after the scheduled quit date. Note that the 2 participants with missing data were among the first half of the participants enrolled in the study. Overall, most participants found the intervention to be easy to use and helpful (for details, see Table 3). In addition, participants reported that they had problems with the app: “never” (4/16, 25%), “once or twice” (4/16, 25%), “a few times” (4/16, 25%), “several times” (1/16, 6%), and “daily or every other day” (1/16, 6%). Participants reported that the smartphone app and smoking monitor were correct in determining whether or not they were smoking: “never” (2/16, 13%), “some of the time” (2/16, 13%), “about half of the time” (1/16, 6%), “most of the time” (4/16, 25%) and “always” (5/16, 31%). Participants reported that it was difficult to find the iCO monitor when they needed to complete a smoking assessment: “never” (8/16, 50%), “a few times” (2/16, 13%), and “several times” (4/16, 25%). Finally, in response to the statement “earning financial incentives for quitting helped me to quit again after I had smoked” participants chose “I never smoked/lapsed after I quit” (3/16, 19%), “disagree” (1/16, 6%), “neither agree nor disagree” (5/16, 31%), “agree” (3/16, 19%), and “strongly agree” (2/16, 1%).

Table 3. Participants’ perceptions of the automated mobile contingency management intervention 4 weeks after a scheduled quit attempt (N=16).

Perceptions[a]	Agree or strongly agree, n (%)	Neither agree or disagree, n (%)	Disagree or strongly disagree, n (%)
The smartphone app was easy to use overall	12 (75)	2 (13)	0 (0)
My overall opinion of the smartphone app was positive	11 (69)	2 (13)	1 (6)
The smoking monitor was easy to use	11 (69)	1 (6)	2 (13)
It was difficult to blow into the smoking monitor while keeping my face in front of the smartphone screen	4 (25)	3 (19)	7 (44)
It was easy to tell how much I had earned for quitting each day/week by checking the payment screen	12 (75)	1 (6)	1 (6)
The opportunity to earn financial incentives for quitting helped keep my motivation for quitting high	11 (69)	3 (19)	0 (0)
Earning financial incentives for quitting helped me to feel more confident in my ability to quit	11 (69)	3 (19)	0 (0)
Earning financial incentives for quitting smoking helped me to successfully quit smoking	5 (31)	7 (44)	2 (13)

[a]A total of 88% (14/16) participants completed the perception survey 4 weeks after the scheduled quit date. As a result, the frequencies across the rows do not add up to 100%. The omitted 13% reflects the missing responses of 2 participants who were among the first half of the participants enrolled.

Smoking Abstinence and Incentives Earned

See Table 2 for details about smoking abstinence and incentive earnings across the study visits. Participants were biochemically confirmed abstinent on a median of 5 days (range 0-20 days) out of 21 assessment days during the first 4 weeks postquit, with the last 8 participants achieving more abstinent days relative to the first 8 enrolled (median 8.5 abstinent days, range 1-20 vs 3 abstinent days, range 0-11; P=.07). Owing to low completion rates, evaluations of smoking status via smartphone assessments during weeks 8 and 12 were problematic. Over the entire 12-week incentive period, participants earned a median of US $28 (range US $0-US $167, out of US $250 possible) in abstinence-contingent incentives. The average earnings were greater among the last 8 participants enrolled than the first 8 participants (median US $60, range US $16-US $167 vs $20, range US $0-US $72; P=.05).

Participants were asked to attend 2 in-person follow-up visits at 12 and 26 weeks after the scheduled quit date to assess smoking status. Self-reported, biochemically confirmed (CO breath sample <6 ppm) 7-day point prevalence abstinence rates were 19% (3/16) and 13% (2/16), respectively, at the 12- and 26-week in-person follow-up visits. Attendance rates at the in-person follow-up visits were 81% (13/16) and 69% (11/16), respectively, at 12 and 26 weeks postquitting date, with those

who did not attend considered to be smoking. As with the smartphone metrics, abstinence and attendance rates at in-person visits were higher among the latter half of participants enrolled relative to the first half of participants, suggesting that the delivery of the intervention may have improved over time (Table 2).

Discussion

Principal Findings

Overall, preliminary data suggest that this remote approach to verifying smoking status and participant identity and automating the delivery of abstinence-contingent incentives to a credit card is feasible for use with socioeconomically disadvantaged adults seeking smoking cessation treatment. During the study, improvements made to the protocol and smartphone app corresponded with improvements on most study metrics between the first half and the latter half of participants (Table 2) including (1) smoking status assessments completed, (2) facial recognition assessments *working as expected*, (3) telephone counseling sessions completed, (4) biochemically verified abstinent days, and (5) incentives earned for abstinence. Notably, most participants reported that they would recommend this intervention to their friends and family. At the in-person follow-up visits, CO-confirmed 7-day point prevalence abstinence rates were 19% (3/16) and 13% (2/16) at 12 and 26 weeks postquit date, respectively, which was high relative to other studies with socioeconomically disadvantaged individuals [9-12]. Attendance at follow-up visits was good, with in-person follow-up rates of 81% (13/16) and 69% (11/16) at 12 and 26 weeks postquit date, respectively.

Opportunities for Improvement

Despite positive indicators of overall feasibility, there is still much opportunity for improvement. In total, 4 out of 20 participants enrolled (4/20, 20%) did not initiate engagement with the app, and therefore did not earn abstinence-contingent financial incentives. Approaches will be needed in future research to monitor lack of engagement and automate assistance early in the intervention as needed. For example, participant app use could be monitored, possibly through automated notifications to staff, to prompt staff outreach with the goal of increasing initial engagement for those who need it. It is possible that participants who did not engage with the app were less comfortable utilizing the technology, had experienced a problem that they were unable to resolve, or simply needed a reminder or encouragement to initiate the intervention. Although all participants were provided with phones for this study, 2 of the 4 participants who did not initiate the app component of the intervention reported that they did not own a mobile phone at the time they enrolled in the study. In contrast, among the 16 participants who engaged with the app, only 1 participant reported that he/she did not own a mobile phone.

Notably, although the mobile smoking status assessment completion rate was very high during the first 4 weeks postquit among the latter half of participants (ie, median of 19.5 out of 21 completed assessments), assessment completion rates declined substantially over time. The completion of these assessments is crucial to this type of intervention; therefore, a continued focus on improving compliance is warranted. It is possible that, on occasion, participants were unprepared for smoking status assessments because they were prompted by the app at inconvenient times. An alternative approach used by Dallery et al [44,48] requires that participants self-initiate CO breath samples twice per day at least eight hours apart. This strategy could maximize convenience for participants because they choose their assessment times, thus reducing missed assessments, while also limiting participants' ability to smoke without detection. The percentage of completed smoking status assessments decreased dramatically between the first 4 weeks and the 8- and 12-week assessments, suggesting that participants became disengaged with the app during the *break* periods. These follow-up assessment/incentive weeks could be eliminated and/or the continuous incentive period could be extended to 6 weeks to increase continuous engagement with the app and maximize the influence of the intervention. Alternatively, smartphone assessments and/or messaging could be added between the incentive weeks to encourage sustained participant engagement.

Challenges Associated With Facial Recognition

Incorporating facial recognition software posed challenges. There were several technical *bugs* that interfered with proper facial recognition processing. This had the effect of preventing incentive payments for participants and subsequently required attention from study staff to manually pay participants. During internal testing, we found that photos could be used to pass the identity verification component of the smoking status assessments. However, this deception becomes obvious when the stored photos are reviewed, and to our knowledge, this did not occur in this feasibility study. A random review of stored photos will be incorporated in future studies to identify attempted deceptions and other problems. Future iterations of the smartphone intervention will benefit from the resolution of problems in this study, and presumably, facial recognition metrics will continue to improve in future studies.

It is noteworthy that a recent evaluation of facial recognition software uncovered bias in facial recognition systems; for example, greater false match rates have been identified among women than men, and among African Americans than Caucasians [64]. Furthermore, false positives are more likely to occur when comparing images of individuals of the same sex, age, and ethnicity. Current recommendations to mitigate these problems include the use of training data that are diverse and globally derived, assessment of both the face and the iris, and modification of matching thresholds based on demographics [64]. As facial recognition software developers begin following these recommendations, the accuracy of the software may be expected to improve.

Limitations

This study has limitations. First, the sample size was small, and the study did not include a comparison group, thus limiting the ability to draw conclusions about intervention effectiveness. Nevertheless, the sample size and design are appropriate for the initial feasibility testing, which will inform improvements to future versions of the intervention. Regarding biochemical verification of abstinence, it is possible that participants could

have smoked between assessments given the short half-life of CO. However, submitting negative CO samples on a near-daily basis while continuing to smoke would have required participant knowledge of CO half-life in combination with discipline in the timing of their daily smoking. For these reasons, it seems unlikely that smoking would go undetected for an extended period of time. Notably, the current intervention includes more frequent smoking status assessments than many other intervention studies utilizing incentives, which have included biochemical verification at weekly intervals or at key follow-up points [34,37,38]. Cotinine assessment provides a longer window of smoking detection, and could be included in future studies of mobile CM when in-person follow-up assessments are part of the protocol (ie, identity could not be easily verified if cotinine was assessed remotely).

Another potential concern about remote biochemical verification of abstinence relates to the limited ability to verify if a participant is actually exhaling through the iCO monitor. The iCO monitor must be plugged-in during the smoking status assessment; otherwise, the app records an error code, and the assessment cannot be completed. In addition, second-to-second variability in CO readings during the exhalation period (20 seconds) can be assessed and is currently being explored as a means of verifying exhalation. Specifically, the variability in CO is expected to be greater during the exhalation period than during the period before exhalation (ie, when participants are instructed to hold their breath) even among those who are not smoking.

Notably, participants were provided with smartphones in this feasibility study to increase consistency and control over app functionality. In future testing, participants' own phones will be used to deliver the intervention to verify usability across multiple types of phones and smartphone service plans. Similarly, in-person visits will be eliminated in future iterations of this research, in favor of completely remote smartphone setup, assessment, and treatment to demonstrate the scalability of the intervention.

Costs and Potential for Real-World Utility

If this intervention is ultimately shown to be efficacious in a fully powered trial, this automated CM approach has numerous potential applications in settings where smokers may have limited access to smoking cessation resources or transportation. Plausibly, mobile CM could be paired with the services offered through state tobacco cessation quitlines and health care clinics to increase the reach of incentives-based interventions to socioeconomically disadvantaged individuals. For example, Mundt et al [39] demonstrated that incentivizing both quitline counseling calls and biochemically verified smoking cessation among Medicaid recipients improved cessation rates relative to a nonincentivized control group and the costs of the intervention compared favorably with other treatments. Nevertheless, real-world factors may impact the feasibility of implementation in different settings, including the availability of funding for the intervention and operational factors unique to each setting.

Notably, the incentive schedule utilized in this study and the investigators' previous work [34,36] is low-cost and potentially cost-effective and compares favorably with the cost of other empirically supported and commonly used tobacco cessation treatments (eg, 10 weeks of generic nicotine patch costs approximately US $130). Previously, we found that cessation rates among socioeconomically disadvantaged adults incentivized for smoking cessation were more than double those of the standard treatment only control group for an additional cost of only US $63 (on average) per participant [34]. Cost-effectiveness research is needed to evaluate the practicality of this approach, and it is necessary to evaluate CM interventions more broadly. The optimal magnitude of abstinence-contingent incentives and the length of time that incentives should be offered must be explored.

Future research will focus on continued refinement and evaluation of this automated mobile CM approach to smoking cessation among socioeconomically disadvantaged individuals. Effective interventions with the potential for broad reach are especially important in places such as Oklahoma where nonmetropolitan residence is common, and rates of poverty and lack of health insurance are elevated. Once refined and tested in a fully powered trial, this fully automated CM approach to smoking cessation has the potential to facilitate intervention delivery to socioeconomically disadvantaged individuals across settings and locations.

Acknowledgments

This research was primarily supported by an administrative supplement to Grant R01CA197314 (to DK) awarded by the National Cancer Institute (NCI). Additional support was provided by Oklahoma Tobacco Settlement Endowment Trust Grant 092-016-0002, and the Mobile Health Technology Shared Resource, which is a component of the NCI Cancer Center Support Grant P30CA225520 awarded to the Stephenson Cancer Center. This study was registered at ClinicalTrials.gov (NCT03517397).

Conflicts of Interest

DK, MB, AM, DG, and DV earn royalties for the use of the INSIGHT mobile health platform by institutions external to the University of Oklahoma. However, note that royalties were not earned for the research described in this published manuscript because the Principal Investigator (DK) is appointed at the University of Oklahoma Health Sciences Center.

References

1. Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, et al. Tobacco product use among adults-United States, 2017. MMWR Morb Mortal Wkly Rep 2018 Dec 9;67(44):1225-1232 [FREE Full text] [doi: 10.15585/mmwr.mm6744a2] [Medline: 30408019]
2. Siahpush M, Carlin JB. Financial stress, smoking cessation and relapse: results from a prospective study of an Australian national sample. Addiction 2006 Jan;101(1):121-127. [doi: 10.1111/j.1360-0443.2005.01292.x] [Medline: 16393198]
3. Fernández E, Schiaffino A, Borrell C, Benach J, Ariza C, Ramon JM, et al. Social class, education, and smoking cessation: long-term follow-up of patients treated at a smoking cessation unit. Nicotine Tob Res 2006 Mar;8(1):29-36. [doi: 10.1080/14622200500264432] [Medline: 16497597]
4. Wetter DW, Cofta-Gunn L, Irvin JE, Fouladi RT, Wright K, Daza P, et al. What accounts for the association of education and smoking cessation? Prev Med 2005 May;40(4):452-460. [doi: 10.1016/j.ypmed.2004.07.001] [Medline: 15530598]
5. Businelle MS, Kendzor DE, Reitzel LR, Costello TJ, Cofta-Woerpel L, Li Y, et al. Mechanisms linking socioeconomic status to smoking cessation: a structural equation modeling approach. Health Psychol 2010 May;29(3):262-273 [FREE Full text] [doi: 10.1037/a0019285] [Medline: 20496980]
6. Fagan P, Shavers VL, Lawrence D, Gibson JT, O'Connell ME. Employment characteristics and socioeconomic factors associated with disparities in smoking abstinence and former smoking among US workers. J Health Care Poor Underserved 2007 Dec;18(4 Suppl):52-72. [doi: 10.1353/hpu.2007.0119] [Medline: 18065852]
7. Kendzor DE, Businelle MS, Costello TJ, Castro Y, Reitzel LR, Cofta-Woerpel LM, et al. Financial strain and smoking cessation among racially/ethnically diverse smokers. Am J Public Health 2010 May;100(4):702-706 [FREE Full text] [doi: 10.2105/AJPH.2009.172676] [Medline: 20167886]
8. Kotz D, West R. Explaining the social gradient in smoking cessation: it's not in the trying, but in the succeeding. Tob Control 2009 Mar;18(1):43-46 [FREE Full text] [doi: 10.1136/tc.2008.025981] [Medline: 18936053]
9. Froelicher ES, Doolan D, Yerger VB, McGruder CO, Malone RE. Combining community participatory research with a randomized clinical trial: the protecting the hood against tobacco (PHAT) smoking cessation study. Heart Lung 2010;39(1):50-63. [doi: 10.1016/j.hrtlng.2009.06.004] [Medline: 20109986]
10. Okuyemi KS, James AS, Mayo MS, Nollen N, Catley D, Choi WS, et al. Pathways to health: a cluster randomized trial of nicotine gum and motivational interviewing for smoking cessation in low-income housing. Health Educ Behav 2007 Mar;34(1):43-54. [doi: 10.1177/1090198106288046] [Medline: 16778147]
11. Hahn EJ, Rayens MK, Chirila C, Riker CA, Paul TP, Warnick TA. Effectiveness of a quit and win contest with a low-income population. Prev Med 2004 Oct;39(3):543-550. [doi: 10.1016/j.ypmed.2004.02.012] [Medline: 15313094]
12. Kendzor DE, Reitzel LR, Mazas CA, Cofta-Woerpel LM, Cao Y, Ji L, et al. Individual- and area-level unemployment influence smoking cessation among African Americans participating in a randomized clinical trial. Soc Sci Med 2012 May;74(9):1394-1401 [FREE Full text] [doi: 10.1016/j.socscimed.2012.01.013] [Medline: 22405506]
13. Anderson M. Pew Research Center. 2019. Mobile Technology and Home Broadband 2019 URL: https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2019-08-16]
14. Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. Am J Psychiatry 2008 Mar;165(2):179-187. [doi: 10.1176/appi.ajp.2007.06111851] [Medline: 18198270]
15. Stitzer M, Petry N. Contingency management for treatment of substance abuse. Annu Rev Clin Psychol 2006;2:411-434. [doi: 10.1146/annurev.clinpsy.2.022305.095219] [Medline: 17716077]
16. Prendergast M, Podus D, Finney J, Greenwell L, Roll J. Contingency management for treatment of substance use disorders: a meta-analysis. Addiction 2006 Dec;101(11):1546-1560. [doi: 10.1111/j.1360-0443.2006.01581.x] [Medline: 17034434]
17. Dunn KE, Sigmon SC, Reimann EF, Badger GJ, Heil SH, Higgins ST. A contingency-management intervention to promote initial smoking cessation among opioid-maintained patients. Exp Clin Psychopharmacol 2010 Mar;18(1):37-50 [FREE Full text] [doi: 10.1037/a0018649] [Medline: 20158293]
18. Tevyaw T, Colby S, Tidey J, Kahler C, Rohsenow D, Barnett N, et al. Contingency management and motivational enhancement: a randomized clinical trial for college student smokers. Nicotine Tob Res 2009 Jul;11(6):739-749 [FREE Full text] [doi: 10.1093/ntr/ntp058] [Medline: 19443788]
19. Heil S, Higgins S, Bernstein I, Solomon L, Rogers R, Thomas C, et al. Effects of voucher-based incentives on abstinence from cigarette smoking and fetal growth among pregnant women. Addiction 2008 Jul;103(6):1009-1018 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02237.x] [Medline: 18482424]
20. Volpp KG, Levy AG, Asch DA, Berlin JA, Murphy JJ, Gomez A, et al. A randomized controlled trial of financial incentives for smoking cessation. Cancer Epidemiol Biomarkers Prev 2006 Jan;15(1):12-18 [FREE Full text] [doi: 10.1158/1055-9965.EPI-05-0314] [Medline: 16434580]
21. Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA, et al. A randomized, controlled trial of financial incentives for smoking cessation. N Engl J Med 2009 Mar 12;360(7):699-709. [doi: 10.1056/NEJMsa0806819] [Medline: 19213683]
22. Dunn K, Sigmon S, Thomas C, Heil S, Higgins S. Voucher-based contingent reinforcement of smoking abstinence among methadone-maintained patients: a pilot study. J Appl Behav Anal 2008;41(4):527-538 [FREE Full text] [doi: 10.1901/jaba.2008.41-527] [Medline: 19192857]

23. Alessi S, Petry N, Urso J. Contingency management promotes smoking reductions in residential substance abuse patients. J Appl Behav Anal 2008;41(4):617-622 [FREE Full text] [doi: 10.1901/jaba.2008.41-617] [Medline: 19192865]
24. Hunt YM, Rash CJ, Burke RS, Parker JD. Smoking cessation in recovery: comparing 2 different cognitive behavioral treatments. Addict Disord Treat 2010;9(2):64-74. [doi: 10.1097/adt.0b013e3181bf0310]
25. Correia CJ, Benson TA. The use of contingency management to reduce cigarette smoking among college students. Exp Clin Psychopharmacol 2006 May;14(2):171-179. [doi: 10.1037/1064-1297.14.2.171] [Medline: 16756421]
26. Higgins S, Bernstein I, Washio Y, Heil S, Badger G, Skelly J, et al. Effects of smoking cessation with voucher-based contingency management on birth outcomes. Addiction 2010 Dec;105(11):2023-2030 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03073.x] [Medline: 20840188]
27. Higgins ST, Washio Y, Heil SH, Solomon LJ, Gaalema DE, Higgins TM, et al. Financial incentives for smoking cessation among pregnant and newly postpartum women. Prev Med 2012 Dec;55(Suppl):S33-S40 [FREE Full text] [doi: 10.1016/j.ypmed.2011.12.016] [Medline: 22227223]
28. Ledgerwood DM, Arfken CL, Petry NM, Alessi SM. Prize contingency management for smoking cessation: a randomized trial. Drug Alcohol Depend 2014 Jul 1;140:208-212 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.03.032] [Medline: 24793364]
29. Secades-Villa R, García-Rodríguez O, López-Núñez C, Alonso-Pérez F, Fernández-Hermida JR. Contingency management for smoking cessation among treatment-seeking patients in a community setting. Drug Alcohol Depend 2014 Jul 1;140:63-68. [doi: 10.1016/j.drugalcdep.2014.03.030] [Medline: 24768410]
30. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
31. Donatelle RJ, Prows SL, Champeau D, Hudson D. Randomised controlled trial using social support and financial incentives for high risk pregnant smokers: significant other supporter (SOS) program. Tob Control 2000;9(Suppl 3):III67-III69 [FREE Full text] [doi: 10.1136/tc.9.suppl_3.iii67] [Medline: 10982912]
32. Etter J, Schmid F. Effects of large financial incentives for long-term smoking cessation: a randomized trial. J Am Coll Cardiol 2016 Aug 23;68(8):777-785 [FREE Full text] [doi: 10.1016/j.jacc.2016.04.066] [Medline: 27539168]
33. Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. Cochrane Database Syst Rev 2019 Jul 17;7:CD004307. [doi: 10.1002/14651858.CD004307.pub6] [Medline: 31313293]
34. Kendzor DE, Businelle MS, Poonawalla IB, Cuate EL, Kesh A, Rios DM, et al. Financial incentives for abstinence among socioeconomically disadvantaged individuals in smoking cessation treatment. Am J Public Health 2015 Jul;105(6):1198-1205. [doi: 10.2105/AJPH.2014.302102] [Medline: 25393172]
35. Rash CJ, Petry NM, Alessi SM. A randomized trial of contingency management for smoking cessation in the homeless. Psychol Addict Behav 2018 Mar;32(2):141-148 [FREE Full text] [doi: 10.1037/adb0000350] [Medline: 29461070]
36. Businelle MS, Kendzor DE, Kesh A, Cuate EL, Poonawalla IB, Reitzel LR, et al. Small financial incentives increase smoking cessation in homeless smokers: a pilot study. Addict Behav 2014 Mar;39(3):717-720. [doi: 10.1016/j.addbeh.2013.11.017] [Medline: 24321696]
37. Lasser KE, Quintiliani LM, Truong V, Xuan Z, Murillo J, Jean C, et al. Effect of patient navigation and financial incentives on smoking cessation among primary care patients at an urban safety-net hospital: a randomized clinical trial. JAMA Intern Med 2017 Dec 1;177(12):1798-1807 [FREE Full text] [doi: 10.1001/jamainternmed.2017.4372] [Medline: 29084312]
38. Fraser D, Fiore M, Kobinsky K, Adsit R, Smith S, Johnson M, et al. A randomized trial of incentives for smoking treatment in medicaid members. Am J Prev Med 2017 Dec;53(6):754-763 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.027] [Medline: 29079405]
39. Mundt M, Baker T, Fraser D, Smith S, Piper M, Fiore M. Paying low-income smokers to quit? The cost-effectiveness of incentivizing tobacco quit line engagement for medicaid recipients who smoke. Value Health 2019 Feb;22(2):177-184 [FREE Full text] [doi: 10.1016/j.jval.2018.08.001] [Medline: 30711062]
40. Bonevski B, Bryant J, Lynagh M, Paul C. Money as motivation to quit: a survey of a non-random Australian sample of socially disadvantaged smokers' views of the acceptability of cash incentives. Prev Med 2012 Aug;55(2):122-126 [FREE Full text] [doi: 10.1016/j.ypmed.2012.06.001] [Medline: 22691939]
41. Sindelar JL, O'Malley SS. Financial versus health motivation to quit smoking: a randomized field study. Prev Med 2014 Mar;59:1-4 [FREE Full text] [doi: 10.1016/j.ypmed.2013.10.008] [Medline: 24139975]
42. Haff N, Patel MS, Lim R, Zhu J, Troxel AB, Asch DA, et al. The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: a meta-analysis. Am J Health Promot 2015;29(5):314-323. [doi: 10.4278/ajhp.140714-LIT-333] [Medline: 25928816]
43. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JP, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis. Prev Med 2015 Jul;75:75-85 [FREE Full text] [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]
44. Dallery J, Raiff BR, Grabinski MJ. Internet-based contingency management to promote smoking cessation: a randomized controlled study. J Appl Behav Anal 2013 Dec;46(4):750-764 [FREE Full text] [doi: 10.1002/jaba.89] [Medline: 24114862]

45. Raiff BR, Jarvis BP, Turturici M, Dallery J. Acceptability of an internet-based contingency management intervention for smoking cessation: views of smokers, nonsmokers, and healthcare professionals. Exp Clin Psychopharmacol 2013 Jul;21(3):204-213 [FREE Full text] [doi: 10.1037/a0032451] [Medline: 23750691]
46. Stoops WW, Dallery J, Fields NM, Nuzzo PA, Schoenberg NE, Martin CA, et al. An internet-based abstinence reinforcement smoking cessation intervention in rural smokers. Drug Alcohol Depend 2009 Dec 1;105(1-2):56-62 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.06.010] [Medline: 19615830]
47. Dallery J, Meredith S, Jarvis B, Nuzzo PA. Internet-based group contingency management to promote smoking abstinence. Exp Clin Psychopharmacol 2015 Jul;23(3):176-183 [FREE Full text] [doi: 10.1037/pha0000013] [Medline: 25821915]
48. Dallery J, Raiff BR, Kim SJ, Marsch LA, Stitzer M, Grabinski MJ. Nationwide access to an internet-based contingency management intervention to promote smoking cessation: a randomized controlled trial. Addiction 2017 May;112(5):875-883 [FREE Full text] [doi: 10.1111/add.13715] [Medline: 27923264]
49. Hertzberg JS, Carpenter VL, Kirby AC, Calhoun PS, Moore SD, Dennis MF, et al. Mobile contingency management as an adjunctive smoking cessation treatment for smokers with posttraumatic stress disorder. Nicotine Tob Res 2013 Dec;15(11):1934-1938 [FREE Full text] [doi: 10.1093/ntr/ntt060] [Medline: 23645606]
50. Carpenter VL, Hertzberg JS, Kirby AC, Calhoun PS, Moore SD, Dennis MF, et al. Multicomponent smoking cessation treatment including mobile contingency management in homeless veterans. J Clin Psychiatry 2015 Jul;76(7):959-964 [FREE Full text] [doi: 10.4088/JCP.14m09053] [Medline: 25699616]
51. Kong G, Goldberg AL, Dallery J, Krishnan-Sarin S. An open-label pilot study of an intervention using mobile phones to deliver contingency management of tobacco abstinence to high school students. Exp Clin Psychopharmacol 2017 Oct;25(5):333-337 [FREE Full text] [doi: 10.1037/pha0000151] [Medline: 29048181]
52. SCRAM: Systems Alcohol and Location Monitoring. 2020. Automated Facial Intelligence (AFI) URL: https://www.scramsystems.com/products/alcohol-monitoring/automated-facial-intelligence/ [accessed 2020-01-03]
53. SoberLink: The Experts in Remote Alcohol Monitoring. 2020. URL: https://www.soberlink.com/ [accessed 2020-01-03]
54. Oklahoma Tobacco Research Center. 2019. Insight Platform: About Insight URL: https://otrc.stephensoncancercenter.org/Mobile-Health-Technology/Insight-Platform [accessed 2019-08-02]
55. Bedfont Scientific. 2018. iCO Smokerlyzer: The World's First Smokerlyzer Breath CO Monitor for Your Smartphone/Tablet URL: https://www.bedfont.com/ico [accessed 2019-08-02]
56. Microsoft Docs. 2019. What is the Azure Face Service? URL: https://docs.microsoft.com/en-us/azure/cognitive-services/face/overview [accessed 2019-08-16]
57. Greenphire. 2019. ClinCard: Automated Participant Payments URL: https://greenphire.com/clincard/ [accessed 2019-08-02]
58. Oklahoma Tobacco Research Center. 2020. Tobacco Treatment Research Program URL: https://otrc.stephensoncancercenter.org/Free-Help-to-Quit [accessed 2020-02-06]
59. Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007 Dec;45(11):1026-1033. [doi: 10.1097/MLR.0b013e3180616c1b] [Medline: 18049342]
60. Benowitz N, Bernert J, Foulds J, Hecht S, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2019 Oct 1:- epub ahead of print. [doi: 10.1093/ntr/ntz132] [Medline: 31570931]
61. US Department of Health and Human Services, Public Health Service. Treating Tobacco Use and Dependence: 2008 Update. Scotts Valley, California, US: CreateSpace Independent Publishing Platform; 2008.
62. Harris P, Taylor R, Minor B, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
63. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 May;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
64. Grother P, Ngan M, Hanaoka K. National Institute of Standards and Technology. 2019. Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects URL: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf [accessed 2020-01-07]

Abbreviations

API: application programming interface
CM: contingency management
CO: carbon monoxide
CTTS: certified tobacco treatment specialist
EMA: ecological momentary assessment
mHealth: mobile health
NCI: National Cancer Institute
NRT: nicotine replacement therapy
ppm: parts per million

SES: socioeconomic status
TTRP: Tobacco Treatment Research Program

Edited by G Eysenbach; submitted 21.08.19; peer-reviewed by B Raiff, S Lepore, P Keller; comments to author 18.09.19; revised version received 17.01.20; accepted 22.03.20; published 15.04.20.

Please cite as:
Kendzor DE, Businelle MS, Waring JJC, Mathews AJ, Geller DW, Barton JM, Alexander AC, Hébert ET, Ra CK, Vidrine DJ
Automated Mobile Delivery of Financial Incentives for Smoking Cessation Among Socioeconomically Disadvantaged Adults: Feasibility Study
JMIR Mhealth Uhealth 2020;8(4):e15960
URL: http://mhealth.jmir.org/2020/4/e15960/
doi:10.2196/15960
PMID:32293568

Original Paper

Impact of a Mobile Phone App to Increase Vegetable Consumption and Variety in Adults: Large-Scale Community Cohort Study

Gilly A Hendrie[1], PhD; M Sazzad Hussain[2], PhD; Emily Brindal[1], PhD; Genevieve James-Martin[1], BSc; Gemma Williams[1], BSc; Anna Crook[3]

[1]Nutrition and Health Program, CSIRO Health & Biosecurity, Adelaide, Australia

[2]The Australian eHealth Research Center, CSIRO Health & Biosecurity, Sydney, Australia

[3]SP Health Co. Pty Ltd, Sydney, Australia

Corresponding Author:
Gilly A Hendrie, PhD
Nutrition and Health Program
CSIRO Health & Biosecurity
PO BOX 10041
Adelaide
Australia
Phone: 61 88305 ext 0662
Email: gilly.hendrie@csiro.au

Abstract

Background: Large-scale initiatives to improve diet quality through increased vegetable consumption have had small to moderate success. Digital technologies have features that are appealing for health-related behavior change interventions.

Objective: This study aimed to describe the implementation and evaluation of a mobile phone app called VegEze, which aims to increase vegetable intake among Australian adults.

Methods: To capture the impact of this app in a real-world setting, the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework was utilized. An uncontrolled, quantitative cohort study was conducted, with evaluations after 21 and 90 days. The app was available in the Apple App Store and was accompanied by television, radio, and social media promotion. Evaluation surveys were embedded into the app using ResearchKit. The primary outcomes were vegetable intake (servings per day) and vegetable variety (types per day). Psychological variables (attitudes, intentions, self-efficacy, and action planning) and app usage were also assessed. Descriptive statistics and multiple linear regression were used to describe the impact of the app on vegetable intake and to determine the characteristics associated with the increased intake.

Results: Data were available from 5062 participants who completed the baseline survey; 1224 participants completed the 21-day survey, and 273 completed the 90-day survey. The participants resided across Australia and were mostly women (4265/5062, 84.3%) with a mean age of 48.2 years (SD 14.1). The mean increase in intake was 0.48 servings, from 3.06 servings at baseline to 3.54 servings at the end of the 21-day challenge (t_{1223}=8.71; P<.001). The variety of vegetables consumed also increased by 0.35 types per day (t_{1123}=9.59; P<.001). No changes in intake and variety were found from day 21 to the 90-day follow-up. Participants with the highest app usage increased their vegetable intake by 0.63 (SD 2.02) servings per day compared with 0.32 (SD 1.69) servings per day for those with the lowest app usage. On the basis of multiple linear regression, gender; age; BMI; psychological variables of self-efficacy, attitudes, intentions, and action planning specific to vegetable intake; baseline vegetable intake; and active days of app usage accounted for 23.3% of the variance associated with the change in intake ($F_{9,1208}$=42.09; P<.001). Baseline vegetable intake was the strongest predictor of change in intake (beta=−.495; P<.001), with lower baseline intake associated with a greater change in intake. Self-efficacy (beta=.116; P<.001), action planning (beta=.066; P=.02), BMI (beta=.070; P=.01), and app usage (beta=.081; P=.002) were all significant predictors of the change in intake.

Conclusions: The VegEze app was able to increase intake by half a serving in a large sample of Australian adults. Testing the app in a real-world setting and embedding the consent process allowed for greater reach and an efficient, robust evaluation. Further work to improve engagement is warranted.

(JMIR Mhealth Uhealth 2020;8(4):e14726) doi:10.2196/14726

KEYWORDS

mHealth; vegetables; healthy diet; intervention study

Introduction

Vegetables as a Target for Nutrition Intervention

Poor diet quality is a risk factor for the development of chronic disease [1] and, along with physical activity, a key modifiable behavioral target for health interventions. Poor quality diets are generally characterized by the inadequate consumption of fruits and vegetables, which are thought to account for 16.0 million (1.0%) disability-adjusted life years and 1.7 million (2.8%) deaths worldwide [2]. Hence, there have been many population-level initiatives addressing inadequate fruit and vegetable consumption that have achieved small to moderate improvements in intake [3-5]. Sustained improvements are more difficult to achieve, and in many countries such as Australia, the United States, and the United Kingdom, gaps between population intakes and national dietary recommendations persist [6-9].

Most large-scale population campaigns [7] and research interventions target fruit and vegetable consumption concurrently [10,11], with very few interventions focusing solely on vegetable intake. Targeting vegetables in isolation is needed as 95% of the Australian population do not meet the recommended intake levels for vegetables (compared with about 50% for fruits) [6]; there are more barriers for the consumption of vegetables than for fruits; there is a poorer understanding of serving sizes for vegetables [12], and when fruits and vegetables are targeted together, increases in consumption are largely associated with fruit and not vegetable intake [13].

Benefits of Digital Technology for Behavior Change

Digital technologies are appealing for health-related behavior change interventions as they may overcome some of the limitations of traditional delivery approaches. For example, the ubiquitous nature of mobile phone ownership and its broad application and usage mean that mobile health (mHealth) interventions have the potential to reach large audiences at nearly any time or place. Mobile technology can also be highly interactive and can be used to deliver health-related information in a way that is engaging and rewarding. The ability to tailor content over time based on user inputs or objective measures (eg, wearable devices) can create personalization, which may increase engagement—the equivalent of exposure to or a dose of a traditional intervention—increasing the likelihood of success [14]. Mobile phones can also make self-monitoring and tracking easy, which, along with timely feedback [15,16], are key drivers and predictors of behavior change. From an implementation perspective, digital technologies may provide a cost-effective and scalable approach to deliver health interventions. This aspect is also appealing to researchers and practitioners as traditional public health interventions are often resource intensive, affecting the service delivery and potential impact.

Mobile Health Interventions: Effectiveness and Validity

Preliminary evidence from mHealth interventions shows promise as they appear to be feasible and acceptable to users [17], but their success in changing behavior is less well established. To maximize the likelihood of success, digital interventions should be based on behavior change theories and utilize evidence-based content [18,19]. Many commercially available mobile phone apps are not evidence-based or grounded in behavior change theories [16,19], and those which do utilize the theories report larger intervention effects [16]. Even nutrition research apps that have been published and evaluated within the scientific domain rarely examine and adequately report the effects on dietary behavior change [20].

Research apps are typically evaluated using approaches such as randomized controlled trials, which emphasize internal validity as opposed to external validity. Such approaches provide robust evidence for efficacy in the study sample but have limited reach and provide little insight into their generalizability to the target population [21]. Apps that have been scientifically developed and tested also face challenges related to the slow pace of traditional research and lack of resources available for the eventual translation of these mHealth interventions for more widespread use [21,22].

Objective

This study aimed to implement and evaluate the impact of a mobile phone app called VegEze, which aims to improve vegetable intake among Australian adults. The Australian Dietary Guidelines recommend “plenty of vegetables, including different types and colours” [23]; therefore, this app focused on increasing the amount and variety of vegetables consumed. Increasing the variety of vegetables has been shown to be an effective strategy to increase the consumption of vegetables in a single meal [24-26] but has not been targeted in previous large-scale digital interventions for adults. The target behavior for the VegEze app was *having 3 different types of vegetables at dinner.* A detailed description of the VegEze app and its development has been published [27].

To better capture the impact of this app in a real-world setting, the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework was utilized. The framework has 5 components—reach, effectiveness, adoption, implementation, and maintenance—which are commonly used to translate research and understand impact and generalizability across populations [28,29]. This framework informed our key research questions, which are as follows:

- How many Australian adults are willing to participate in the VegEze intervention? (Reach)
- What is the impact of VegEze on increasing vegetable variety and intake? (Effectiveness)
- Who used the VegEze app most? (Adoption)
- How do participants use the features of the VegEze app and is app usage associated with success? (Implementation)

- Does the VegEze app support participants to maintain their consumption for a longer term? (Maintenance)

Methods

Study Design and Participant Recruitment

An uncontrolled, quantitative cohort study was used to assess the impact of the VegEze app on the daily intake and variety of vegetables, with an evaluation conducted after 21 and 90 days of the program. The VegEze research study launched in the Apple App Store (free of charge) on November 8, 2017. To facilitate timely dissemination of the results, data for inclusion in this evaluation were extracted after 6 months (app download and baseline survey were completed between November 2017 and May 2018).

There was associated media coverage, including free-to-air television and radio interviews as well as social media promotions from November 13, 2017. Emails were also sent to an existing database of people who had opted in to receive nutrition-related newsletters.

Participants were eligible for the study if they were aged 18 years and above, were living in Australia, owned a compatible iPhone Operating System (iOS)10 or 11 device (including iPhone or iPad) with an internet connection, and were willing and able to download the app and participate in the trial. Those who had any condition or self-prescribed diet that prevented them from consuming vegetables were excluded. This study was approved by the Commonwealth Scientific & Industrial Research Organisation (CSIRO) Health and Medical Human Research Ethics Committee Low Risk Review Panel (proposal number 13/2017) and was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12618000481279).

Data Collection

All data were collected via mobile phones. We leveraged the Apple ResearchKit framework with the app because of its innovative features in transforming the way a large-scale cohort study can be delivered via a mobile phone. The onboarding process, including participant information and electronic consent, and the surveys were embedded into the app. Evaluation surveys were administered at 3 time points: baseline, end of the 21-day challenge, and at the 90-day follow-up. Although the surveys were designed to be as short as possible, they were still able to capture critical outcomes and predictors of behavior. Standard demographic questions about age, gender, and self-reported height and weight were also asked in the baseline survey only.

Outcomes

Primary Outcomes: Vegetable Consumption

The primary outcomes were vegetable intake (reported in servings per day) and vegetable variety (reported in types per day) and were assessed at each time point. Vegetable intake was assessed using a series of short questions from the previously validated CSIRO Healthy Diet Score survey [30,31], which asks about the usual frequency (daily, weekly, monthly, and never) and the amount consumed, in standard servings, within the timeframe selected. From this, servings per day was calculated and compared with the age- and gender-specific daily intake targets provided in the Australian Dietary Guidelines and was coded as meeting the guidelines if the intake was greater than or equal to the recommendation.

Vegetable variety was assessed in a single question asking about the number of different types of vegetables consumed in the past 2 days. Changes in vegetable intake and variety were calculated as end of challenge minus baseline consumption, where a positive value indicated an increase in consumption. One question asked about the frequency (always, usually, sometimes, and never) of achieving the target behavior, ie, having 3 different types of vegetables at dinner.

Psychological Variables

Attitudes, intentions, and self-efficacy are 3 well-established personal factors that have been shown to predict changes in health behavior [32,33]. More recently, it has been suggested that having action plans about how to implement the target behavior is also a critical predictor of behavior [32]. These 4 constructs were measured at each time point using previously validated scales, and participants responded on Likert scales. Nutrition self-efficacy was assessed using 5 questions [34]. Intentions [35] and attitudes [36] toward consuming more vegetables were assessed using 3 questions each, and action planning was assessed in 2 questions [37]. Reponses to questions were summed, and a higher score represented a higher amount of each construct.

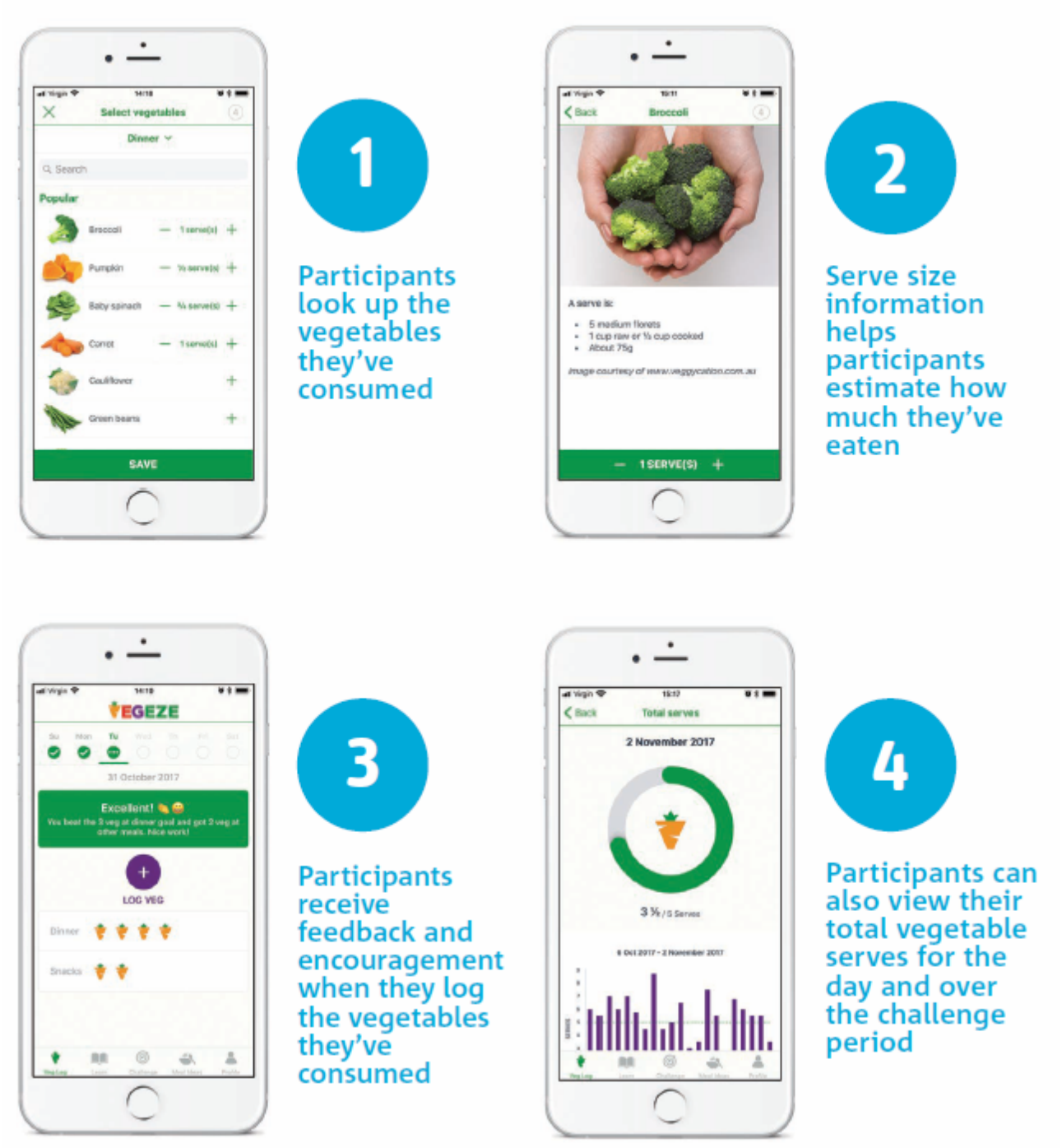
App Usage

Firebase by Google was used to collect extensive app-related data, including screen views, engagement with notifications, and app events (ie, interactions such as tapping navigation buttons), which helped to explain the behavior of participants when using the app.

Intervention: VegEze App Design

A detailed description of the development process has been published previously [27]. VegEze is an Apple iOS–based app that was launched as a 21-day challenge to get Australians in the healthy habit of eating more vegetables, starting with 3 different types of vegetables in their evening or main meal. The app contained easy and fun ways to help establish this habit and encouraged individuals to monitor their vegetable intake with an easy-to-use tracker for logging the amount and types of vegetables consumed at each meal. Tracking intake was done through the vegetable log, which was the core feature of the app (Figure 1). Here, participants could search for vegetables they had consumed using an alphabetized list or a keyword search and record the amount they consumed in servings. Serving size information was available to assist with portion size estimation.

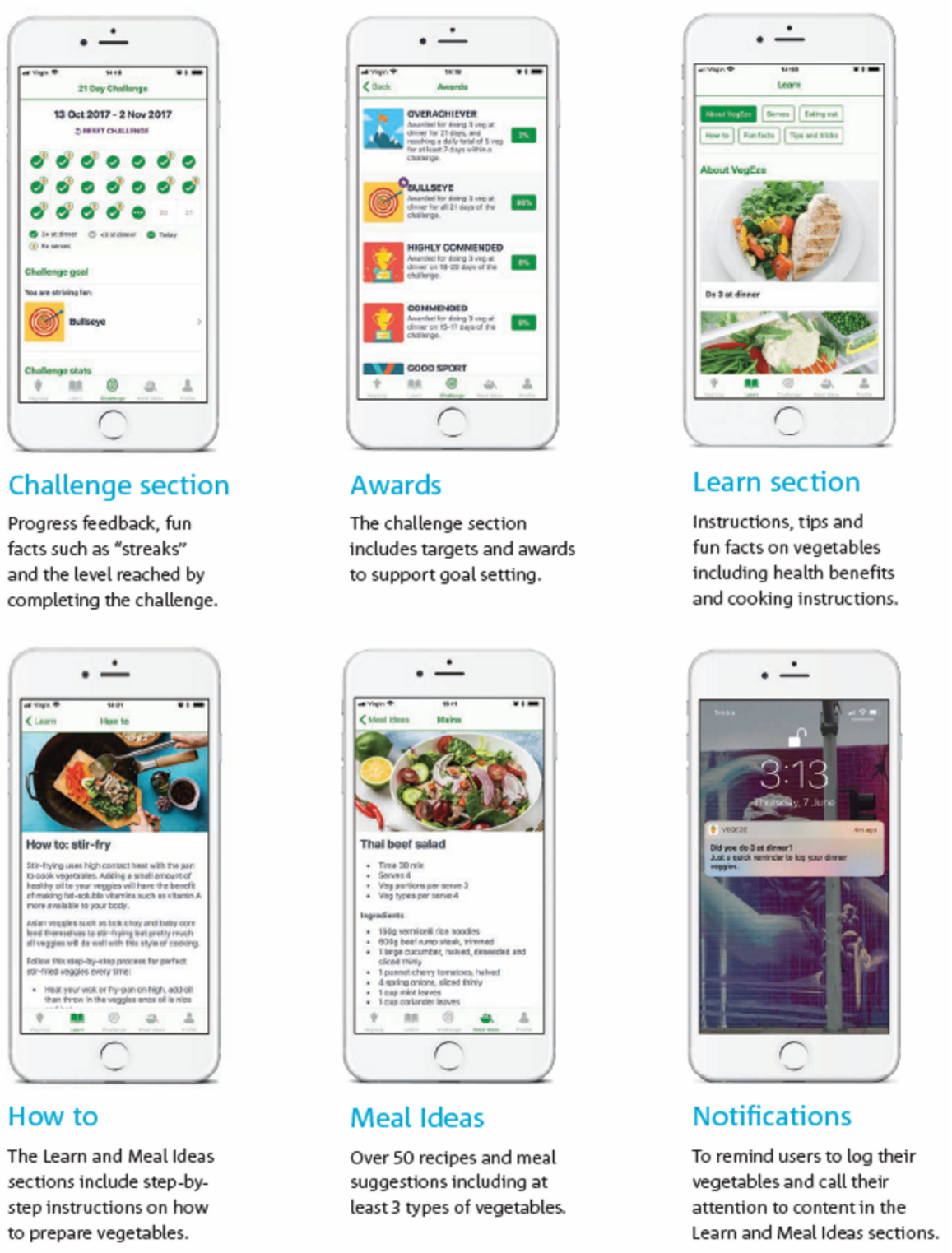
Figure 1. Vegetable log feature as a core component of the VegEze app.



Two-way user feedback was also central to the app. On the home screen, users could review their daily progress for the current day (Figure 1). Feedback and motivational messages were also displayed here when participants logged their vegetable intake. The previous day's log could be reviewed with a *check mark* indicating that the goal of 3 types was met. With a swipe from the home screen, users could also review their progress toward reaching the recommended number of servings per day (Figure 1).

Other features of the app included a challenge and awards section that provided awards to support goal setting and further feedback on the level reached in the challenge. Content sections were divided into *Learn*, which provided tips and fun, evidence-based facts about vegetables; *How to*, which provided step-by-step instructions on how to prepare vegetables; and *Meal Ideas*, which provided over 50 recipes and meal suggestions for meeting the target behavior. Notifications to log intake and view content were also included at varying frequencies (Figure 2).

Figure 2. Other features of the VegEze app.



Data Processing

The app usage and interaction logs were extracted from Big Query by Google into Excel files and processed in R. Any logs related to app updates, iOS updates, and errors were removed, and time stamps were converted from microseconds to the date and time format. Participants' app log data for 21 days from the date of starting the challenge were used for the analysis.

Google Analytics was used to interpret high-level app usage for the entire cohort. The app usage data were linked via a unique ID to a custom database that recorded participation data such as vegetable logs, demographic information, and survey submissions. Structured Query Language queries were used to extract summarized participation data into Comma-Separated Values files for analysis.

Membership duration was calculated for individuals as the length of time (in days) between the start date of the challenge

and the last date of app usage. The attrition rate of the app over 21 days of the program was calculated using the membership duration. The number of active days for which individuals visited the app (ie, app usage) and the feature usage (Home, Veg Lookup, Challenges, Notifications, Meal Ideas, and Learn) were calculated for users who had at least one log instance for that day.

Survey data were extracted from a service called SurveyGizmo into IBM SPSS Statistics files. Extreme outliers were removed based on vegetable intake that was greater than 12 servings of vegetables per day, reported at any time point (equivalent to the mean±3SD).

Statistical Analysis

Descriptive statistics (means, standard deviations, and percentages) were used to describe the characteristics of the sample. We applied a per-protocol analysis to calculate the significance of the change in vegetable consumption between baseline and the end of the 21-day challenge. This change was tested using paired samples *t* tests for the sample as a whole, and within demographic subgroups of gender, age, and weight status. The differences in the change in consumption among demographic subgroups were assessed using 1-way analysis of variance. Multiple linear regression was used to assess the user characteristics (gender, age, BMI, baseline vegetable consumption, self-efficacy, attitudes, intentions, action planning, and app usage) that were associated with the change in consumption. User characteristics (predictors) were added to the model simultaneously. Descriptive statistics were used to describe app usage. Tertiles of app usage were created, and the mean changes in vegetable intake and variety were examined to understand how adaptation related to effectiveness. As an indication of maintenance of behavior change, the change in intake was examined between baseline and 90-day follow-up using paired samples *t* tests. Significance was assumed at a level of $P<.05$. Analyses were conducted using IBM SPSS Statistics version 25.

Power

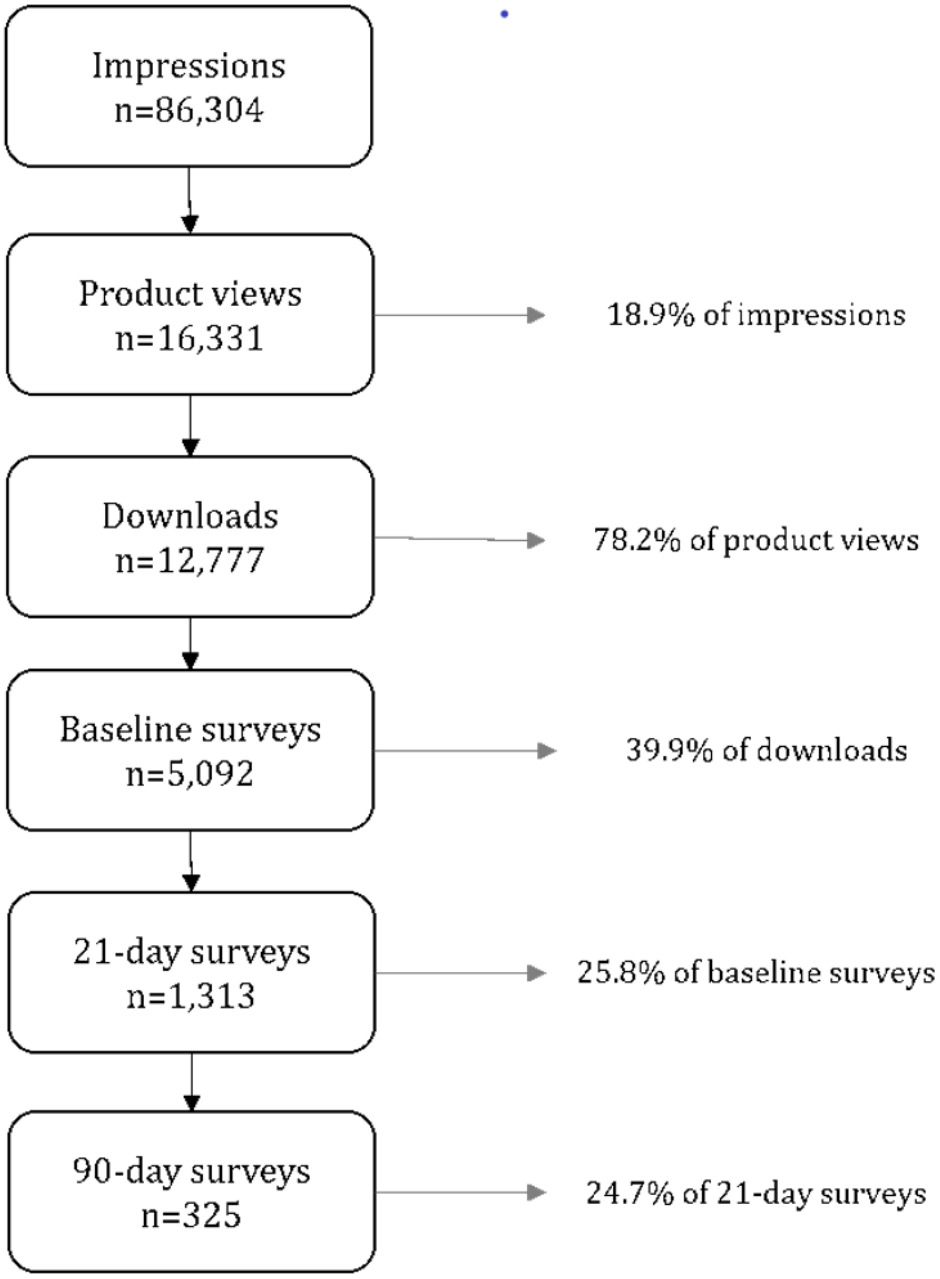
A previous review of digital interventions to increase vegetable intake, albeit in an adolescent population, reported increases of 0.1 to 0.4 servings of vegetables per day [10]. An a priori power calculation using the G*Power software [38] indicated that a sample of 272 people would be needed to detect a small effect size ($d=0.2$) with 95% power using a *t* test for the difference between 2 dependent means with alpha at .05. This was considered as the minimum number of participants needed per group to allow for a subgroup analysis. Variables such as age group and weight status have up to 4 subgroups; therefore, we aimed to retain 1100 participants until the end of the 21-day challenge. We expected 20% of the baseline sample to complete the 21-day survey; therefore, we aimed to recruit approximately 5000 participants.

Results

Description of the Sample

The media coverage resulted in over 86,000 impressions within the App Store, over 16,000 product views, and 12,777 people downloading the VegEze app (Figure 3). Data were available from 5092 participants who completed the baseline survey, 1313 participants who had completed the 21-day end-of-challenge survey, and 325 participants who had completed the 90-day follow-up survey during this period. Given that the study was purely Web-based and there was no contact with participants outside of the app, reasons for nonparticipation at each stage were not collected.

Figure 3. A flow chart of participants in the VegEze study.



Following data cleaning, survey data, including vegetable intake measures, were available for 5062 participants at baseline, 1224 participants at the end of the 21-day challenge, and 273 participants at the 90-day follow-up. At baseline, 4683 users had both app log and survey data, with 1219 having app and survey data at the end of the 21-day challenge.

Participants who installed the app and completed the baseline survey were from across Australia, were mostly women (4265/5062, 84.3%), and were aged between 31 and 70 years (4183/5062, 82.7%; the mean age of the sample was 48.2 years, SD 14.1). The sample had a greater proportion of obese adults than the Australian population (590/5062, 31.4% in the sample vs 27.5% in the Australian population) and consequently a lower proportion in the normal weight category (Table 1).

Table 1. Demographic characteristics of the baseline sample of participants and their comparison with the Australian population.

Demographic characteristics	Sample (N=5062), n (%)	Australian population[a], %
Gender		
Male	774 (15.29)	49.4
Female	4265 (84.25)	50.6
Unisex	23 (0.45)	—[b]
Age group (years)		
18-30	675 (13.33)	18.6
31-50	1997 (39.45)	37.7
51-70	2186 (43.18)	30.5
≥71	204 (4.03)	13.1
Weight status		
Underweight	52 (1.02)	1.7
Normal weight	1587 (31.35)	35.5
Overweight	1833 (36.21)	35.3
Obese	1590 (31.41)	27.5
Australian state or territory		
New South Wales	1529 (30.21)	32.2
Victoria	1484 (29.32)	24.9
Queensland	916 (18.10)	20.1
Western Australia	446 (8.81)	10.4
South Australia	435 (8.60)	7.4
Tasmania	76 (1.50)	2.3
Northern Territory	12 (0.24)	1.0
Australian Capital Territory	93 (1.84)	1.7

[a]Australian population estimates were taken from the 2016 census, available from the Australian Bureau of Statistics [39].

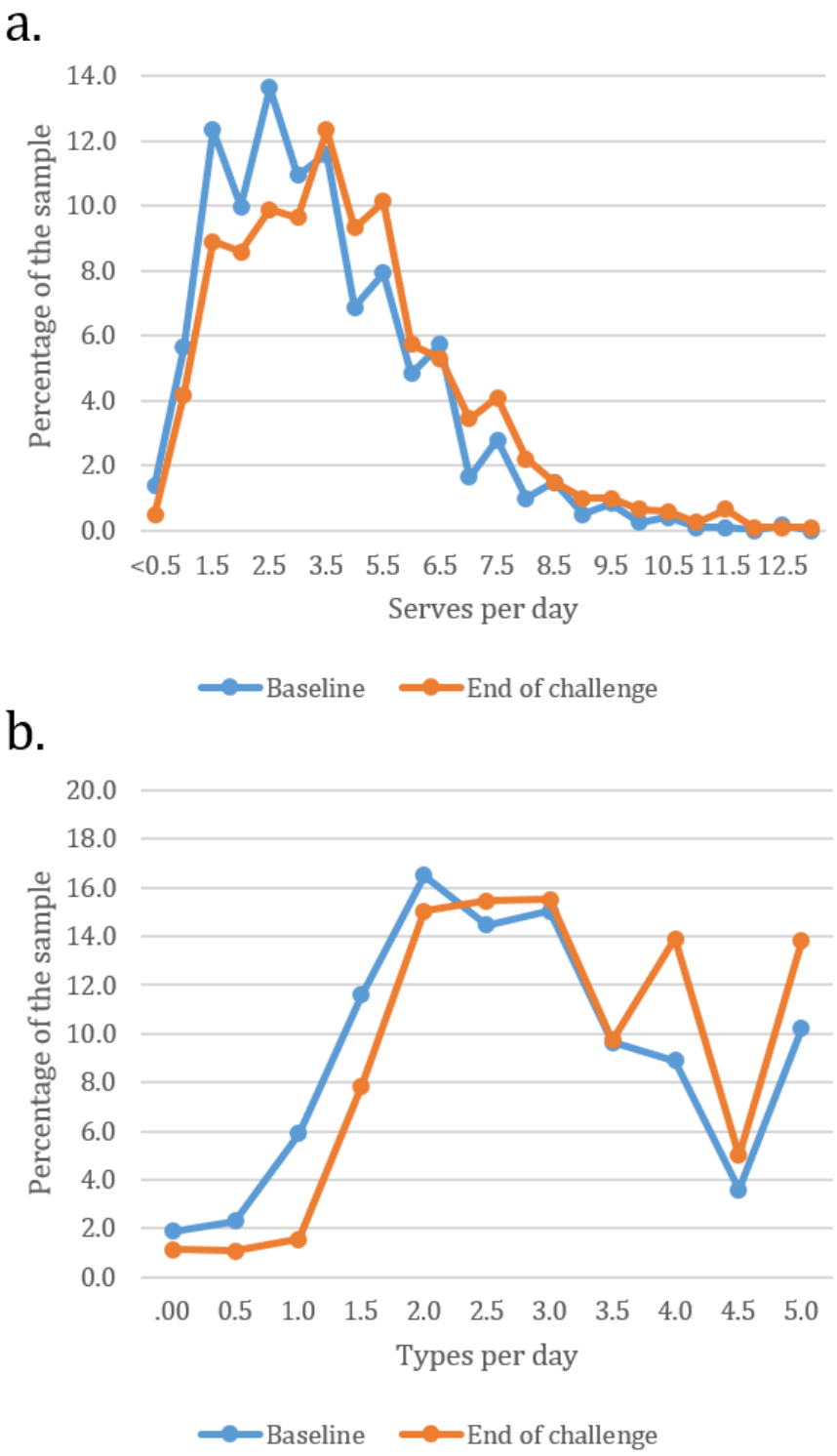
[b]Data unavailable.

Average vegetable intake per day was 3.1 servings (SD 1.9) and 2.5 types (SD 1.3) at baseline. This translates to 14.4% (729/5062) meeting the Australian Dietary Guidelines recommendation for vegetable intake. Furthermore, 22.1% (1119/5062) of the sample reported of *always* having 3 different types of vegetable at dinner, and 64.5% (3265/5062) *always* or *usually* achieved this target.

Impact on Vegetable Consumption

In the sample that completed the 21-day survey (n=1224), the distribution of vegetable intake, in servings and types, at baseline and at the end of the 21-day challenge is shown in Figure 4. The mean increase in consumption was 0.48 servings, from 3.06 servings at baseline to 3.54 servings at the end of the challenge (t_{1223}=8.71; *P*<.001). The variety of vegetables consumed also increased by 0.35 types per day (t_{1123}=9.59; *P*<.001).

Figure 4. Distribution of (a) vegetable intake and (b) vegetable variety at baseline and at the end of the 21-day challenge.



The changes in vegetable intake varied for different demographic groups. Women significantly increased their number of servings (0.51 servings; t_{1077}=8.74; *P*<.001) and types (0.36 types; t_{1077}=9.31; *P*<.001). The change in consumption for men was significant for the types (0.23 types; *P*=.02) but not the amount (0.25 servings; *P*=.13;). Significant increases in intake were also observed among participants in the age groups of 31 to 50 years and 51 to 70 years (0.50-0.53 servings; *P*<.001). All categories of weight status significantly increased intake (ranging from 0.40-0.53 servings; all *P*<.001; Table 2). The between-group differences for change in consumption were not significant for gender, age group, or weight status.

Table 2. Change in vegetable consumption (servings and types) at baseline and at the end of the 21-day challenge by demographic characteristics.

Demographic characteristics	Sample (N=1224), n (%)	Amount in servings				Number of varieties			
		Baseline consumption, mean (SD)	End of challenge (day 21), mean (SD)	Change, mean (SD)	*P* value	Baseline consumption, mean (SD)	End of challenge (day 21), mean (SD)	Change, mean (SD)	*P* value
Gender[a]									
Male	139 (11.36)	3.06 (1.98)	3.30 (2.00)	0.25 (1.92)	.134	2.55 (1.26)	2.78 (1.13)	0.23 (1.17)	.02
Female	1078 (88.07)	3.07 (1.72)	3.58 (1.92)	0.51 (1.92)	<.001	2.78 (1.24)	3.15 (1.19)	0.36 (1.29)	<.001
Total	1224 (100.00)	3.06 (1.76)	3.54 (1.93)	0.48 (1.92)	<.001	2.75 (1.25)	3.10 (1.19)	0.35 (1.27)	<.001
Age group (years)									
18-30	95 (7.76)	3.08 (1.68)	3.30 (1.87)	0.22 (1.74)	.225	2.73 (1.33)	3.10 (1.17)	0.38 (1.30)	.01
31-50	411 (33.58)	2.75 (1.63)	3.28 (1.76)	0.53 (1.83)	<.001	2.75 (1.26)	3.09 (1.20)	0.34 (1.32)	<.001
51-70	654 (53.43)	3.24 (1.81)	3.74 (2.01)	0.50 (2.01)	<.001	2.77 (1.22)	3.12 (1.19)	0.35 (1.26)	<.001
≥71	63 (5.15)	3.23 (1.88)	3.54 (2.00)	0.31 (1.79)	.172	2.57 (1.32)	2.94 (1.08)	0.37 (1.03)	.01
Weight status[a]									
Normal weight	392 (32.03)	3.01 (1.68)	3.40 (1.89)	0.40 (1.84)	<.001	2.90 (1.22)	3.27 (1.17)	0.37 (1.19)	<.001
Overweight	436 (35.62)	3.08 (1.78)	3.61 (1.98)	0.53 (1.99)	<.001	2.69 (1.26)	3.05 (1.19)	0.36 (1.29)	<.001
Obese	386 (31.53)	3.10 (1.80)	3.61 (1.91)	0.51 (1.94)	<.001	2.67 (1.24)	2.99 (1.19)	0.32 (1.33)	<.001

[a]An inadequate sample size to report on unisex (n=6) or underweight categories (n=9).

In this sample, the proportion of the sample meeting the dietary guidelines recommendation for vegetable consumption increased from 15.9% (195/1224) at baseline to 22.6% (277/1224) at the end of 21 days. The proportion of the sample reporting of *always* having 3 different types of vegetables at dinner increased from 23.2% (284/1224) at baseline to 28.8% (353/1224) at the end of 21 days, and the percentage who were *always* or *usually* achieving this target behavior increased from 68.2% (835/1224) to 81.3% (995/1224).

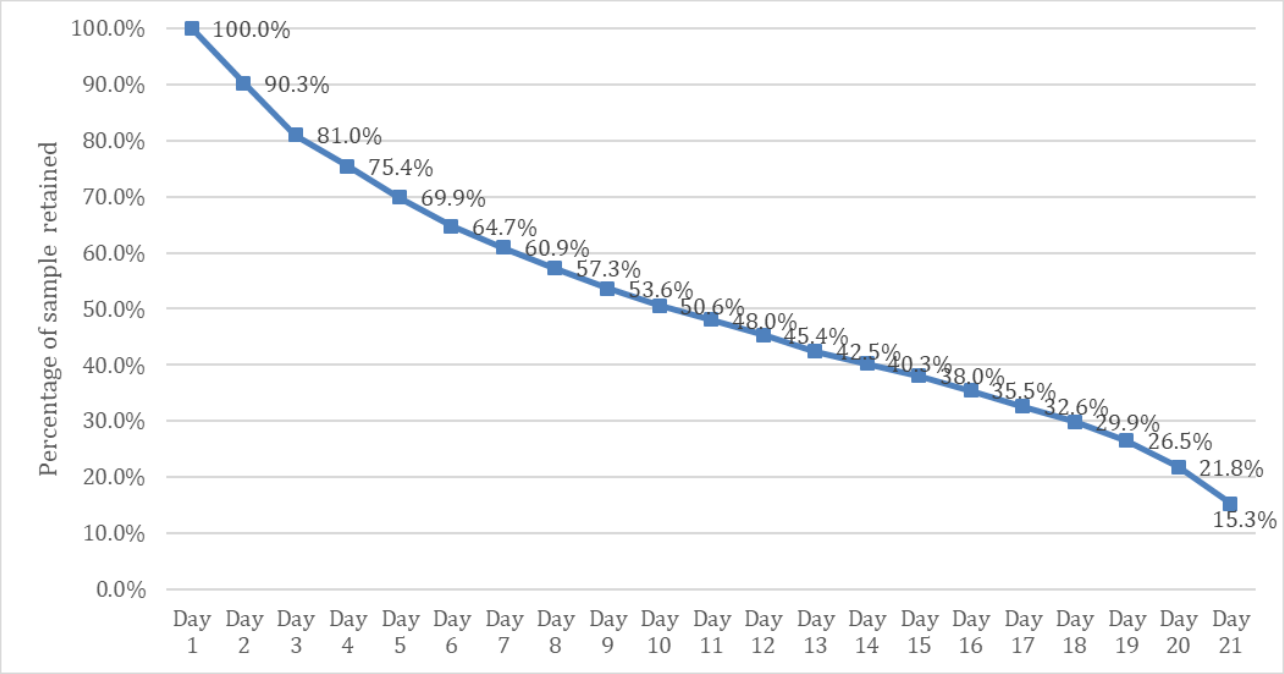
Impact on Psychological Factors

There was a small but significant increase in positive attitudes toward eating a greater variety of vegetables during the challenge period, albeit there was a highly positive attitude at baseline (3.84 to 3.87 points out of 4; t_{1223}=2.85; *P*=.005). Nutrition self-efficacy also increased significantly (14.54 to 15.04 out of 20; t_{1223}=6.09; *P*<.001), as did planning related to eating vegetables (3.27 to 3.60 points out of 6; t_{1223}=7.01; *P*<.001). The intention to eat a greater variety of vegetables (4.91 to 4.60 points out of 6; t_{1223}=−9.48; *P*<.001) and intention to use the app (5.33 to 4.49 points out of 6; t_{1223}=−19.20; *P*<.001) decreased significantly during the 21-day challenge.

App Usage Statistics

The attrition curve for the usage of the app is shown in Figure 5. There was a gradual reduction in the percentage of the sample retained. By day 10 of the challenge, about half the sample was using the app, and 21 days after completing the survey, 15.3% (719/4683) of participants were still using the app.

Figure 5. Participant attrition over the 21-day challenge (n=4683).



On average, participants actively used the app for 6.3 days out of the 21 days of the challenge. Furthermore, 49.2% (2304/4683) actively used the app for 2 to 7 days, 19.1% (894/4683) used the app for 8 to 14 days, 11.6% (543/4683) for 15 to 20 days, and 1.2% (56/4683) used the app every day during the 21-day challenge period (Table 3). Women were more likely than men to use the app for more than 7 days during the challenge (1302/3956, 32.9% vs 184/707, 26.0%; data not shown). On average, women actively used the app for 6.4 days during the challenge compared with 5.6 days for men (P=.001). Those in the oldest 2 age groups used the app more than those in the 2 younger age groups (5.1 days vs 7.1 days for age groups 19-50 years and ≥51 years, respectively; P<.001). App usage did not differ among categories of weight status.

Table 3. Percentage of the sample actively using the VegEze app, for the baseline sample and for those that completed the program.

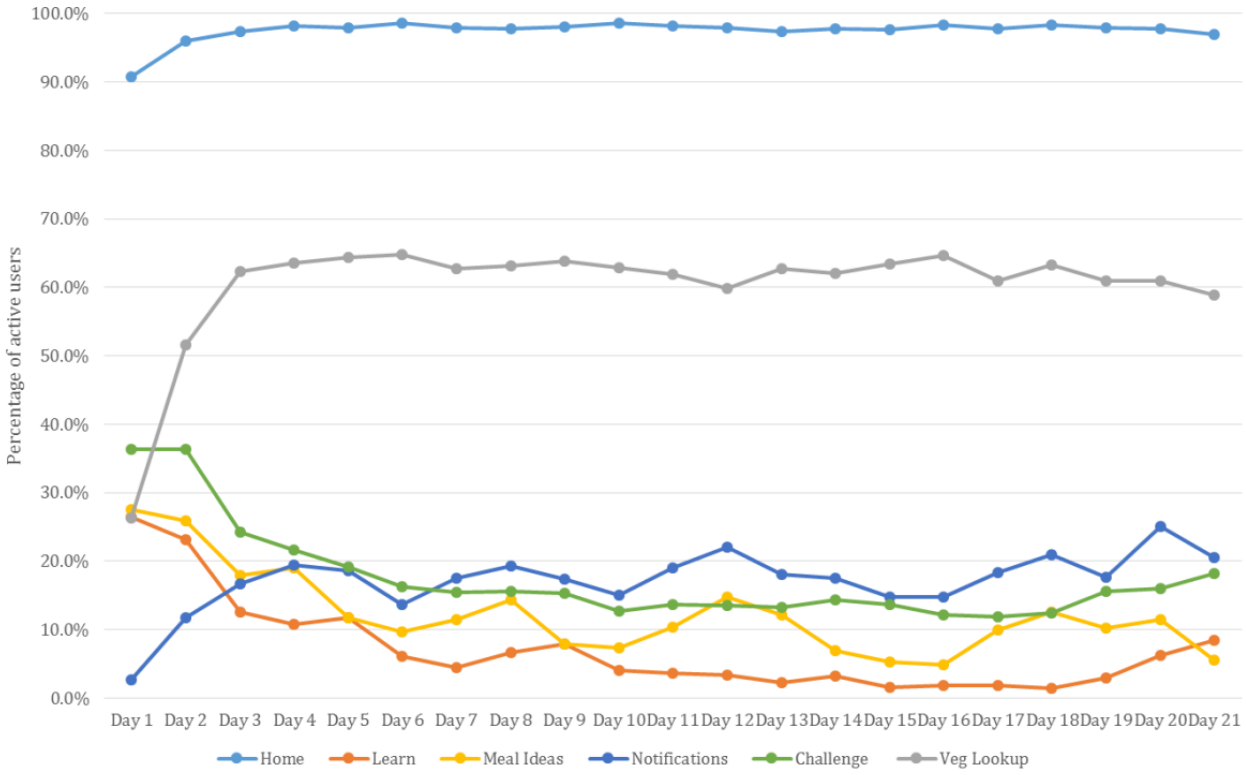
App usage	Baseline (n=4683)	Completers (n=1219)
Frequency of usage (days), %		
1	19.0	2.1
2-7	49.2	18.4
8-14	19.1	38.6
15-20	11.6	36.8
21	1.2	4.1
Feature usage, days		
Total usage	6.3	12.5
Home screen	6.1	12.3
Veg Lookup	3.7	8.3
Challenges	1.3	2.4
Meal Ideas	0.9	1.6
Notifications	1.0	2.0
Learn	0.6	1.1

The average active usage was higher among the participants who had app data and completed the 21-day survey (n=1219/1224), ie, 12.5 days out of the 21-day challenge. Furthermore, 36.8% (449/1219) of this sample used the app for 15 to 20 days, and 4.1% (50/1219) used the app every day during the 21-day challenge (Table 3).

Figure 6 shows the percentage of active users (n=4683) for each feature. Aside from the home screen, the most commonly used feature of the app was the Veg Lookup section. The usage of this feature increased from 26.3% (1232/4683) on day 1 to 62.3% (2918/4683) on day 3 and then stayed relatively stable, with about 60% to 63% of active users using this feature on any

day of the challenge. Viewing the notifications appeared to increase in the second half of the challenge period from about 15% of active users in the first half of the challenge period to about 20% in the later part. In contrast, the percentage of users viewing the challenge screen dipped from 36% of users on days 1 and 2 to between 12% and 15% at roughly day 10 and then peaked slightly again in the last few days of the challenge. The Meal Ideas and Learn sections of the app were used least frequently, with usage ranging from 5% to 25% of active users on any particular day during the challenge (Figure 6).

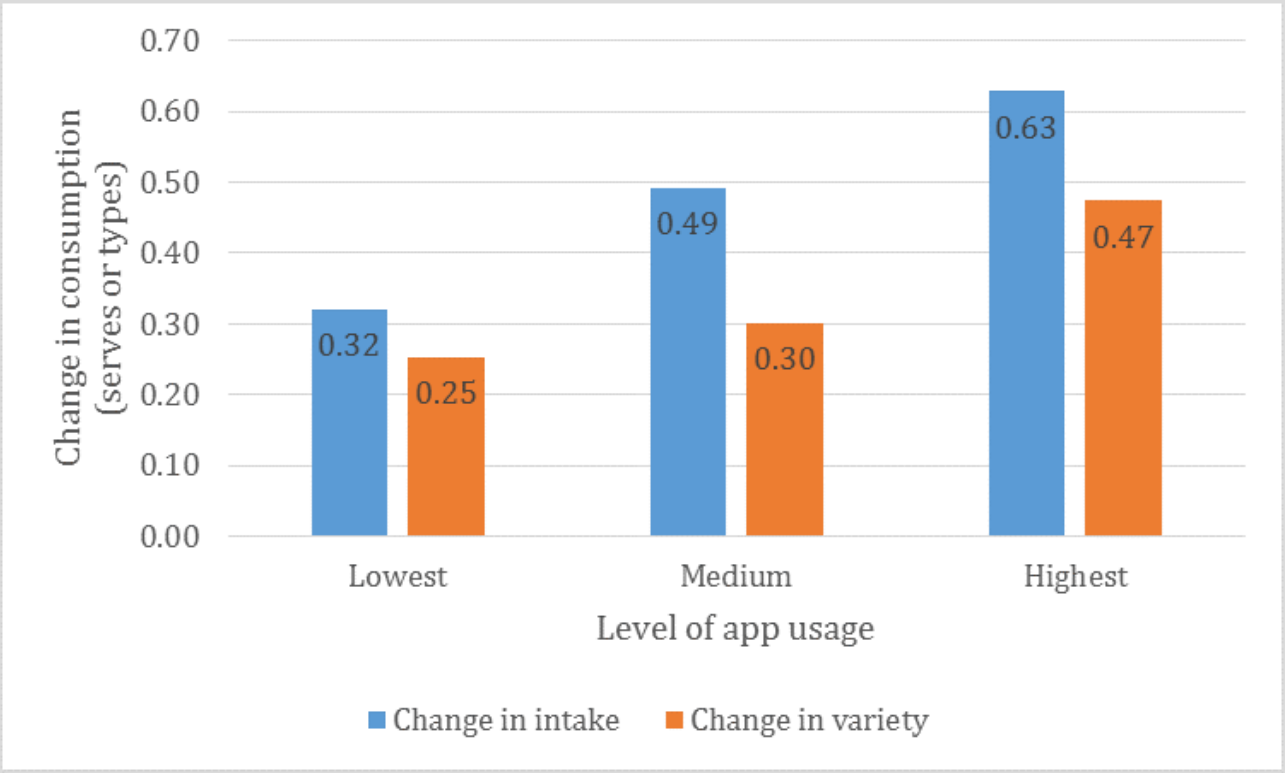
Figure 6. Percentage of active users using app features throughout the 21-day challenge period (n=4683).



Associations Between App Usage and Vegetable Consumption

When app usage was divided into tertile groups, it was found that participants with the highest usage had actively used the app almost every day during the 21-day challenge. This group increased their vegetable intake by 0.63 servings (SD 2.02) per day over the 21-day challenge compared with 0.32 servings (SD 1.69) per day for those with the lowest app usage (difference=0.31 servings; *P*=.06, Figure 7). Changes in the variety of vegetables consumed also increased in a stepwise manner with increasing app usage. Participants who used the app the most increased their consumption variety by 0.47 types (SD 1.26) per day, which was significantly more than those who used the app the least and increased their intake by 0.25 (SD 1.28) types per day (difference=0.22 types; *P*=.03, Figure 7).

Figure 7. Change in vegetable consumption by level of app usage* (n=1219). *Lowest tertile of app usage: active days ranged from 1 to -10 days, average 6.4 days; medium app usage: active days ranged from 11 to -15 days, average 13.1 days; and high app usage: active days ranged from 16 to -21 days, average 18.3 days.



On the basis of multiple linear regression, gender; age; BMI; psychological variables of self-efficacy, attitudes, intentions, and action planning specific to vegetables; baseline vegetable intake; and active days of app usage accounted for 23.3% of the variance associated with the change in intake ($F_{9,1208}$=42.09; P<.001). Baseline vegetable intake was the strongest predictor of change in intake (beta=−.495; P<.001), with lower baseline intake associated with greater change in intake. Self-efficacy (beta=.116; P<.001), action planning (beta=.066; P=.02), BMI (beta=.070; P=.01), and app usage (beta=.08; P=.002) were all significant predictors of change in intake. The same model was used to predict a change in variety. This model explained 32.3% of the variance ($F_{9,1208}$=65.46; P<.001). Baseline vegetable variety (beta=−.582; P<.001) and app usage (beta=.110; P<.001) were the strongest predictors of change in variety. Baseline attitudes toward vegetables (beta=.058; P=.02) and gender (beta=.050; P=.04) were the other significant predictors.

Maintaining Behavior Change for a Longer Term

Of those who completed the 90-day survey (n=273), 93% (254/273) were women, 25.6% (70/273) were aged between 31 and 50 years, 65.2% (178/273) were aged between 51 and 70 years, 34.1% (93/273) were overweight, and 36.6% (100/273) were obese. In this sample, vegetable intake increased significantly from 3.1 servings at baseline to 3.8 servings (average increase 0.68 servings; P<.001). Interestingly, this increase occurred during the first 21 days and was maintained during the follow-up period with an increase of 0.53 servings from baseline to the end of the 21-day challenge (P<.001) and a nonsignificant increase of 0.15 servings between days 21 and 90. Variety also increased significantly between baseline and the end of 21 days (average increase 0.48 types; P<.001) but not further during the extended follow-up period.

In this sample, the percentage of participants meeting the Australian Dietary Guidelines for vegetables increased from 16% (44/273) at baseline to 25% (68/273) and 26% (72/273) at the end of 21 and 90 days, respectively. Participants reporting of *always* having 3 different types of vegetables at dinner increased from 22.3% (61/273) at baseline to 32.2% (88/273) at the end of the 21-day challenge and then to 36.3% (99/273) at the end of the 90-day follow-up period. The percentage who *always* or *usually* had 3 different types of vegetables increased from 71.8% (196/273) at baseline to 83.5% (228/273) and 87.5% (239/273) at the end of the 21-day challenge and 90-day follow-up, respectively.

Discussion

Principal Findings

The VegEze app was designed to be an engaging 21-day challenge to increase the amount and variety of vegetables consumed by Australian adults. Central to this was a clear and specific behavioral target of having 3 different types of vegetables at dinner each day. At baseline, 22% of the sample reported *always* doing this and 68% reported *always or usually* doing this, which increased to 29% and 81%, respectively, at the end of the challenge. The dissemination pathway for this study was chosen to maximize its reach, and subsequently, the evaluation framework aimed to understand the impact and generalizability of the results within the population more broadly. The results of this study are discussed using the

RE-AIM framework [28] and within the context of other scientific literature.

Reach—How Many People Were Willing to Participate?

This research app was made available through the Apple App Store, and the use of ResearchKit negated the need for any face-to-face contact and streamlined the consent process. This novel approach combined with a structured promotional campaign allowed us to reach a large national sample of over 5000 participants within a relatively short period. That said, only 40% of those who downloaded the app completed the baseline survey. Therefore, including the evaluation surveys as a compulsory part of the onboarding process appeared to be a barrier to the overall uptake. This may be because of the time burden and delayed gratification associated with completing the surveys for people who simply wanted to download and explore the app.

Compared with the Australian population, the sample of people reached by the recruitment process was largely women (84% in the sample vs 51% in the Australian population) and slightly younger (4% aged over 70 years in the sample vs 13% of the Australian population). Other characteristics were fairly similar to the broader population. Furthermore, the 1200 participants who completed surveys at the beginning and at the end of the challenge may represent a biased sample of those more motivated to participate in research, and the results may overstate the likely impact on intake. Therefore, whether the reported changes in consumption are generalizable to the population more broadly is unknown.

Effectiveness—What Was the Impact on Vegetable Intake?

In terms of the impact of the intervention on critical outcomes, the app showed great promise. Evaluation data indicated that the 1224 users who completed the survey at the end of the challenge reported an average increase of 0.5 servings in daily vegetable intake and an increase in variety of 0.35 types over 21 days. This change is consistent with, or greater than, the changes reported by large-scale population campaigns [4] and other digital interventions focused on increasing vegetable intake [10,40]. A recent systematic review of electronic health and mHealth interventions for young adults reported increases in vegetable intake of between 0.1 and 0.4 servings per day [10]. Another study not included in the aforementioned review, developed and tested an app to increase vegetable intake among overweight adults who were already participating in a weight loss study. Participants with access to the app (n=68) increased their vegetable consumption by 0.8 servings over 35 days (5 weeks) compared with a control group who reported a decrease in consumption [15]. It is possible that people participating in weight management programs may be more motivated to change their eating behaviors as they are gaining the added and immediate reinforcement of weight loss.

We found positive changes in attitudes, self-efficacy, and action planning during the challenge period, which can help change behavioral intention into action [41]. However, it should be noted that attitudes toward eating a greater variety of vegetables were high at baseline, a likely indication of a motivated sample. We also found that higher levels of self-efficacy and action planning at baseline were associated with greater increase in vegetable consumption. Self-efficacy has been directly related to health behavior [33]. Its effect can also be indirect through its impact on goals. Individuals with higher self-efficacy, ie, who are more confident in their ability, may challenge themselves more by setting higher goals and focus on opportunities rather than obstacles in carrying out the specific behavior [42]. Literature also suggests that those with higher nutrition-related self-efficacy are less likely to relapse to their previous unhealthy habits [43]. Therefore, changes in all these variables are considered promising for future vegetable consumption. It is interesting to note that we also found a decrease in individuals' reported intention to eat a greater variety of vegetables. These findings are consistent with other evaluations of app-delivered programs [44]. There are several possible reasons for this. For example, the VegEze app centered around a short challenge, and at the end, participants may have felt as though they had completed the program and no longer needed to keep improving and/or keep using the app and therefore reported a lower intention. They may have also felt successful in their behavior change endeavors and therefore no longer required as much motivation to continue performing the target behavior. Finally, motivation is critical for initiating new behaviors, but as the behavior becomes routine, different factors may be needed to continue to support performing the behavior.

Adoption—Who Was Most Likely to Use the App?

It was a purposeful decision to test this app in the real world; therefore, participants of this study used the app in representative settings. With regard to the representativeness of those who engaged with the program, women and those aged 51 years and above used the app more than men and younger adults. Despite recruiting participants through multiple channels—such as free-to-air television and radio coverage, social media, and an existing database—similar to the findings of previous health-related interventions, the majority of this study sample was women. Traditionally, mobile phone usage has been higher in younger adults, but more recently, the largest growth has been among older Australians [45]. Consistent with this, we found that, on average, the older participants used VegEze for an extra 2 days in the 21-day challenge period. The benefits of consuming more vegetables may also be perceived as more beneficial for older participants as they are seeking to improve their health. Engagement with technology is increasing across all age groups, and the appeal of apps as a delivery platform technology should be considered in health intervention regardless of the age profile of targeted users.

Implementation—App Usage Patterns and Association With Success

Baseline vegetable consumption and app usage were the 2 strongest predictors of increased vegetable intake and variety. Participants with the highest usage, ie, using the app almost every day (16-21 days; average 18 days), increased their consumption by twice as much as those with the lowest app usage. The cause and effect cannot be assumed; however, it is encouraging that higher engagement with the app was associated with a more positive outcome in the behavior of interest. The

fact that lower vegetable consumers reported greater increases in consumption was also promising. Although all Australians need to increase their vegetable intake to meet recommendations [6], arguably those with the lowest intakes have the greatest need for support and will benefit the most from intervention. It is not always the case that interventions work in those that need it most. For example, Mummah et al [15] reported that baseline consumption was a significant moderator of intervention effect whereby the impact of their app on vegetable intake was the greatest among those who reported higher intakes at baseline.

Maintenance—Is Behavior Change Sustained?

Long-term maintenance of behavior is important for realizing the health benefits associated with higher vegetable intake. Initial increases in vegetable intake were sustained up till a 3-month follow-up for which 87% of the sample reported of *always* or *usually* including 3 different types of vegetables at dinner, and one-quarter of the sample was meeting the dietary guidelines recommendations. Nationally, it is estimated that less than 5% of Australian adults have adequate vegetable intake [6], and public health and policy initiatives are seeking novel approaches to shift this persistent population trend. To determine whether this app has been truly effective in long-term behavior change, we would need to extend the follow-up to 6 months or more.

Other digital interventions have not been able to sustain changes in vegetable intake over 3 months [40]. We found that the increases in intake were maintained despite a reduction in engagement with the app. It is possible that users may reduce their reliance on the app once their goal has been achieved and a new way of eating has been formed. This is possible given the purposeful simplicity of the target behavior. Participants started the challenge with a high level of intention to increase their vegetable intake, and it is possible that as they progressed; they felt they were successful in their behavior change endeavors and therefore no longer required as much motivation from the app features, which may be why the intention to use the app fell. As a behavior becomes routine, different factors may be needed to continue maintenance [32]. Similarly, apps may need to evolve with the users' stage of change.

Limitations

There are limitations to this study that warrant discussion. The study design was chosen to maximize reach and replicate a real-world setting, but a cohort study evaluation is not as strong to show efficacy as other methods such as a randomized controlled trial, and therefore caution is required in the interpretation of results. The subsample of participants who completed the follow-up survey was less than half of those doing the baseline survey, and data imputation was not used to account for missing data. Therefore, the change in vegetable consumption reported here is likely to overstate the potential of this app on intake more generally. Australia is largely a mobile phone market dominated by Apple and Samsung, with 42% and 35% of the market share, respectively [45]. For pragmatic and financial reasons, we decided to initially test this app on the iOS only, which limited the potential sample. Despite making the app as accessible as possible through the App Store, we still recruited a largely female sample, and those who completed the evaluation surveys are likely to represent the more motivated, health-interested users. It remains a challenge to engage with other subgroups of the population who are also likely to benefit from technology-based nutrition interventions.

Maintaining interest in an app is difficult. Only 15% of the sample actively used the app for the entire 21 days of the challenge. Nonetheless, the attrition rate observed here was similar to what has been reported elsewhere [46]. Furthermore, to facilitate timely dissemination of results, the recruitment window and study time frames for this study were short. We do not know the optimal exposure time needed for an intervention message to achieve the greatest result, and in fact, this may differ depending on the behavior of interest. Although the follow-up data suggest that behavior change can be maintained, we would require additional time to demonstrate a more sustained impact on vegetable intake in a larger, more diverse sample.

Conclusions

Inadequate vegetable intake is a global problem, and the health impact of increased vegetable intake is well known, meaning, there is a real need for novel strategies and interventions that achieve successful increase in vegetable consumption. The VegEze app was designed as a 21-day challenge to increase vegetable consumption, and results indicate that focusing on a specific behavioral target around increasing variety was successful in increasing the amount of vegetables consumed. We were able to shift the distribution of intake in a large sample and increase the average consumption by half a serving. Utilizing Apple's ResearchKit allowed us to embed the evaluation process in the app and still place it in the App Store, contributing to the overall reach, generalizability, and rigorous evaluation of outcomes. Future research apps may also choose this approach to allow for more frequent evaluation of mHealth nutrition interventions in an uncontrolled real-world setting. Given that app usage was associated with successful behavior change, further work to improve engagement is warranted.

Acknowledgments

The authors would like to thank the SP Health technical team for their role in supporting the app development company in building the VegEze app databases.

This project was funded by Hort Innovation through the research and development levy. Hort Innovation had no involvement in the evaluation or interpretation of results or preparation of the manuscript.

Authors' Contributions

GH and AC led the development of the app. GM and GW contributed to the design and content of the app. GH, MH, and EB designed and conducted the analysis and wrote the results. GH, MH, and AC contributed to the presentation of results. GH wrote the initial draft of the manuscript. All authors have critically reviewed the iterations of the manuscript and approved it for publication. The team of authors listed were involved in the development of the app and responsible for planning and conducting the evaluation.

Conflicts of Interest

None declared.

References

1. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
2. World Health Organisation.: World Health Organisation Promoting Fruit and Vegetable Consumption Around the World URL: http://www.who.int/dietphysicalactivity/fruit/en/ [accessed 2017-10-17]
3. Produce for Better Health Foundation. Have a Plant. State of the Plate URL: https://fruitsandveggies.org/state-of-the-plate/ [accessed 2017-07-15]
4. Pollard CM, Miller MR, Daly AM, Crouchley KE, O'Donoghue KJ, Lang AJ, et al. Increasing fruit and vegetable consumption: success of the Western Australian Go for 2&5 campaign. Public Health Nutr 2008 Mar;11(3):314-320. [doi: 10.1017/S1368980007000523] [Medline: 17612423]
5. Woolcott Research, Australia. Department of Health and Ageing. Evaluation of the National Go for 2 & 5 Campaign. Canberra, ACT: Dept of Health and Ageing; 2007.
6. Australian Bureau of Statistics. 2016. 4364.0.55.012 - Australian Health Survey: Consumption of Food Groups from the Australian Dietary Guidelines, 2011-12 URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4364.0.55.0122011-12 [accessed 2020-02-11]
7. Rekhy R, McConchie R. Promoting consumption of fruit and vegetables for better health. Have campaigns delivered on the goals? Appetite 2014 Aug;79:113-123. [doi: 10.1016/j.appet.2014.04.012] [Medline: 24751917]
8. Lee-Kwan SH, Moore LV, Blanck HM, Harris DM, Galuska D. Disparities in state-specific adult fruit and vegetable consumption - United States, 2015. MMWR Morb Mortal Wkly Rep 2017 Nov 17;66(45):1241-1247 [FREE Full text] [doi: 10.15585/mmwr.mm6645a1] [Medline: 29145355]
9. NHS Digital. 2019 Nov 27. Health Survey for England 2017: Summary of Key Findings URL: https://files.digital.nhs.uk/4D/F4FF7D/HSE17-report-summary-v2.pdf [accessed 2019-11-30]
10. Nour M, Chen J, Allman-Farinelli M. Efficacy and external validity of electronic and mobile phone-based interventions promoting vegetable intake in young adults: Systematic review and meta-analysis. J Med Internet Res 2016 Apr 8;18(4):e58 [FREE Full text] [doi: 10.2196/jmir.5082] [Medline: 27059765]
11. Appleton KM, Hemingway A, Saulais L, Dinnella C, Monteleone E, Depezay L, et al. Increasing vegetable intakes: rationale and systematic review of published interventions. Eur J Nutr 2016 Apr;55(3):869-896 [FREE Full text] [doi: 10.1007/s00394-015-1130-8] [Medline: 26754302]
12. Glasson C, Chapman K, James E. Fruit and vegetables should be targeted separately in health promotion programmes: differences in consumption levels, barriers, knowledge and stages of readiness for change. Public Health Nutr 2011 Apr;14(4):694-701. [doi: 10.1017/S1368980010001643] [Medline: 20576194]
13. Knai C, Pomerleau J, Lock K, McKee M. Getting children to eat more fruit and vegetables: a systematic review. Prev Med 2006 Feb;42(2):85-95. [doi: 10.1016/j.ypmed.2005.11.012] [Medline: 16375956]
14. Serrano KJ, Yu M, Coa KI, Collins LM, Atienza AA. Mining health app data to find more and less successful weight loss subgroups. J Med Internet Res 2016 Jun 14;18(6):e154 [FREE Full text] [doi: 10.2196/jmir.5473] [Medline: 27301853]
15. Mummah S, Robinson TN, Mathur M, Farzinkhou S, Sutton S, Gardner CD. Effect of a mobile app intervention on vegetable consumption in overweight adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Sep 15;14(1):125 [FREE Full text] [doi: 10.1186/s12966-017-0563-2] [Medline: 28915825]
16. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
17. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
18. 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 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]

19. 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]
20. Kankanhalli A, Shin J, Oh H. Mobile-based interventions for dietary behavior change and health outcomes: scoping review. JMIR Mhealth Uhealth 2019 Jan 21;7(1):e11312 [FREE Full text] [doi: 10.2196/11312] [Medline: 30664461]
21. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, et al. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. J Med Internet Res 2013 Oct 4;15(10):e224 [FREE Full text] [doi: 10.2196/jmir.2745] [Medline: 24095951]
22. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
23. National Health & Medical Research Council, Department of Health and Aging. Eat for Health. Australian Dietary Guidelines: Summary. Australia: Commonwealth of Australia; 2013.
24. Meengs JS, Roe LS, Rolls BJ. Vegetable variety: an effective strategy to increase vegetable intake in adults. J Acad Nutr Diet 2012 Aug;112(8):1211-1215 [FREE Full text] [doi: 10.1016/j.jand.2012.05.013] [Medline: 22818729]
25. Bucher T, Siegrist M, van der Horst K. Vegetable variety: an effective strategy to increase vegetable choice in children. Public Health Nutr 2014 Jun;17(6):1232-1236. [doi: 10.1017/S1368980013002632] [Medline: 24107607]
26. Bucher T, van der Horst K, Siegrist M. Improvement of meal composition by vegetable variety. Public Health Nutr 2011 Aug;14(8):1357-1363. [doi: 10.1017/S136898001100067X] [Medline: 21557868]
27. Hendrie GA, James-Martin G, Williams G, Brindal E, Whyte B, Crook A. The development of VegEze: smartphone app to increase vegetable consumption in Australian adults. JMIR Form Res 2019 Mar 27;3(1):e10731 [FREE Full text] [doi: 10.2196/10731] [Medline: 30916653]
28. RE-AIM: Improving Public Health Relevance and Population Impact. URL: http://www.re-aim.org/ [accessed 2018-03-09]
29. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
30. Hendrie GA, Baird D, Golley RK, Noakes M. The CSIRO Healthy Diet Score: an online survey to estimate compliance with the Australian Dietary Guidelines. Nutrients 2017 Jan 9;9(1):pii: E47 [FREE Full text] [doi: 10.3390/nu9010047] [Medline: 28075355]
31. Hendrie GA, Rebuli MA, Golley RK. Reliability and relative validity of a diet index score for adults derived from a self-reported short food survey. Nutr Diet 2017 Jul;74(3):291-297. [doi: 10.1111/1747-0080.12303] [Medline: 28731597]
32. Schwarzer R. Health Action Process Approach (HAPA) as a theoretical framework to understand behavior change. Actual Psicol 2016;30(121):119-130. [doi: 10.15517/ap.v30i121.23458]
33. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
34. Schwarzer R, Renner B. Userpage. Health-Specific Self-Efficacy Scales URL: http://userpage.fu-berlin.de/health/healself.pdf [accessed 2016-11-16]
35. Luszczynska A, Tryburcy M, Schwarzer R. Improving fruit and vegetable consumption: a self-efficacy intervention compared with a combined self-efficacy and planning intervention. Health Educ Res 2007 Oct;22(5):630-638. [doi: 10.1093/her/cyl133] [Medline: 17060349]
36. van Assema P, Martens M, Ruiter RA, Brug J. Framing of nutrition education messages in persuading consumers of the advantages of a healthy diet. J Hum Nutr Diet 2001 Dec;14(6):435-442. [doi: 10.1046/j.1365-277x.2001.00315.x] [Medline: 11906585]
37. Teal CR, Haidet P, Balasubramanyam AS, Rodriguez E, Naik AD. Measuring the quality of patients' goals and action plans: development and validation of a novel tool. BMC Med Inform Decis Mak 2012 Dec 27;12:152 [FREE Full text] [doi: 10.1186/1472-6947-12-152] [Medline: 23270422]
38. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
39. Australian Bureau of Statistics. Canberra, Australia 2016 Census URL: https://www.abs.gov.au/websitedbs/censushome.nsf/home/2016 [accessed 2020-02-11]
40. Nakamura S, Inayama T, Harada K, Arao T. Reduction in vegetable intake disparities with a web-based nutrition education intervention among lower-income adults in Japan: randomized controlled trial. J Med Internet Res 2017 Nov 24;19(11):e377 [FREE Full text] [doi: 10.2196/jmir.8031] [Medline: 29175810]
41. Bailey RR. Goal setting and action planning for health behavior change. Am J Lifestyle Med 2019;13(6):615-618 [FREE Full text] [doi: 10.1177/1559827617729634] [Medline: 31662729]
42. Schwarzer R, Luszczynska A. Division of Cancer Control and Population Sciences. Perceived Self-Efficacy URL: https://cancercontrol.cancer.gov/brp/research/constructs/self-efficacy.pdf [accessed 2017-05-12]
43. Bagozzi RP, Edwards EA. Goal setting and goal pursuit in the regulation of body weight. Psychol Health 1998;13(4):593-621. [doi: 10.1080/08870449808407421]

44. Brindal E, Hendrie GA, Freyne J, Noakes M. A mobile phone app designed to support weight loss maintenance and well-being (MotiMate): Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Sep 4;7(9):e12882 [FREE Full text] [doi: 10.2196/12882] [Medline: 31486407]
45. Deloitte. Mobile Consumer Survey 2019 URL: https://www2.deloitte.com/au/mobile-consumer-survey [accessed 2019-10-15]
46. van der Mispel C, Poppe L, Crombez G, Verloigne M, de Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]

Abbreviations

CSIRO: Commonwealth Scientific & Industrial Research Organisation
iOS: iPhone Operating System
HIAL: Horticulture Innovation Australia Ltd
mHealth: mobile health
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

Edited by G Eysenbach; submitted 19.05.19; peer-reviewed by L Poppe, M Gaman; comments to author 12.07.19; revised version received 16.09.19; accepted 12.11.19; published 17.04.20.

Please cite as:
Hendrie GA, Hussain MS, Brindal E, James-Martin G, Williams G, Crook A
Impact of a Mobile Phone App to Increase Vegetable Consumption and Variety in Adults: Large-Scale Community Cohort Study
JMIR Mhealth Uhealth 2020;8(4):e14726
URL: http://mhealth.jmir.org/2020/4/e14726/
doi:10.2196/14726
PMID:32301739

Original Paper

Temporal Dynamics of Treatment Receipt in a Text Message Intervention for Physical Activity: Single-Group, Within-Person Trial

David E Conroy[1], PhD; Chih-Hsiang Yang[2], PhD; Stephanie T Lanza[3], PhD; Joshua M Smyth[4], PhD; Constantino M Lagoa[5], PhD

[1]Department of Kinesiology, The Pennsylvania State University, University Park, PA, United States

[2]Department of Exercise Science, University of South Carolina, Columbia, SC, United States

[3]The Edna Bennett Pierce Prevention Research Center, Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, United States

[4]Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, United States

[5]Department of Electrical Engineering & Computer Science, The Pennsylvania State University, University Park, PA, United States

Corresponding Author:
David E Conroy, PhD
Department of Kinesiology
The Pennsylvania State University
266 Rec Hall
University Park, PA, 16802
United States
Phone: 1 814 863 3451
Fax: 1 814 865 1275
Email: conroy@psu.edu

Abstract

Background: Mobile technology has increased the reach of health behavior interventions but raised new challenges in assessing the fidelity of treatment receipt. Fidelity can be compromised if participant fatigue or burden reduces engagement, leading to missed or delayed treatments for just-in-time interventions.

Objective: This study aimed to investigate the temporal dynamics of text message receipt confirmations.

Methods: Community-dwelling adults (N=10) were sent five text messages daily for 4 months (5598 messages sent in total), with a financial incentive to confirm receipt of 75% or more messages.

Results: Overall, the message receipt confirmation rate was very high (5504/5598, 98.32%) and timely (eg, two-thirds of confirmations within 2 min). Confirmation times were slightly slower on weekends (vs weekdays) and as a function of the cumulative time in the study. Neither time of message delivery nor message content was associated with message confirmation latencies.

Conclusions: Participants receiving financial incentives to confirm text message receipt exhibit extremely high and fast confirmation rates, although receipt confirmations were somewhat less timely on weekends (vs weekdays) and later in the intervention. The social calendar and treatment fatigue should be considered when planning text message–based interventions, especially if treatments are intended for a just-in-time delivery that requires extended engagement and precise timing.

(JMIR Mhealth Uhealth 2020;8(4):e14270) doi:10.2196/14270

KEYWORDS

short message service; patient engagement; mHealth; physical activity; sedentary behavior

Introduction

Background

Digital tools, such as mobile phones, have transformed the form and reach of many health behavior interventions. It is now possible to deliver intensive treatments to people in the natural context of their daily lives [1]. Mobile health (mHealth) interventions have shown considerable promise as being efficacious for modifying a number of preventive health behaviors [2]. Text messages (ie, SMS) have emerged as a popular mode of mHealth intervention delivery because over 90% of American adults own a cell phone and over 80% use their phones for text messaging [3,4]. SMS intervention effects have been quite variable across a range of health behaviors and, more specifically, have not produced consistent changes in physical activity [2,5,6]. One possible source of that variation involves the fidelity of treatment receipt: If people do not receive or read messages in a timely fashion, their full effects are unlikely to be realized, particularly for just-in-time interventions that are sensitive to synchronizing the timing of message receipt with moments of opportunity or vulnerability. Little is known about the temporal dynamics of treatment receipt in SMS interventions generally or for just-in-time messaging interventions specifically. This knowledge gap presents a barrier to optimally engaging patients in just-in-time messaging interventions. This study documented the rate of message confirmations within a variety of time windows and estimated associations between time-varying (within-person) factors and an important indicator of timely treatment receipt, namely, the latency of message receipt confirmations.

Treatment Fidelity

The National Institutes of Health (NIH) Behavioral Change Consortium defined the scope of fidelity as spanning study design; staff training; and intervention delivery, receipt, and enactment [7,8]. High fidelity is essential for drawing causal inferences about intended intervention effects. One prominent threat to mHealth treatment fidelity is insufficient user engagement with the behavior change intervention or its techniques [9]. A recent study of mobile apps for mental health indicated that users' daily rate of opening the apps dropped by 80% in the first 10 days after initially installing and opening the app [10]. This decline in engagement reduces mHealth treatment fidelity but is somewhat unavoidable when users are required to initiate their interactions with intervention content. One way to reduce that friction and boost engagement is to push intervention content to users using SMS messages or notifications. With just-in-time or ecological momentary intervention approaches, the intervention can be conceptualized as a series of individual intervention doses (eg, SMS messages) delivered over time [1,11]. Intervention doses that are not delivered, received, or enacted will introduce discrepancies between intended and enacted treatment doses and undermine treatment fidelity. If the individual intervention doses require precise timing to coincide with moments of opportunity or vulnerability, it is essential for those messages to be received when delivered [12]. Doses that are not received as intended are unlikely to change behavioral target actions or clinical outcomes [13]. Moreover, when mHealth interventions fail, a lack of treatment fidelity makes it difficult to understand whether that failure was because of failures in study design (eg, inert intervention content and insufficient dosing), failures in treatment delivery, failures in treatment receipt, or failures to enact prescribed actions. Yet granular data on treatment receipt from just-in-time interventions have not been widely reported to date, and little is known about the factors that influence those dynamics.

It can be challenging to verify whether participants read SMS messages as they are received (or at all). Simply verifying that a message was opened in a timely fashion would eliminate the need to assume that messages are always received as sent. Such data would help to rule out the possibility that messages were accumulating undelivered or that long delays existed between message transmission and receipt. If participants were asked to confirm message receipt with a response, those responses could signal that messages were at least delivered, opened, and inspected enough to determine whether a confirmation was needed (ie, discriminating between intervention-related messages and personal messages). Such message confirmations could indicate that some level of treatment was received (although they do not provide direct evidence that the message was understood or enacted). Thus, overall confirmation response rates and latencies can serve as approximate indicators of the fidelity of text message treatment receipt.

Treatment Fidelity as a Dynamic Process

Several dynamic factors may influence the latency of text message receipt confirmations, including treatment fatigue and timing. Treatment fatigue describes "psychological fatigue associated with treatment engagement" [14]. The effort required to open and read a single text message is of course minimal, but this activity requires one to interrupt other ongoing activities and shift attention so cumulative effort may be substantial (ie, across multiple messages). Unless messages are highly rewarding (ie, interesting or enjoyable), intervention burden can accumulate to create fatigue and may reduce adherence to reading messages in a timely fashion; eg, self-reports of cessation fatigue during smokers' quit attempts have been associated with reduced success [15]. We are not aware of any studies on behavioral indicators of treatment fatigue experienced during intensive SMS interventions for physical activity or for extended periods of time. This study examined changes in overall response confirmation rates across a long-term SMS intervention when participants are vulnerable to treatment fatigue and related processes that would undermine the fidelity of treatment receipt (eg, habituation).

In an intensive SMS intervention, the timing of message distribution may also vary on daily and hourly time scales. For example, differences in discretionary time availability because of work or personal obligations may produce weekday-weekend differences in the dynamics of treatment receipt. Similarly, the time of day when messages are sent may impact how quickly people can read and respond to confirm receipt. These hypotheses about differences in response confirmation times by day-of-week or time of message delivery were treated as exploratory.

Objective

This study was designed to investigate the dynamics of treatment receipt in an intensive and long-term SMS intervention to promote physical activity. For 4 months, participants received five SMS messages daily, each randomly drawn from large pools of messages that focused on (1) social-cognitive processes associated with physical activity (*move more*, 101 messages), (2) social-cognitive processes associated with limiting sedentary behavior (*sit less*, 101 messages), or (3) non–health-related general facts (trivia, 254 messages). We hypothesized that the latency of SMS receipt confirmations would increase as a function of the number of days since the beginning of the intervention. We also hypothesized that receipt confirmation latency would vary as a function of the day of the week and the time of day when messages were delivered, but we made no hypothesis about the direction of these associations. Finally, we evaluated whether message content (*move more* and *sit less*) influenced message confirmations relative to the non–health-related general fact messages.

Methods

Participants and Procedures

Community advertisements were used to recruit participants for a study to evaluate a 16-week SMS intervention to promote physical activity. Participants (9/10, 90% female) ranged from 22 to 47 years in age with a mean of 34.3 (SD=8.99) years. This sample was mostly white (9/10, 90%) and not Hispanic or Latino (10/10, 100%) with full-time employment (8/10, 80%). Participants varied in their marital status (5/10, 50% single; 4/10, 40% married, and 1/10, 10% divorced) and status as parents (6/10, 60% had children). Maximum educational attainments ranged from a high school diploma to a doctoral degree, but most of the sample (60%) had not completed a Bachelor's degree.

For the 16 weeks of the study, participants received five text messages daily. A static *Do Not Disturb* window was fixed for all participants between 8:00 PM and 8:00 AM. Messages were delivered using an online service on a semirandom schedule within equal-sized segments under the constraint that consecutive messages were separated by at least 60 min [16]. For all messages, participants were instructed to confirm receipt by replying to each message as quickly as possible after reading it. Responses could be as simple as a single letter to minimize burden. As an incentive, participants received a weekly payment (US $15/week) for responding to 75% of the text messages each week. Participants who responded to at least 75% of the messages within 2 min during a month entered a drawing for a US $100 bonus. Details about the study that are not relevant to this report (eg, activity monitoring) have been reported elsewhere [17].

Measures

Response confirmation latencies were measured in seconds (to the second decimal place) from the time the intervention SMS was sent until the corresponding SMS confirmation was received. Confirmations received in less than 1 second (n=94; <1%) were deemed implausible and attributed to technical errors, and so they were recoded as missing values.

Data Analysis

Response latencies were positively skewed. A natural log transformation was implemented to normalize the distribution of scores. Multilevel models were estimated to model the duration of confirmation latencies [18]. A dummy variable was created to represent weekend days (weekdays were set as the reference category). Time of day was coded as minutes since 3 AM because of the observed SMS inactivity in the early morning hours. A pair of dummy variables were created to represent message content related to increasing physical activity or decreasing sedentary behavior (general fact messages were set as the reference category).

Results

A total of 5598 SMS messages were sent and 5504 confirmations (5504/5598, 98.3%) were received. Confirmation responses were received very quickly overall, with most within 33 seconds (2690/5504, 49.87%), 2 min (3676/5504. 66.79%), or 5 min (4108/5504, 74.64%; percentages are cumulative). The distribution of response times had a heavy tail that was normalized after implementing a natural logarithm transformation. The intraclass correlation for the transformed score was 0.08. Given the minimal proportion of confirmation time variance that existed between people and our limited sample size, the multilevel model regressed transformed confirmation latencies on the hypothesized within-person predictors.

As shown in Table 1, inferred treatment fatigue (as indexed by day in study) was associated with a small, but statistically significant, increase in the transformed latency of response confirmations. Saturdays and Sundays were also associated with significantly longer transformed response confirmation latencies than weekdays (the untransformed equivalent of approximately 3.8 seconds longer for weekend response confirmations), but these differences were relatively small in absolute terms. Neither message delivery time nor message content was associated with transformed response confirmation latencies.

Table 1. Multilevel regressions of message confirmation latencies.

Variable	Beta	SE	*P* value
Treatment fatigue (study day)	.003[a]	0.001	.003
Message delivery time of day	.00	0.00	.62
Weekends (vs weekdays)	.40[a]	0.10	<.001
Physical activity (vs general) content	−.007	0.03	.83
Sedentary behavior (vs general) content	.07	0.06	.26
Residual variance (within person)	3.09[a]	0.25	<.001
Mean (between person)	−.09	0.20	.67
Variance (between person)	.28[a]	0.09	.002

[a]*P*<.01.

Discussion

The objective of this brief report was to examine temporal correlates of the inferred fidelity of treatment receipt in an intensive SMS intervention. Results indicated that (1) there was a very high and sustained rate of message confirmations, likely because of (in full or in part) the incentive program, and (2) the social calendar and treatment fatigue, but not message timing or content, were associated with small but reliable latency differences in participant confirmations via SMS.

Principal Findings

The very high message confirmation rate (5504/5598, 98.3%) was impressive given the duration (4 months) and intensity (five times daily) of the SMS intervention. Treatment fatigue was evidenced by the small, but statistically significant, increased delays before message confirmations were received as cumulative exposure to the intervention increased. This finding provides evidence of a potential barrier to treatment fidelity in SMS-based interventions requiring high temporal precision, particularly those with lengthy or intensive SMS schedules that require extended engagement, although we note that overall response rates were outstanding and the delays were of very small magnitude in the majority of cases. It should be noted that it is also not clear whether the delays were due to the burden of mentally processing message content, the workflow for responding to individual messages, or other factors (eg, being out of a service area or in airplane mode). Regardless of the source, a lack of fidelity in terms of treatment receipt can reduce confidence in inferences about intended treatment effects, particularly if fidelity was not evaluated and accounted for when estimating effects [8]. Research on chronic disease management also has shown that other indicators of treatment fatigue adversely impact adherence [19,20]. The relatively weak treatment fatigue observed in this study raises questions about whether such fatigue could be sufficient to reduce treatment enactment or modify behavior change. This issue should be monitored when implementing future SMS interventions to provide an empirical basis for answering this question.

As expected, the social calendar was associated with differences in the fidelity of treatment receipt: message confirmations were, on average, 3.8 seconds slower on weekends than weekdays, possibly because of reduced availability. Prior work has defined availability as a state in which a person "is capable of engaging in an incoming, unplanned activity" (p. 913), and—based on multiple wearable sensors used to infer availability—demonstrated greater availability on weekends than weekdays [21]. Taken together, these findings suggest that participants may be available for—but slightly less motivated to receive—interventions on weekends than on weekdays. However, the magnitude of differences observed in this study (and in the context of a financial reward for compliance) does not appear to be sufficient to represent a substantial threat to the fidelity of intervention receipt. Thus, this conclusion may well be limited to approaches that incentivized responses and may differ substantially when participants are not incentivized. It also may not generalize to people with different social calendars (eg, shift workers and stay-at-home parents). Future research can benefit from mixed method approaches to identify reasons for differential availability or motivation that can explain these within-person associations.

Neither message delivery time nor message content was associated with the fidelity of treatment receipt. Incentivized SMS confirmation receipts may be resistant to self-regulatory fluctuations because of changes in desire or impaired affective and cognitive functioning as the day progresses [22]. This study compared three types of SMS messages across a fixed 12-hour period of the day. To strengthen conclusions about message properties, future work should investigate the dynamics of treatment receipt for a more diverse array of health behaviors and across a wider time interval.

Limitations

This study provided insights into how the fidelity of SMS treatment receipt can vary over time but also had several limitations. First, the design was not experimental, so causal inferences cannot be drawn based on these findings. Second, the sample size was small and homogenous, so results may not generalize to more representative populations, populations with different motivational profiles, or, as mentioned earlier, individuals who have different social calendars. Owing to the limited sample size, a two-level model was estimated (repeated measures of message confirmations within person). Future work with larger samples should consider a three-level model to

separate the day-in-study variance from the time-of-day variance. Third, the financial incentive to confirm message receipt and the overall SMS confirmation rate were high. These findings may not generalize to contexts without financial incentives (although it is worth noting that, even with these incentives, treatment receipt was still weakly associated with treatment fatigue and day of week). It is possible that different patterns exist when participants are not incentivized to confirm message receipt or if incentives have different contingencies (eg, targeting latency or frequency alone). It seems likely that overall confirmation rates would likely be lower under either of those conditions, but research is needed to test that hypothesis. Fourth, participants' availability was not screened to determine whether they could receive messages (eg, Is the person sleeping? Is the person driving? Is the person exercising already?), but high response rates were obtained using static timeframes for message delivery. Finally, message content was drawn from a scripted message bank with limited repetition; it is possible that fatigue may be a greater threat in contexts with less novelty and greater message repetition.

Conclusions

We conclude that sustained high treatment fidelity, even in contexts with high-density responding (5 times/day) for sustained durations (4 months), is possible—at least under circumstances where there are financial incentives for confirming SMS treatment receipt. Although overall rates were very high, we also found some evidence that treatment fidelity remains is a dynamic process. Depending on the pattern of message receipt, treatment may be slightly biased toward certain situations. Specifically, SMS-based treatment receipt was slightly faster on weekdays (compared with weekends), and slowly decayed as the intervention continued over a 4-month period. Failures to receive interventions as intended may be one factor that helps explain the lack of consistent effects of mHealth physical activity interventions and variation in SMS-based effects for other health behaviors [2,5], although we again note that overall response in this context was exceptionally high. As such, the dynamics of treatment receipt should be monitored because this process is not static. These preliminary findings can help to inform best practices for using SMS (or a similar technology such as smartphone notifications) to deliver just-in-time interventions that depend on people receiving interventions at specific moments of vulnerability or opportunity.

Acknowledgments

The research reported in this publication was supported in part by the Penn State Social Science Research Institute and the National Heart, Lung, and Blood Institute of the NIH under award number R01HL142732. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

DC, SL, JS, and CL designed the study. DC and CY were involved with data acquisition. DC conducted the analysis. Interpretation of data was performed by DC, SL, JS, and CL. All authors were involved in interpreting the results and in reviewing and revising the manuscript.

Conflicts of Interest

None declared.

References

1. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
2. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: A meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
3. Anderson M. Pew Research Center. Washington, DC: Pew Research Center; 2015 Oct 29. The Demographics of Device Ownership URL: https://www.pewresearch.org/internet/2015/10/29/the-demographics-of-device-ownership/ [accessed 2016-01-15]
4. Duggan M. Pew Research Center. 2012 Nov 25. Cell Phone Activities 2012 URL: http://www.pewinternet.org/2012/11/25/cell-phone-activities-2012/ [accessed 2017-03-11]
5. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: Behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
6. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
7. 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]
8. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(s1):S52-S63 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]
9. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
10. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
11. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
12. Smyth JM, Heron KE. Is Providing Mobile Interventions 'Just-in-time' Helpful? An Experimental Proof of Concept Study of Just-in-time Intervention for Stress Management. In: Proceedings of the 2016 IEEE Wireless Health Conference. 2016 Presented at: WH'16; October 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764561]
13. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019 Dec;38(12):1168-1174. [doi: 10.1037/hea0000805] [Medline: 31580127]
14. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 1;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
15. Liu X, Li R, Lanza ST, Vasilenko SA, Piper M. Understanding the role of cessation fatigue in the smoking cessation process. Drug Alcohol Depend 2013 Dec 1;133(2):548-555 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.07.025] [Medline: 23954071]
16. SurveySignal. URL: https://www.surveysignal.com/ [accessed 2020-02-19]
17. Ashour M, Bekiroglu K, Yang C, Lagoa C, Conroy D, Smyth J, et al. On the Mathematical Modeling of the Effect of Treatment on Human Physical Activity. In: Proceedings of the 2016 IEEE Conference on Control Applications. 2016 Presented at: CCA'16; September 19-22, 2016; Buenos Aires, Argentina. [doi: 10.1109/cca.2016.7587951]
18. Snijders TA, Bosker RJ. Multilevel Analysis: An Introduction To Basic And Advanced Multilevel Modeling. Thousand Oaks, CA: Sage; 1999.
19. Claborn KR, Meier E, Miller MB, Leffingwell TR. A systematic review of treatment fatigue among HIV-infected patients prescribed antiretroviral therapy. Psychol Health Med 2015;20(3):255-265 [FREE Full text] [doi: 10.1080/13548506.2014.945601] [Medline: 25110152]
20. Pyatak EA, Florindez D, Weigensberg MJ. Adherence decision making in the everyday lives of emerging adults with type 1 diabetes. Patient Prefer Adherence 2013;7:709-718 [FREE Full text] [doi: 10.2147/PPA.S47577] [Medline: 23935361]
21. Sarker H, Sharmin M, Ali AA, Rahman MM, Bari R, Hossain SM, et al. Assessing the availability of users to engage in just-in-time intervention in the natural environment. Proc ACM Int Conf Ubiquitous Comput 2014;2014:909-920 [FREE Full text] [doi: 10.1145/2632048.2636082] [Medline: 25798455]
22. Millar BM. Clocking self-regulation: why time of day matters for health psychology. Health Psychol Rev 2017 Dec;11(4):345-357. [doi: 10.1080/17437199.2017.1316673] [Medline: 28385063]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 04.04.19; peer-reviewed by B Loo Gee, A Pfammatter, K Wolin; comments to author 01.10.19; revised version received 09.11.19; accepted 04.02.20; published 22.04.20.

Please cite as:
Conroy DE, Yang CH, Lanza ST, Smyth JM, Lagoa CM
Temporal Dynamics of Treatment Receipt in a Text Message Intervention for Physical Activity: Single-Group, Within-Person Trial
JMIR Mhealth Uhealth 2020;8(4):e14270
URL: http://mhealth.jmir.org/2020/4/e14270/
doi:10.2196/14270
PMID:32319957

Original Paper

Use of a Smartphone-Based Mobile App for Weight Management in Obese Minority Stroke Survivors: Pilot Randomized Controlled Trial With Open Blinded End Point

Nneka L Ifejika[1,2,3], MPH, MD; Minal Bhadane[4], PhD; Chunyan C Cai[5], PhD; Elizabeth A Noser[6,7], MD; James C Grotta[8], MD; Sean I Savitz[6,7], MD

[1]Department of Physical Medicine and Rehabilitation, University of Texas Southwestern Medical Center, Dallas, TX, United States

[2]Department of Neurology and Neurotherapeutics, University of Texas Southwestern Medical Center, Dallas, TX, United States

[3]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States

[4]College of Optometry, University of Houston, Houston, TX, United States

[5]Division of Clinical and Translational Sciences, Department of Internal Medicine, McGovern Medical School at the University of Texas Health Science Center at Houston, Houston, TX, United States

[6]Institute for Stroke and Cerebrovascular Disease, University of Texas Health Science Center at Houston, Houston, TX, United States

[7]Department of Neurology, McGovern Medical School at the University of Texas Health Science Center at Houston, Houston, TX, United States

[8]Clinical Innovation and Research Institute, Memorial Hermann Hospital – Texas Medical Center, Houston, TX, United States

Corresponding Author:
Nneka L Ifejika, MPH, MD
Department of Physical Medicine and Rehabilitation
University of Texas Southwestern Medical Center
5323 Harry Hines Blvd
Dallas, TX
United States
Phone: 1 2146488762
Email: nneka.ifejika@utsouthwestern.edu

Abstract

Background: Minorities have an increased incidence of early-onset, obesity-related cerebrovascular disease. Unfortunately, effective weight management in this vulnerable population has significant barriers.

Objective: Our objective was to determine the feasibility and preliminary treatment effects of a smartphone-based weight loss intervention versus food journals to monitor dietary patterns in minority stroke patients.

Methods: Swipe out Stroke was a pilot prospective randomized controlled trial with open blinded end point. Minority stroke patients and their caregivers were screened for participation using cluster enrollment. We used adaptive randomization for assignment to a behavior intervention with (1) smartphone-based self-monitoring or (2) food journal self-monitoring. The smartphone group used Lose it! to record meals and communicate with us. Reminder messages (first 30 days), weekly summaries plus reminder messages on missed days (days 31-90), and weekly summaries only (days 91-180) were sent via push notifications. The food journal group used paper diaries. Both groups received 4 in-person visits (baseline and 30, 90, and 180 days), culturally competent counseling, and educational materials. The primary outcome was reduced total body weight.

Results: We enrolled 36 stroke patients (n=23, 64% African American; n=13, 36% Hispanic), 17 in the smartphone group, and 19 in the food journal group. Mean age was 54 (SD 9) years; mean body mass index was 35.7 (SD 5.7) kg/m2; education, employment status, and family history of stroke or obesity did not differ between the groups. Baseline rates of depression (Patient Health Questionnaire-9 [PHQ-9] score median 5.5, IQR 3.0-9.5), cognitive impairment (Montreal Cognitive Assessment score median 23.5, IQR 21-26), and inability to ambulate (5/36, 14% with modified Rankin Scale score 3) were similar. In total, 25 (69%) stroke survivors completed Swipe out Stroke (13/17 in the smartphone group, 12/19 in the food journal group); 1 participant in the smartphone group died. Median weight change at 180 days was 5.7 lb (IQR –2.4 to 8.0) in the smartphone group versus 6.4 lb (IQR –2.2 to 12.5; *P*=.77) in the food journal group. Depression was significantly lower at 30 days in the smartphone group than in the food journal group (PHQ-9 score 2 vs 8; *P*=.03). Clinically relevant depression rates remained in the zero to minimal range for the smartphone group compared with mild to moderate range in the food journal group at day 90 (PHQ-9 score 3.5 vs 4.5; *P*=.39) and day 180 (PHQ-9 score 3 vs 6; *P*=.12).

Conclusions: In a population of obese minority stroke survivors, the use of a smartphone did not lead to a significant difference in weight change compared with keeping a food journal. The presence of baseline depression (19/36, 53%) was a confounding variable, which improved with app engagement. Future studies that include treatment of poststroke depression may positively influence intervention efficacy.

Trial Registration: ClinicalTrials.gov NCT02531074; https://www.clinicaltrials.gov/ct2/show/NCT02531074

(JMIR Mhealth Uhealth 2020;8(4):e17816) doi:10.2196/17816

KEYWORDS

smartphone; stroke; obesity; telemedicine; minority groups; cognitive dysfunction; outcome assessment, health care

Introduction

Importance of the Problem

Decades of research have shown a disproportionate incidence of obesity among African American and Hispanic populations in the United States [1-3]. Minorities have a median age at first stroke, which is highly correlated with obesity, 10 to 13 years earlier than non-Hispanic whites [4,5]. The early onset of stroke in minorities is expensive. Direct medical costs for African American stroke patients are estimated to top US $16 billion in 2020; by 2030, stroke prevalence is expected to rise the most among Hispanic men, with direct costs of care increasing over 300% since 2012 [6].

Adherence to evidence-based therapies for obesity management in minority stroke patients is insufficient. Previous studies have demonstrated success through restrictive health promotion interventions [7] and stroke navigators [8], both of which require close follow-up by a medical professional. Unfortunately, inadequate finances and lack of health insurance are barriers to care in minorities [9]; therefore, study protocols that require frequent medical follow-up are not generalizable.

Prior Work

The efficacy of smartphone-based self-monitoring to facilitate weight loss is well established. Wang et al [10] conducted a randomized controlled trial in a largely African American population, comparing a smartphone-based behavioral intervention for weight loss and glycemic control versus paper diaries or usual care. The smartphone group experienced a 2.73% weight loss at 6 months, compared with 0.13% weight loss in the paper diary group and a 0.49% weight gain in the usual-care group [10]. Minority populations are willing adopters of mobile health (mHealth) technology; a recent analysis by Asan et al of the Health Information National Trends Survey of American adults found that minorities were less likely to indicate they had no interest in exchanging health tips electronically with a health care provider [11].

However, there may be an attenuated effect of electronic device use in patients with chronic diseases, such as stroke. A study by Robbins et al showed that those with self-reported "very good or excellent" health or those who engaged in physical activity were more likely to download a health app [12]. Furthermore, in mHealth studies involving high-need, high-cost populations, Singh et al found that only 30.3% of the apps were identifiable and available to the public [13]. As stroke is a chronic end-organ disease with a high incidence of physical impairments, stroke survivors may have lower use of electronic interventions.

Objectives

The Swipe out Stroke study tested the feasibility and preliminary treatment effects of using a readily available smartphone-based weight loss intervention in obese minority stroke patients versus food journals. Stroke affects the cognitive domains of attention, memory, language, and orientation [14]; approximately 30% of stroke patients develop dementia within 1 year [15]. These cognitive challenges may lead to limited mHealth research in the minority stroke population, which is increasing in volume and health care costs. We hypothesized that personal engagement with health care professionals through the mobile app would provide positive reinforcement and give support when participants experienced difficulty with program maintenance.

Methods

Study Design

Swipe out Stroke was a phase 1, pilot, prospective, randomized controlled trial with open blinded end point study. The blinded end point committee consisted of the senior authors (JG, SS) and the statistician (CC). We evaluated 2 groups—(1) a group receiving a behavior intervention with smartphone-based self-monitoring and (2) a group receiving a behavior intervention with food journal self-monitoring—for differences in weight loss at day 30, day 90, and day 180 of the study period. We used a mixed-methods design with quantitative measures. We distributed exit surveys to determine participants' acceptance of the intervention; we report the quantitative results in this paper.

Ethics

This study was approved by the Institutional Review Board at the McGovern Medical School at the University of Texas Health Science Center at Houston, Houston, Texas, USA (HSC-MS-13-0608). We obtained informed consent from both study participants and their caregivers, detailing the purpose of the study, frequency of visits, descriptors of the smartphone and food journal interventions, benefits, risks, protection of privacy, and study withdrawal procedures. The Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) checklist served as a guide for reporting this study (Multimedia Appendix 1 [16]). The trial is registered with ClinicalTrials.gov (NCT02531074).

Patient Population and Setting

We screened obese African American or Hispanic patients age 18 years and older, acutely hospitalized for ischemic or hemorrhagic stroke. Textbox 1 outlines the inclusion and exclusion criteria. The setting was a Joint Commission–accredited Comprehensive Stroke Center in Houston, Texas, USA with a 322-km telemedicine radius, reaching areas with a large underinsured population. The Joint Commission is a US nonprofit organization that accredits more than 22,000 US health care organizations and programs. Joint Commission accreditation is recognized by the majority of US state governments as a condition of licensure for receiving Medicare and Medicaid reimbursements.

Textbox 1. Swipe out Stroke inclusion and exclusion criteria.

Inclusion criteria

- Ischemic or hemorrhagic stroke.
- Age 18-85 years.
- African American or Hispanic ethnicity.
- Poststroke modified Rankin Scale (mRS) score 0-3.
- Poststroke body mass index >30 kg/m^2.
- Prescription medication for diabetes mellitus, hypertension or hyperlipidemia.
- Willing to follow a healthy eating pattern and NOT use weight loss medications.
- Personal or caregiver ownership of a computer, smartphone or other smart device (iPhone or Android platform) with internet access.
- If patient has alexia, agraphia, acalculia, dementia or blindness, caregiver must be willing to complete the intervention.

Exclusion criteria

- Preexisting disability with mRS score ≥4.
- Contraindications to weight loss (planning to become pregnant, history of an eating disorder).
- Steroid use for suspected vasculitis.
- Current or recent (past 6 months) participation in a weight loss program or use of weight loss medication.

Sample Size Estimate

We aimed to recruit 50 patients in Swipe out Stroke, with retention of 15 patients in each group (n=30). Based on clinical estimates, we anticipated a 50% completion rate in the smartphone group and a 10% completion rate in the food journal group. Due to several large regional flooding events during the study period (March 2015 to December 2016), we closed enrollment upon achieving 17 patients in the smartphone group and 19 patients in the food journal group (n=36). Completion rates in both groups were higher than clinical estimates.

Randomization

We randomly assigned stroke survivors to either the smartphone group or food journal group in a 1 to 1 ratio, using an adaptive covariate randomization algorithm. The adaptive randomization schema used the Pearson chi-square statistic to measure treatment imbalances in stroke severity (modified Rankin Scale score of 0-1: no symptoms to no significant disability; or modified Rankin Scale score of 2-3: slight to moderate disability), age, and sex. We determined the presence of depression at randomization using the Patient Health Questionnaire-9 (PHQ-9) survey; we equally allocated participants with depression to the smartphone and food journal groups in 3 categories: zero to minimal (score 0-4), mild to moderate (5-14), and moderately severe to severe (>14). We recalculated treatment assignment probability for each new patient; random assignment using a Web-based management app achieved the best balance between the smartphone and food journal groups. Caregivers joined the group of the study participant to facilitate participation in the protocol and compliance with the study intervention.

Intervention

Screening

Participants were screened by the stroke social worker and a member of the Swipe out Stroke research team during the acute hospitalization. After informed consent was obtained, the baseline clinic visit occurred within 2 weeks of hospital discharge.

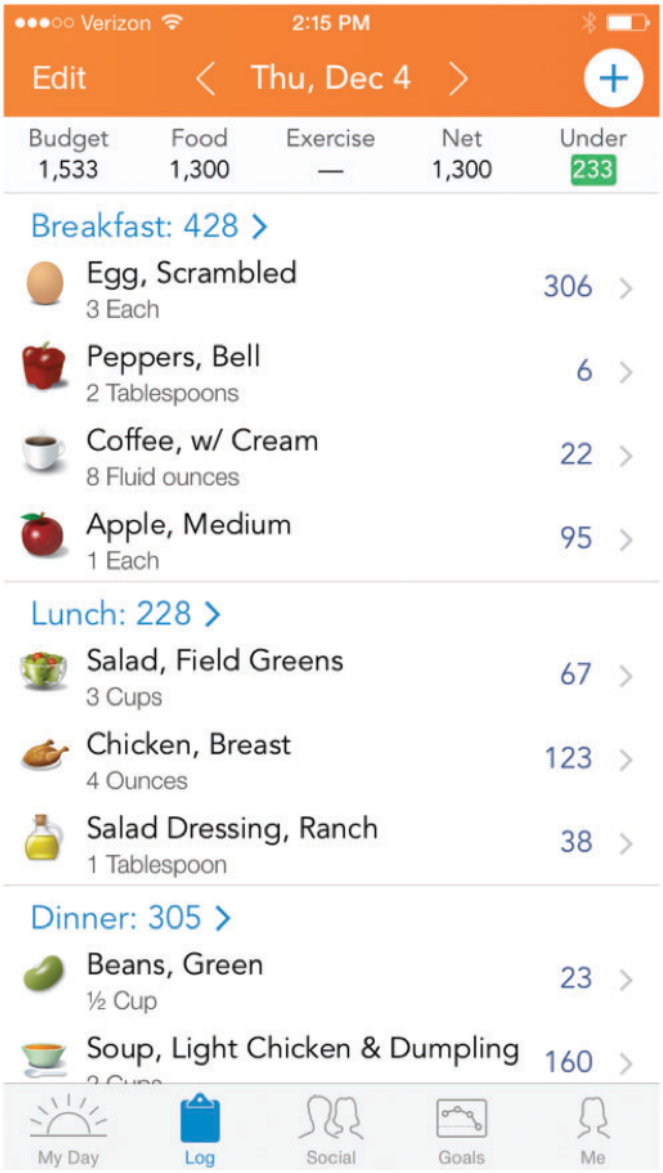
Smartphone Group

During the baseline visit, a member of the study team created a user account for the mobile app (Lose It! version 5.2.1; FitNow, Inc, Boston, MA, USA) by using the participant's email address. Each participant and caregiver (if applicable) received a tutorial on how to download and use Lose It! on his or her personally owned smartphone. Lose It! is a free mobile app.

For each participant, we implemented a 10% weight loss goal and provided a value for maximum daily caloric intake under the Budget column (Figure 1). Participants could search for foods, including restaurant items, or scan the bar codes of grocery items, with direct upload onto the Lose It! platform. Nutritional data were uploaded from food labels of grocery items as mandated by the US Food and Drug Administration

since 1994, and from restaurants as mandated by each state. The number of calories consumed could be increased with physical activity and logged in the Exercise column, which varied based on exercise duration and intensity. The net calories each day equaled the total of food calories minus exercise calories (Figure 1).

Figure 1. Example of a smartphone group Lose it! weight loss app screenshot.



During the initiation phase (baseline to day 30), we created a pattern of compliance. Participants received daily monitoring of caloric intake, with reminder messages via push notifications. During the application phase (day 31-90), weekly push notification summaries of compliance were sent, with reminder messages on missed days. The maintenance phase (days 91-180) consisted of weekly summaries with no reminder messages. Details of the smartphone intervention protocol and message examples have been published [17].

Food Journal Group

Similar to the smartphone group, for the food journal group we calculated a goal of 10% body weight reduction during the baseline visit. We used the pocket-sized *CalorieKing Food & Exercise Journal* [18]; each journal contained space to document 10 weeks of food intake; a weekly summary page; areas to record calories, fat, protein, and carbohydrate; and a 4-page mini-calorie, fat, and carbohydrate counter to reference beverages, meats, fruits, vegetables, cereals, snack foods, and popular fast food establishments.

Interventions for Both Groups

Participants in both groups received culturally competent dietary counseling by a clinician of their ethnic group, measuring cups, the US Department of Health and Human Services Dietary Approaches to Stop Hypertension eating plan guide, American Heart Association (AHA) cookbooks, and AHA reading materials (Suggested Servings from Each Food Group, Choosing a Restaurant, Dining Out Tips by Cuisine, and Ordering your Meal). At the end of each phase, both groups returned to the clinic for weight measurements and counseling. Study visits were at no cost to participants and caregivers.

Feasibility and Treatment Fidelity

We measured study feasibility using retention rates at day 180 and adherence to the self-monitoring intervention at days 30, 90, and 180. We measured adherence to the smartphone intervention through a coaching tool integrated with the Lose It! platform (Ascend for Lose It!). Smartphone participants, with caregiver assistance if needed, were instructed to enter food daily, though either the functionality of the mobile app or free text. We measured adherence to the food journal intervention through review of written entries, with similar caregiver assistance.

Measures

Demographics and Self-Reported Cerebrovascular Risk Factors

We collected data on participants' age, sex, ethnicity, marital status, employment status, educational level, weight and height (to calculate body mass index), and stroke-related disability level in a sociodemographic questionnaire. We assessed cognitive impairments using the Montreal Cognitive Assessment (MoCA); a score of 26 and higher was considered normal [19]. We collected personal health, family history, and past medical history information using a general health history form.

Objective Cerebrovascular Risk Factors

We collected cerebrovascular risk factors in both groups, with normative ranges derived from AHA guidelines [20,21]. Baseline study measurements were collected within 2 weeks of the index stroke.

Primary Outcome

The primary outcome was a reduction in total body weight. The United States Preventive Services Task Force guidelines indicate that a weight reduction of 5% to 10% yields measurable improvement in risk factors for cerebrovascular disease [22]. We measured overall weight reduction in both the smartphone and food journal groups.

We planned a subgroup analysis of weight loss in depression screen–positive and depression screen–negative participants a priori. Poststroke depression is one of the most frequent neuropsychiatric consequences of stroke, affecting approximately 30% of stroke survivors [23]. The presence of poststroke depression has been linked to reduced participation in structured rehabilitation programs [24,25]. Swipe out Stroke was a type of structured dietary rehabilitation intervention; therefore, we analyzed the combined group (smartphone and food journal).

Secondary Outcomes

Secondary outcomes included compliance with the weight loss intervention, improvement in depression, and, if abnormal, normalization of systolic blood pressure, serum low-density lipoprotein value, proportion of total hemoglobin, and proportion of serum coagulation factor VIII.

Data Management

Paper case report forms (screening, feasibility, and tracking forms) were collected and stored in the Institute for Stroke and Cerebrovascular Disease Research Coordinator office at the McGovern Medical School at University of Texas Health Science Center at Houston. We used REDCap version 6.10 (REDCap Consortium) for online form design, data entry, data verification, and data management. All forms were precoded to minimize errors. During data collection, forms were screened upon receipt for completeness.

Statistical Analysis

We reported descriptive statistics (frequency and percentage for categorical variables, and mean and standard deviation or median and IQR for continuous variables). We used 2-sample *t* test or Wilcoxon rank sum test to compare continuous variables, and Fisher exact test to compare categorical variables.

We obtained demographics and cerebrovascular risk factors by self-report. Cognitive testing was completed at baseline. Depression screening was completed at baseline and days 30, 90, and 180 and compared between the smartphone and food journal groups using Wilcoxon rank sum test. We measured the primary outcome as weight change in pounds; we also compared weight change by the presence of depression (defined as self-reported depression or PHQ-9 score ≥5). We performed all analyses using SAS version 9.4 (SAS Institute); we considered $P<.05$ to be significant.

Results

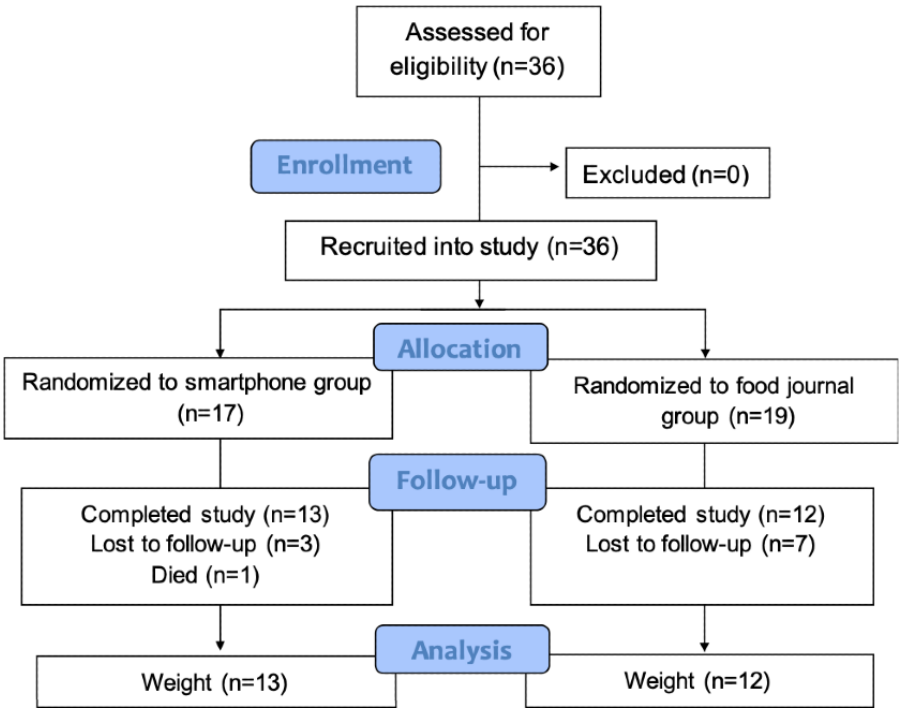
Study Period

Participants were enrolled in Swipe out Stroke between March 2015 and May 2016. The final follow-up visit occurred in December 2016. There were no significant changes in computer hardware or internet delivery resources during the study period. The duration of each in-person study visit was 30 minutes. For the smartphone group, engagement by a member of the study team took approximately 3 minutes per participant during the initiation phase, 2 minutes per participant during the application phase, and 1 minute per participant during the maintenance phase.

Feasibility and Treatment Fidelity

The CONSORT flow diagram in Figure 2 details Swipe out Stroke participant retention. A total of 25 (69%) stroke survivors completed Swipe out Stroke (13/17 in the smartphone group, 12/19 in the food journal group); 1 participant died in the smartphone group. The overall retention rate was 78% (28/36) at day 30, 72% (26/36) at day 90, and 69% (25/36) at day 180.

Figure 2. Swipe out Stroke Consolidated Standards of Reporting Trials (CONSORT) flow diagram.



Intervention adherence was significantly higher at day 90 in the smartphone group than in the food journal group (10/17, 59% vs 4/19, 21%; *P*=.04). In the smartphone group, adherence to 1 self-monitored diet entry per day was 59% (10/17) at day 30 and 35% (6/17) at day 180. Adherence to 1 self-monitored diet entry per day in the food journal group was 42% (8/19) at day 30 and 11% (2/19) at day 180.

Demographics and Self-Reported Cerebrovascular Risk Factors

Table 1 describes participant demographics and self-reported cerebrovascular risk factors by treatment group. Equal allocation was achieved between the smartphone and food journal groups. The mean participant age was 54 years, and mean education level was completion of 12th grade, some College, or General Education Development. Of the 36 participants, 20 (56%) were male and 23 (64%) were African American. The majority of participants had a source of income, either through retirement or disability (14/36) or employment (self or significant other, 20/36). The participants' overall median PHQ-9 scores (5.5) signified mild to moderate depression; the median MoCA score (23.5, IQR 21-26) indicated mild cognitive impairment.

Table 1. Demographics and self-reported cerebrovascular risk factors.

Characteristics	Total (n=36)	Smartphone (n=17)	Food journal (n=19)	*P* value
Age (years), mean (SD)	54.1 (9.4)	54.4 (10.9)	53.8 (8.2)	.87[a]
Race, n (%)				
Black	23 (64)	9 (53)	14 (74)	.30[b]
Hispanic	13 (36)	8 (47)	5 (26)	
Male sex, n (%)	20 (56)	9 (53)	11 (58)	.77[c]
Marital status, n (%)				
Married or domestic partner	21 (58)	10 (59)	11 (58)	.96[c]
Not married	15 (42)	7 (41)	8 (42)	
Employment status, n (%)				
Employed	20 (56)	10 (59)	10 (53)	.87[b]
Not employed	2 (6)	1 (6)	1 (5)	
Retired or receiving disability	14 (39)	6 (35)	8 (42)	
Education, n (%)				
Master's degree	5 (14)	1 (6)	1 (5)	.26[b]
Bachelor's degree	3 (8)	0 (0)	3 (16)	
Associate's degree or other	21 (58)	4 (24)	1 (5)	
High school, some college, or GED[d]	5 (14)	9 (53)	12 (63)	
Less than high school	2 (6)	3 (18)	2 (11)	
Modified Rankin Scale score, n (%)				
0	4 (11)	2 (12)	2 (11)	.92[b]
1	19 (53)	9 (53)	10 (53)	
2	8 (22)	3 (18)	5 (26)	
3	5 (14)	3 (18)	2 (11)	
Patient Health Questionnaire-9, median (IQR)	5.5 (3.0-9.5)	6 (2-10)	5 (3-9)	.74[e]
Patient Health Questionnaire-9, n (%)				.90[b]
0-4	17 (48)	8 (47)	9 (47)	
5-14	14 (39)	6 (35)	8 (42)	
≥15	5 (14)	3 (18)	2 (11)	
Self-reported depression or psychiatric disease, n (%)	12 (33)	5 (29)	7 (37)	.73[b]
Montreal Cognitive Assessment, median (IQR)	23.5 (21-26)	23 (19-25)	25 (21-27)	.15[e]
Medical history, n (%)				
Cardiac disease	11 (31)	4 (24)	7 (37)	.48[b]
Hypertension	35 (97)	17 (100)	18 (95)	>.99[b]
Hyperlipidemia	21 (60)	10 (63)	11 (58)	.78[c]
Family history of obesity	33 (94)	16 (94)	17 (94)	>.99[b]
Family history of stroke	24 (69)	12 (71)	12 (67)	>.99[b]
Personal history of stroke	7 (20)	3 (19)	4 (21)	>.99[b]

[a]*P* values obtained by 2-sample *t* test.

[b]*P* values obtained by Fisher exact test.

[c]*P* values obtained by chi-square test.

[d]GED: General Education Development.

[e]*P* values obtained by Wilcoxon rank sum test.

Objective Cerebrovascular Risk Factors

Table 2 describes objective cerebrovascular risk factors. The mean baseline body mass index was 35.7 (SD 5.7) kg/m^2, systolic blood pressure was 127.7 (SD 17.0) mm Hg, and serum total cholesterol, high-density lipoprotein, low-density lipoprotein, and triglycerides were in the reference ranges. Median proportion of total hemoglobin (0.061, IQR 0.056-0.07) and proportion of factor VIII (2.29, IQR 1.63-3.09) were both elevated.

Table 2. Objective cerebrovascular risk factors.

Risk factors	Total (n=36)	Smartphone (n=17)	Food journal (n=19)	*P* value
Body mass index (kg/m^2), mean (SD)	35.7 (5.7)	35.1 (5.3)	36.2 (6.2)	.57[a]
Weight (lb), mean (SD)	228.3 (47.1)	219.9 (49.6)	235.8 (44.7)	.32[a]
Systolic blood pressure (mmHg), mean (SD)	127.7 (17.0)	125.1 (9.8)	130.1 (14.1)	.38[a]
Total cholesterol (mmol/L), median (IQR)	9.27 (7.77-10.38)	8.94 (7.21-10.16)	9.43 (7.88-10.43)	.51[b]
Low-density lipoprotein (mmol/L), median (IQR)	4.88 (4.05-6.47)	4.94 (3.94-6.16)	4.77 (4.16-6.55)	.81[b]
High-density lipoprotein (mmol/L), median (IQR)	2.44 (2.0-2.66)	2.44 (2.05-2.66)	2.44 (1.94-2.55)	.51[b]
Triglycerides (mmol/L), median (IQR)	7.83 (5.55-9.63)	7.05 (4.5-9.55)	7.88 (5.83-11.77)	.40[b]
Proportion of total hemoglobin, median (IQR)	0.061 (0.056-0.07)	0.062 (0.056-0.096)	0.06 (0.056-0.065)	.39[b]
Proportion of factor VIII, median (IQR)	2.29 (1.63-3.09)	2.64 (1.68-3.70)	2.15 (1.44-3.04)	.26[b]

[a]*P* values by 2-sample *t* test.

[b]*P* values by Wilcoxon rank sum test.

Primary Outcome

Figure 3 summarizes descriptive findings on weight outcomes at each study collection time point. There were no statistically significant differences in weight loss between the smartphone and food journal groups. At day 180, the smartphone group had a median weight loss of 5.7 pounds (IQR –2.4 to 8.0), and the food journal group had a median weight loss of 6.4 pounds (IQR –2.2 to 12.5; *P*=.77). There was a 2.1% median weight change in the smartphone group, compared with a 2.9% median weight change in the food journal group (*P*=.63). In the combined smartphone and food journal analysis at day 180, the depression screen–positive group lost 3.9 pounds (IQR –10.1 to 14), and the depression screen–negative group lost 6.4 pounds (IQR –2.4 to 15; *P*=.49).

Secondary Outcomes

There was no difference in the proportion of total hemoglobin above 0.06 and proportion of factor VIII above 2 in the smartphone and food journal groups at 180 days (*P*=.68).

Depression was significantly lower at 30 days in the smartphone group than in the food journal group (PHQ-9 score 2 vs 8; *P*=.03) (Figure 4). Clinically relevant depression rates remained in the zero to minimal range for the smartphone group compared with the mild to moderate range in the food journal group at day 90 (PHQ-9 score 3.5 vs 4.5; *P*=.39) and day 180 (PHQ-9 score 3 vs 6; *P*=.12) (Figure 4). The IQR for PHQ-9 scores remained lower in the smartphone group than in the food journal group at day 30 (smartphone group IQR 1-4 vs food journal group IQR 4-10.5), day 90 (IQR 1-8 vs IQR 3-10), and day 180 (IQR 0-6 vs IQR 1-9).

Figure 3. Median weight change by intervention group and depression status.

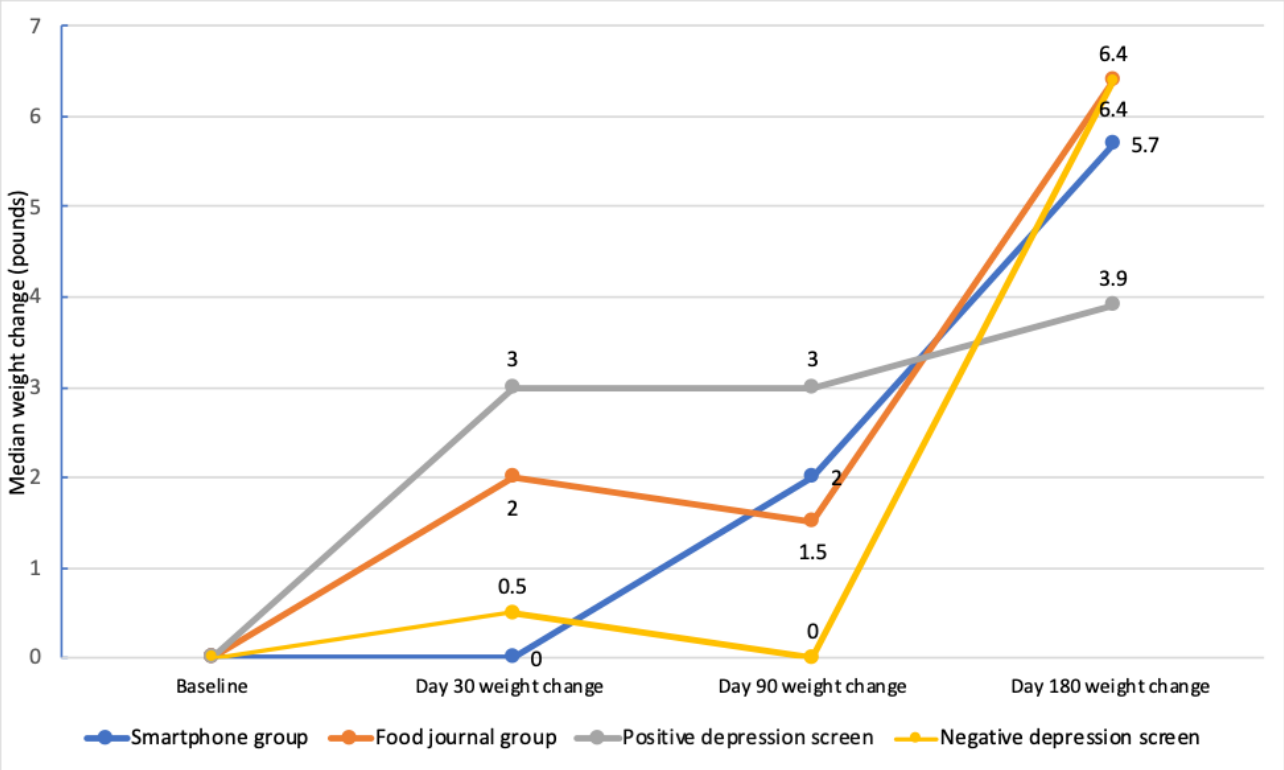
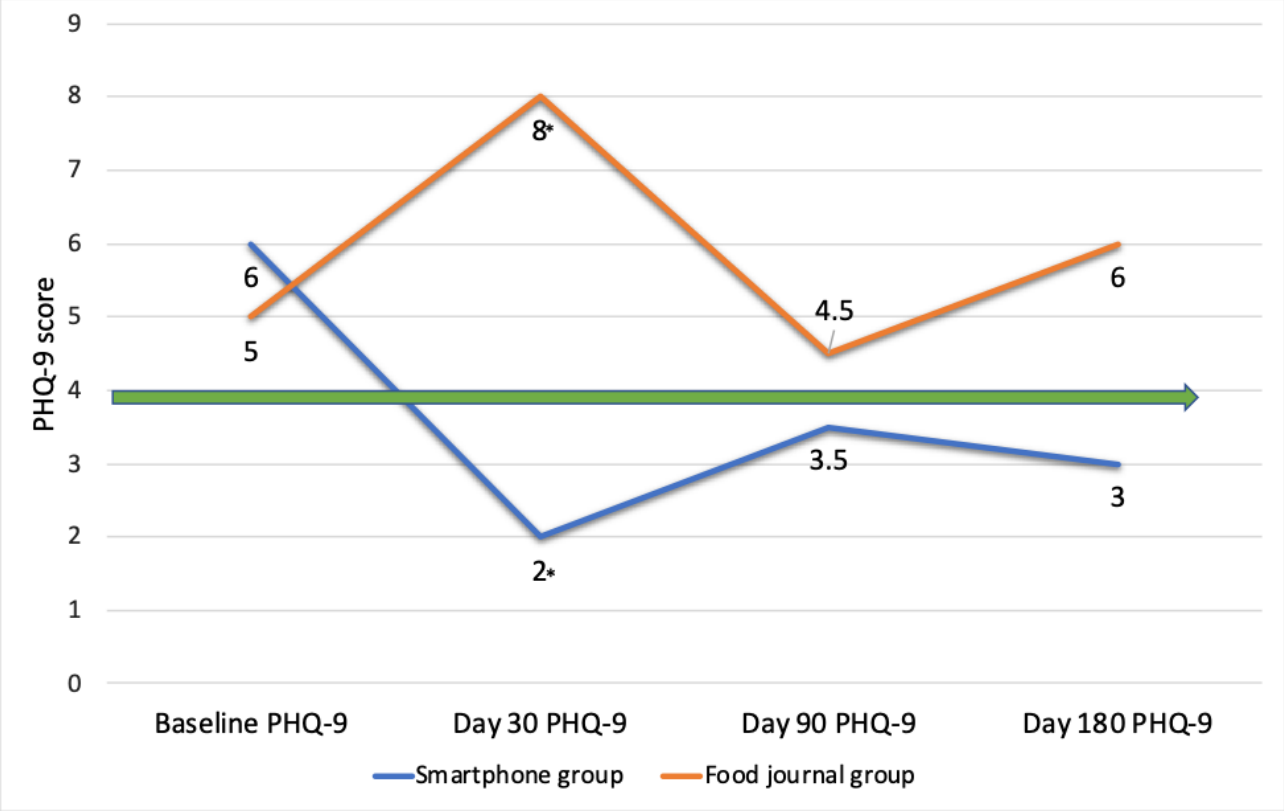


Figure 4. Depression by median Patient Health Questionnaire-9 (PHQ-9) score in the smartphone and food journal groups. Green arrow Indicates zero to minimal depression cutoff (PHQ-9 ≤4). *Statistically significant difference (*P*=.03).



Discussion

Principal Findings

The use of our smartphone-based intervention did not lead to a significant difference in weight change compared with the use of a food journal–based intervention. To our knowledge, this study is the first to implement a self-monitored smartphone-based intervention, with caregiver support, in minority patients with poststroke cognitive impairments. The time interval immediately following an acute stroke presents the opportunity to successfully change dietary patterns in patients and their caregivers, who influence the availability of healthy food choices. The overall retention rates establish an early precedent in this population at increased risk for recurrent stroke. A recent meta-analysis by Lui et al yielded no original clinical trials of the role of mHealth in recurrent stroke prevention [26].

The findings of increased patient engagement and adherence to self-monitoring in the smartphone group are consistent with studies in non–stroke populations [27-31]. Augmenting electronic health tools with non-Web–based components, such as culturally competent dietary counseling, measuring cups, and AHA cookbooks and reading materials, was well received by both the smartphone and food journal groups. Exit interviews reflected appreciation of a template that can be followed to establish healthy eating behaviors. The provision of this heightened supportive care, plus the Hawthorne effect, could have resulted in no significant difference in weight loss between the food journal and smartphone groups.

We found significantly decreased depression rates in the smartphone group at 30 days, which remained in the zero to minimal range; the food journal group had PHQ-9 scores indicative of mild to moderate depression throughout the study. Although both groups had built-in caregiver support, the smartphone group received reminder messages and positive reinforcement on a daily, then weekly, basis, providing support for the weight loss intervention. The use of the Lose It! mobile app as a tool for mitigation of poststroke depression is an interesting, cost-free adjunct in an underresourced minority population. Qu et al found that the top-rated mobile apps for depression were available at a direct cost for more advanced features (up to US $29.99 per month) or an indirect cost in terms of advertisements, which raised privacy concerns [32]. During the first 6 months after a stroke, depression appears to be reactive and is correlated with increased severity of impairment in activities of daily living at 1 year [33]. Kim et al described an improvement in depression at 12 weeks using a smartphone-based mHealth system in a Korean cohort [34]. The effect of direct patient engagement via existing mobile apps, on both poststroke depression and weight loss, is an exciting direction of future studies.

Limitations

There are several limitations to this study. First, the primary focus was feasibility and early treatment effect; there was not sufficient power to detect group differences. Second, study participants used their personally owned smartphones; during exit interviews, several participants noted financial limitations. The adherence to self-monitoring might have been different if smartphones with data plans had been provided. Third, study participation might have been increased within the smartphone group due to lower rates of depression. We did not examine interactions of cognitive impairment or depression with the treatment effect of each intervention, due to the small sample size in this pilot study. Fourth, our study evaluated short-term outcomes; assessment of the long-term effects of a smartphone-based intervention on weight loss and poststroke depression is a goal for future studies.

Conclusions

Swipe out Stroke provided useful data on designing subsequent weight management trials for obese minority stroke survivors. Both the smartphone and the food journal interventions resulted in weight loss. Data from Swipe out Stroke suggest that poststroke depression improves with smartphone-based engagement. Future minority research studies that include treatment of poststroke depression and cognitive rehabilitation may positively influence intervention efficacy.

Acknowledgments

NLI has received a University of Texas Southwestern Medical Center and Texas Health Resources Clinical Scholar Award (#4). NLI's and CCC's previous work at the Center for Clinical and Translational Sciences at the McGovern Medical School at University of Texas Health Science Center at Houston was funded by US National Institutes of Health (NIH)/National Center for Advancing Translational Sciences Clinical and Translational Award UL1 TR000371 and KL2 TR000370. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the NIH. NLI received the NIH/National Institute of Neurological Disorders and Stroke Diversity Supplement to P50 NS 044227, the University of Texas Specialized Program of Translational Research in Acute Stroke.

Authors' Contributions

NLI conceived the study concept and design. CCC, SIS, and JCG analyzed and interpreted the data. CCC conducted all statistical analyses. NLI, MB, EAN, CCC, SIS, and JCG critically revised the manuscript for important intellectual content. SIS and JCG provided administrative, technical, or material support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1482 KB - mhealth_v8i4e17816_app1.pdf]

References

1. National Institutes of Health, National Heart, Lung, and Blood Institute. Obesity Education Initiative: Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report. NIH Publication No 98-4083. Washington, DC: U.S. Department of Health and Human Services; 1998 Sep. URL: http://www.nhlbi.nih.gov/guidelines/obesity/ob_gdlns.pdf [accessed 2019-09-12]
2. Sacco RL, Boden-Albala B, Gan R, Chen X, Kargman DE, Shea S, et al. Stroke incidence among white, black, and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. Am J Epidemiol 1998 Feb 01;147(3):259-268. [doi: 10.1093/oxfordjournals.aje.a009445] [Medline: 9482500]
3. National Center for Health Statistics. Health, United States, 2016: With Chartbook on Long-Term Trends in Health. Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2017. URL: http://www.cdc.gov/nchs/data/hus/hus16.pdf [accessed 2019-11-07]
4. Christian JB, Lapane KL, Toppa RS. Racial disparities in receipt of secondary stroke prevention agents among US nursing home residents. Stroke 2003 Nov;34(11):2693-2697. [doi: 10.1161/01.STR.0000096993.90248.27] [Medline: 14551402]
5. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2017 update: a report from the American Heart Association. Circulation 2017 Mar 07;135(10):e146-e603. [doi: 10.1161/CIR.0000000000000485] [Medline: 28122885]
6. Ovbiagele B, Goldstein LB, Higashida RT, Howard VJ, Johnston SC, Khavjou OA, American Heart Association Advocacy Coordinating Committee and Stroke Council. Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association. Stroke 2013 Aug;44(8):2361-2375 [FREE Full text] [doi: 10.1161/STR.0b013e31829734f2] [Medline: 23697546]
7. Rimmer JH, Braunschweig C, Silverman K, Riley B, Creviston T, Nicola T. Effects of a short-term health promotion intervention for a predominantly African-American group of stroke survivors. Am J Prev Med 2000 May;18(4):332-338. [doi: 10.1016/s0749-3797(00)00129-x] [Medline: 10788737]
8. Ovbiagele B, Saver JL, Fredieu A, Suzuki S, McNair N, Dandekar A, et al. PROTECT: a coordinated stroke treatment program to prevent recurrent thromboembolic events. Neurology 2004 Oct 12;63(7):1217-1222. [doi: 10.1212/01.wnl.0000140493.83607.f1] [Medline: 15477541]
9. Pratt CA, Ha L, Levine SR, Pratt CB. Stroke knowledge and barriers to stroke prevention among African Americans: implications for health communication. J Health Commun 2003;8(4):369-381. [doi: 10.1080/10810730305725] [Medline: 12907401]
10. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. JMIR Mhealth Uhealth 2018 Apr 10;6(4):e92 [FREE Full text] [doi: 10.2196/mhealth.4478] [Medline: 29636320]
11. Asan O, Cooper Ii F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for health information technologies among US adults: analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
12. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
13. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-facing mobile apps to treat high-need, high-cost populations: a scoping review. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e136 [FREE Full text] [doi: 10.2196/mhealth.6445] [Medline: 27993761]
14. Cumming TB, Marshall RS, Lazar RM. Stroke, cognitive deficits, and rehabilitation: still an incomplete picture. Int J Stroke 2013 Jan;8(1):38-45. [doi: 10.1111/j.1747-4949.2012.00972.x] [Medline: 23280268]
15. Al-Qazzaz NK, Ali SH, Ahmad SA, Islam S, Mohamad K. Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment. Neuropsychiatr Dis Treat 2014;10:1677-1691 [FREE Full text] [doi: 10.2147/NDT.S67184] [Medline: 25228808]
16. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
17. Ifejika NL, Noser EA, Grotta JC, Savitz SI. Swipe out Stroke: feasibility and efficacy of using a smart-phone based mobile application to improve compliance with weight loss in obese minority stroke patients and their carers. Int J Stroke 2016 Jul;11(5):593-603. [doi: 10.1177/1747493016631557] [Medline: 26956032]
18. Borushek A. The CalorieKing Food & Exercise Journal. Costa Mesa, CA: Family Health Publications; 2006.

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. Jellinger PS, Handelsman Y, Rosenblit PD, Bloomgarden ZT, Fonseca VA, Garber AJ, et al. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for the management of dyslipidemia and prevention of cardiovascular disease. Endocr Pract 2017 Apr;23(Suppl 2):1-87. [doi: 10.4158/EP171764.APPGL] [Medline: 28437620]
21. Fox CS, Golden SH, Anderson C, Bray GA, Burke LE, de Boer IH, American Heart Association Diabetes Committee of the Council on Lifestyle and Cardiometabolic Health, Council on Clinical Cardiology, Council on Cardiovascular and Stroke Nursing, Council on Cardiovascular Surgery and Anesthesia, Council on Quality of Care and Outcomes Research, and the American Diabetes Association. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American Heart Association and the American Diabetes Association. Circulation 2015 Aug 25;132(8):691-718. [doi: 10.1161/CIR.0000000000000230] [Medline: 26246173]
22. Leblanc ES, O'Connor E, Whitlock EP, Patnode CD, Kapka T. Effectiveness of primary care-relevant treatments for obesity in adults: a systematic evidence review for the U.S. Preventive Services Task Force. Ann Intern Med 2011 Oct 04;155(7):434-447. [doi: 10.7326/0003-4819-155-7-201110040-00006] [Medline: 21969342]
23. Hackett ML, Pickles K. Part I: frequency of depression after stroke: an updated systematic review and meta-analysis of observational studies. Int J Stroke 2014 Dec;9(8):1017-1025. [doi: 10.1111/ijs.12357] [Medline: 25117911]
24. Paolucci S, Di Vita A, Massicci R, Traballesi M, Bureca I, Matano A, et al. Impact of participation on rehabilitation results: a multivariate study. Eur J Phys Rehabil Med 2012 Sep;48(3):455-466 [FREE Full text] [Medline: 22522435]
25. Subramanian SK, Chilingaryan G, Sveistrup H, Levin MF. Depressive symptoms influence use of feedback for motor learning and recovery in chronic stroke. Restor Neurol Neurosci 2015;33(5):727-740. [doi: 10.3233/RNN-150508] [Medline: 26444639]
26. Liu S, Feng W, Chhatbar PY, Liu Y, Ji X, Ovbiagele B. Mobile health as a viable strategy to enhance stroke risk factor control: a systematic review and meta-analysis. J Neurol Sci 2017 Jul 15;378:140-145 [FREE Full text] [doi: 10.1016/j.jns.2017.04.050] [Medline: 28566151]
27. Burke LE, Swigart V, Warziski TM, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
28. Yon BA, Johnson RK, Harvey-Berino J, Gold BC, Howard AB. Personal digital assistants are comparable to traditional diaries for dietary self-monitoring during a weight loss program. J Behav Med 2007 Apr;30(2):165-175. [doi: 10.1007/s10865-006-9092-1] [Medline: 17216341]
29. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
30. 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]
31. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-monitoring as a mediator of weight loss in the SMART randomized clinical trial. Int J Behav Med 2013 Dec;20(4):556-561 [FREE Full text] [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
32. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]
33. Robinson RG. The Clinical Neuropsychiatry of Stroke: Cognitive, Behavioral and Emotional Disorders Following Vascular Brain Injury. 2nd edition. Cambridge, MA: Cambridge University Press; 2006.
34. Kim DY, Kwon H, Nam K, Lee Y, Kwon H, Chung YS. Remote management of post stroke patients with a smartphone-based management system integrated in clinical care: prospective, nonrandomized, interventional study. J Med Internet Res 2020 Feb 27;22(2):e15377 [FREE Full text] [doi: 10.2196/15377] [Medline: 32130140]

Abbreviations

AHA: American Heart Association
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
mHealth: mobile health
MoCA: Montreal Cognitive Assessment
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by C Simons, R Ciptaningtyas, H Jin; comments to author 10.02.20; revised version received 05.03.20; accepted 23.03.20; published 22.04.20.

Please cite as:
Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI
Use of a Smartphone-Based Mobile App for Weight Management in Obese Minority Stroke Survivors: Pilot Randomized Controlled Trial With Open Blinded End Point
JMIR Mhealth Uhealth 2020;8(4):e17816
URL: http://mhealth.jmir.org/2020/4/e17816/
doi:10.2196/17816
PMID:32319963

Original Paper

A Mobile Health Solution Complementing Psychopharmacology-Supported Smoking Cessation: Randomized Controlled Trial

Laura Carrasco-Hernandez[1,2], Dr med; Francisco Jódar-Sánchez[3], PhD; Francisco Núñez-Benjumea[3], MSc; Jesús Moreno Conde[3], MSc; Marco Mesa González[1], MSc; Antón Civit-Balcells[4], Prof Dr; Santiago Hors-Fraile[5], MSc; Carlos Luis Parra-Calderón[3], MSc; Panagiotis D Bamidis[6], Prof Dr, PhD; Francisco Ortega-Ruiz[1], Dr med, PhD

[1]Smoking Cessation Unit, Medical-Surgical Unit of Respiratory Diseases, Virgen del Rocío University Hospital, Seville, Spain

[2]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Carlos III Institute of Health, Madrid, Spain

[3]Research and Innovation Group in Biomedical Informatics, Biomedical Engineering and Health Economy, Institute of Biomedicine of Seville, Virgen del Rocío University Hospital, Spanish National Research Council, University of Seville, Seville, Spain

[4]Department of Architecture and Computer Technology, School of Computer Engineering, Universidad de Sevilla, Seville, Spain

[5]Salumedia Labs, Seville, Spain

[6]Medical Physics Laboratory, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:
Francisco Jódar-Sánchez, PhD
Research and Innovation Group in Biomedical Informatics
Biomedical Engineering and Health Economy, Institute of Biomedicine of Seville
Virgen del Rocío University Hospital, Spanish National Research Council, University of Seville
Avenida Manuel Siurot S/N
Seville, 41013
Spain
Phone: 34 670943651
Email: francisco.jodar@juntadeandalucia.es

Abstract

Background: Smoking cessation is a persistent leading public health challenge. Mobile health (mHealth) solutions are emerging to improve smoking cessation treatments. Previous approaches have proposed supporting cessation with tailored motivational messages. Some managed to provide short-term improvements in smoking cessation. Yet, these approaches were either static in terms of personalization or human-based nonscalable solutions. Additionally, long-term effects were neither presented nor assessed in combination with existing psychopharmacological therapies.

Objective: This study aimed to analyze the long-term efficacy of a mobile app supporting psychopharmacological therapy for smoking cessation and complementarily assess the involved innovative technology.

Methods: A 12-month, randomized, open-label, parallel-group trial comparing smoking cessation rates was performed at Virgen del Rocío University Hospital in Seville (Spain). Smokers were randomly allocated to a control group (CG) receiving usual care (psychopharmacological treatment, n=120) or an intervention group (IG) receiving psychopharmacological treatment and using a mobile app providing artificial intelligence–generated and tailored smoking cessation support messages (n=120). The secondary objectives were to analyze health-related quality of life and monitor healthy lifestyle and physical exercise habits. Safety was assessed according to the presence of adverse events related to the pharmacological therapy. Per-protocol and intention-to-treat analyses were performed. Incomplete data and multinomial regression analyses were performed to assess the variables influencing participant cessation probability. The technical solution was assessed according to the precision of the tailored motivational smoking cessation messages and user engagement. Cessation and no cessation subgroups were compared using t tests. A voluntary satisfaction questionnaire was administered at the end of the intervention to all participants who completed the trial.

Results: In the IG, abstinence was 2.75 times higher (adjusted OR 3.45, *P*=.01) in the per-protocol analysis and 2.15 times higher (adjusted OR 3.13, *P*=.002) in the intention-to-treat analysis. Lost data analysis and multinomial logistic models showed different patterns in participants who dropped out. Regarding safety, 14 of 120 (11.7%) IG participants and 13 of 120 (10.8%) CG participants had 19 and 23 adverse events, respectively (*P*=.84). None of the clinical secondary objective measures showed

relevant differences between the groups. The system was able to learn and tailor messages for improved effectiveness in supporting smoking cessation but was unable to reduce the time between a message being sent and opened. In either case, there was no relevant difference between the cessation and no cessation subgroups. However, a significant difference was found in system engagement at 6 months (P=.04) but not in all subsequent months. High system appreciation was reported at the end of the study.

Conclusions: The proposed mHealth solution complementing psychopharmacological therapy showed greater efficacy for achieving 1-year tobacco abstinence as compared with psychopharmacological therapy alone. It provides a basis for artificial intelligence–based future approaches.

Trial Registration: ClinicalTrials.gov NCT03553173; https://clinicaltrials.gov/ct2/show/NCT03553173

International Registered Report Identifier (IRRID): RR2-10.2196/12464

(JMIR Mhealth Uhealth 2020;8(4):e17530) doi:10.2196/17530

KEYWORDS

smoking cessation; behavioral change; health recommender systems; mHealth; randomized controlled trial

Introduction

Tobacco use presents a major preventable public health problem; it is the leading cause of health deterioration and premature death. The World Health Organization recognizes smoking as a chronic systemic disease among the addictions capable of producing high physical damage related to multiple diseases [1]. Further, smoking kills over 7 million people annually, with global costs estimated at US $1.4 trillion [2].

Validated approaches to facilitate smoking cessation include nicotine replacement therapy, pharmacological treatment (ie, bupropion and varenicline), and behavioral and psychological support. Combining behavioral and psychological support with pharmacological treatment is currently the most effective intervention for achieving tobacco abstinence [3]. Moreover, bupropion was shown to approximately double the likelihood of long-term tobacco abstinence as compared with placebo [4]. Furthermore, varenicline (2 mg, total daily dose) was shown to triple the likelihood of maintaining long-term tobacco abstinence as compared with placebo [5]. Despite the proven efficacy of these treatments, a meaningful number of smokers fail to stop smoking, and many factors influence their success, including the motivation to stop and continuous support.

Mobile and web-based interventions have been used to facilitate effective behavioral changes in different health domains [6-9], including smoking cessation. Scientific evidence has indicated that mobile phone–based text messages supporting smoking cessation were nearly 1.7 times more successful than a control approach at 6 months [10]. It has been proven that the behavioral change impact is higher when users receive multiple tailored health recommendations [11,12]. However, in previous studies, a traditional computer-tailoring approach was used to generate such personalized health recommendations. Computer tailoring involves the generation of patient-specific recommendations, typically in the form of messages, by computers, and it is performed after personal assessment to match the characteristics, needs, and interests of the patients [13-16]. Cupertino et al [17] recently piloted a 12-week smoking cessation program involving text messaging with pharmacotherapy support, which showed promising abstinence rates at 3 months and high participant satisfaction. However, there is no evidence for the long-term abstinence efficacy of mobile-based tailored interventions combined with psychopharmacological therapies.

As part of the SmokeFreeBrain H2020 European Commission project [18], the Social-Local-Mobile (So-Lo-Mo) study investigated mobile and artificial intelligence (AI) technologies (health recommender system [HRS]) as a complementary aid to pharmacological treatments. The experimental intervention focused on providing ubiquitous tailored support to patients willing to stop smoking through a digital therapeutic mobile app solution.

Participants receiving the So-Lo-Mo intervention were provided with access to a code to be introduced in the app. The code activated the app, connecting it with the hospital patient database. The app automatically downloaded the necessary data to initialize the app profile, reducing the burden of having to create a user profile. The mobile app was connected with an AI system designed to learn from patient interests through their interactions with the app to dynamically (1) determine, personalize, and send motivational messages to support smoking cessation; (2) schedule the message delivery frequency according to the transtheoretical model of behavioral change [19]; and (3) calculate the most convenient time to send a motivational message to support smoking cessation for each patient. Specifically, the AI system used to tailor motivational messages is a hybrid HRS, whose design has been presented by Hors-Fraile et al [20].

The motivational messages were designed after conducting semistructured interviews involving two smoking cessation experts (a pulmonologist and a psychologist). Based on their comments, we identified the following five different topics, which were the most relevant for participants to succeed in their smoking cessation process: (1) general motivation; (2) healthy diet recommendations; (3) recommendations for an active lifestyle; (4) positive reinforcement messages to meet activity goals; and (5) benefits of being a nonsmoker. A total of 150 different messages were written for each topic by a health communication and health promotion PhD candidate, and they were validated by the two smoking cessation experts. This number of messages was chosen to ensure that, even in the worst case scenario where the AI system determines that a user needs to receive messages of a single topic, no user would get the same message twice during the intervention, as the maximum

number of messages that could be received is 150 during the 1-year study. This approach minimizes any potential robotic and static feelings.

Messages included health communication and health promotion strategies, such as creating empathy, adding new knowledge, and changing existing misconceptions. The messages were short (less than 100 words), written in simple Spanish, written using a close and friendly tone, and written with easy-to-understand terms to facilitate comprehension by smokers of all educational levels. Other studies have used steps similar to the ones used by us for message creation [21]. However, in our design, the message content was not associated with any specific behavioral change model. Each message was checked to be suitable for all genders. When a message was clearly gender specific (ie, relation between smoking and erectile dysfunction in men), an alternate suitable message for the other gender was also designed (ie, risks of smoking during pregnancy).

HRSs involve self-learning algorithms that can adapt to the constantly evolving needs and interests of users and offer a high level of personalization of recommendations, taking advantage of the so-called "collective intelligence." This is a new recommendation generation paradigm that contrasts the traditional tailoring approach, which provides static recommendations according to user responses to usually lengthy questionnaires. Thus, as this novel approach of tailoring is still in its infancy [22], it is of high interest to determine the relationship between the clinical outcomes of the So-Lo-Mo study and this type of technology.

The main objective of this So-Lo-Mo study was to compare usual psychopharmacological therapy (control group, CG) alone and alongside the aforementioned digital solution (intervention group, IG) for smoking cessation. The secondary objectives were to analyze health-related quality of life (HRQoL) and monitor healthy lifestyle and physical exercise habits. Complementarily, we assessed the impact of the AI-generated motivational messages on smoking cessation outcomes in the IG.

Methods

Study Design

This 12-month, randomized, open-label, parallel-group trial was performed at the Smoking Cessation Unit of Virgen del Rocío University Hospital in Seville (Spain) between October 24, 2016, and October 24, 2018, and it complied with the Declaration of Helsinki and Good Clinical Practice Guidelines. The recruitment period closed on October 23, 2017. The local ethics committee approved the study protocol, and written informed consent was obtained from each participant prior to inclusion. The clinical study design (NCT03553173) and technical study design (NCT03206619) have been published previously [23,24].

Randomization

A technician generated a random-group table (n=240) using computer methods and following a 1:1 ratio between groups. Clinicians enrolled the participants and assigned them to the group mentioned in the table according to their enrolment sequence. Participants were blinded to this allocation, as those in the IG were told that the provided mobile app was part of usual care. Participants in the CG were not informed about the existence of the app and did not have access to it.

Study Population

Smokers were recruited during routine visits to our outpatient clinic. The inclusion criteria were as follows: (1) age over 18 years and desire to stop smoking; (2) owning an Android smartphone (as the mobile app was only available for Android devices owing to time and resource constraints in the development phase of this study and Android phones were more likely to be owned by the target population owing to their lower entry price as compared with iPhones); and (3) ability to interact with the smartphone. Smartphone literacy was assessed by asking the participants if they commonly use other text exchange smartphone apps, such as Mail, SMS, and WhatsApp. The only exclusion criterion was any previous adverse effect related to the present pharmacological treatment.

Power Calculation and Recruitment

A sample size of 236 was calculated during the study design phase, according to the following parameters: CI, 95%; statistical power, 80%; CG success rate, 35%; IG success rate, 55%; and expected dropout rate, 20%. A total of 240 participants were recruited and randomized for the stratified analysis.

Intervention and Control Groups

Usual care in the Smoking Cessation Unit of Virgen del Rocío University Hospital consists of pharmacological therapy with bupropion (ie, Zyntabac 150 mg; GlaxoSmithKline, Brentford, UK) or varenicline (ie, Champix 0.5 mg or 1 mg; Pfizer, New York, New York, USA) plus behavioral therapy. To facilitate recruitment and avoid bias associated with treatment cost, the SmokeFreeBrain project financed the drugs for usual care. Thus, all participants received their assigned treatments free of charge. Behavioral therapy, which was provided during face-to-face follow-up consultations, included various psychological techniques, including motivational interviews and cognitive-behavioral therapy. This psychopharmacological therapy was provided to both CG and IG participants. CG participants (n=120) received psychopharmacological therapy alone, whereas IG participants (n=120) received psychopharmacological therapy and used the digital therapeutic solution. Further information regarding enrollment, allocation, and the study analysis phases is provided in Figure 1. The app utilized behavioral techniques by sending personalized motivational messages generated using AI with the intent to achieve better smoking cessation rates by improving program adherence and abstinence rates [20]. After the participant read each message, the app asked the participant to rate how relevant the message was for him or her, collecting feedback for the AI. Complementing the core messaging feature, the mobile app had other minor sections. These included a user profile linked to the performed physical activity level collected by Google Fit (Google, Mountain View, California, USA), four smoking cessation benefit indicators (savings, smoke-free days, regained life hours by not smoking, and number of nonsmoked cigarettes since quitting), text-based information about smoking cessation,

and a section containing relaxing and distracting elements (breathing exercises and minigames). Figure 2 presents an overview of the app features.

Figure 1. CONSORT diagram of the study.

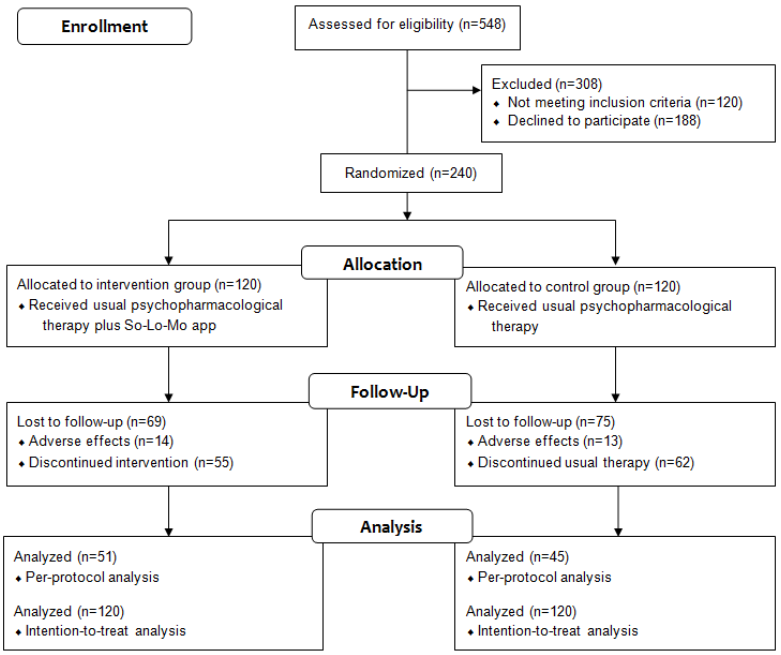
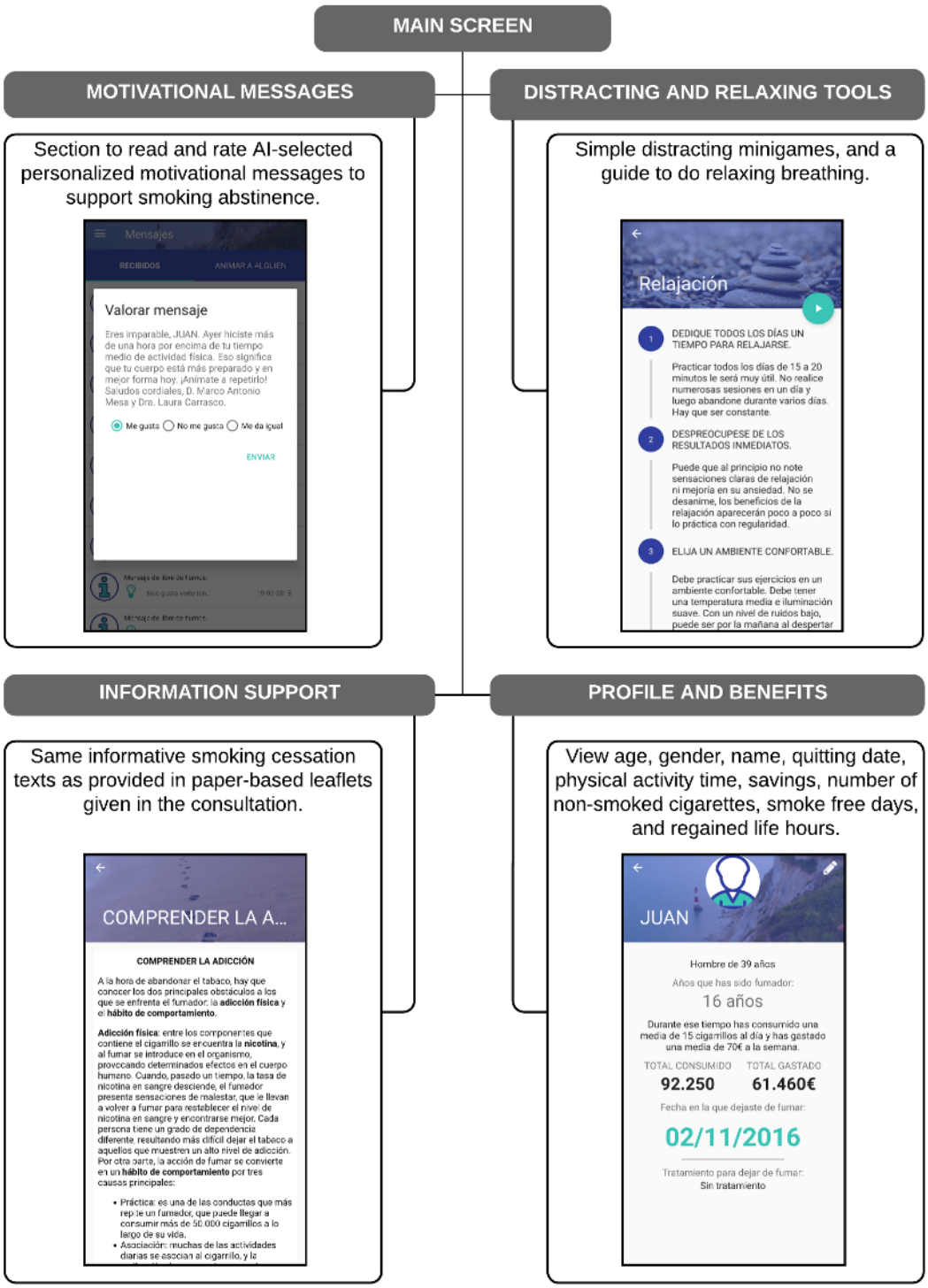


Figure 2. Structure of app features.



Measurements

Information from all participants included demographic (age and sex) and socioeconomic data (profession and employment status), consumption history (daily cigarettes smoked, living with smokers, partner smoking status, smoking cessation attempts, etc), clinical information (weight, height, blood pressure, comorbidities, etc), nicotine dependence measured using the Fagerström test for nicotine dependence [25], and motivation to stop smoking according to the Richmond test [26]. Safety was measured as the number of adverse events related to pharmacological therapy.

The main clinical outcome was the 1-year smoking abstinence rate measured by exhaled carbon monoxide (CO), which was assessed using a CO tester (Micro+ Smokerlyzer; CoVita, Santa Barbara, California, USA) and urine cotinine tests. Participants with an exhaled CO level greater than 6 ppm were considered smokers [27]. Urine cotinine (SmokeScreen test; Concept Smoke

Screen Ltd, Lincolnshire, UK) is a colorimetric test that measures the main nicotine metabolites, including cotinine. Participants with a cotinine concentration greater than 200 ng/ml were considered smokers [28]. Participants were considered smokers when at least one of the aforementioned conditions was met.

The HRQoL was assessed using the 36-item Short-Form Health Survey (SF-36) that was validated in Spanish [29] and the EuroQol 5-dimension 5-level (EuroQoL-5D-5L) questionnaire [30]. Physical activity was measured using the International Physical Activity Questionnaire (IPAQ) [31], and a healthy lifestyle was interpreted via body mass index (BMI) variations during follow-up consultations. A case report form, built upon the OpenClinica [32] tool, was developed to facilitate information management in the study. Information subsets were registered according to the following schedule: basal and 15, 30, 60, 90, 120, 180, and 365 (±5) days after the basal consultation [23].

The HRS impact on smoking cessation was assessed according to the precision of the recommendations sent by the system. Precision was calculated for both the message content recommendations and the time to open the messages, as the HRS aimed to optimize both variables. Further, we considered the generated engagement of the user with the system and the user appreciation of the messages as part of the technical evaluation of HRS impact. The starting point to measure the evolution of these metrics was the first day that a participant rated a message (December 5, 2016), and the assessment continued for the following 18 months to evaluate the progress. Detailed descriptions of these metric calculations are presented below.

Precision of the System

This measurement focused on the system's effectiveness in recommending relevant messages. The mobile app required the user to rate all read messages at least once; otherwise, the user could no longer interact with the app. There were three message rating options (like, dislike, and neutral). The ratings were then coded as 1 for "like," −1 for "dislike," and 0 for "neutral." The precision metric was calculated by dividing the number of hits (messages rated as "like") by the total number of sent messages (precision_p). A modification of this measurement was performed by considering "likes" and "neutrals" as hits (precision_p_n,). We used two variants of precision to obtain different perspectives on how the system performed in selecting relevant messages. The second approach focused on minimizing the negatively rated messages. Regardless of the variant, when we code the hits as 1 and misses as 0, this metric is in fact an arithmetic mean (we sum the number of hits and divide them by the total number of sent messages). The value of this metric ranges from 0 to 1. A higher value is associated with more system precision. This is because on sending messages at random, we expect 33% positive feedback, 33% neutral feedback, and 33% negative feedback. However, when the system learns and becomes more "intelligent," the percentage of negative feedback will decrease over time, increasing the value resulting from this metric.

Time to Open Motivational Messages

This pertained to the arithmetic mean of the elapsed time between the message being sent and opened in a 30-day period. We assumed that this metric would decrease over time because the system was expected to learn and become more "intelligent" in predicting the best time for the user to open and read the sent message.

Engagement With the System

To measure system engagement, we included a time stamp in each sent message and compared it with the time stamp sent by the app to the server when the user opened the given message. Thus, the engagement metric was determined by the ratio of rated messages calculated as the total of all rated messages from one user divided by the total number of messages sent to that user. As we coded each message rated with a "1" value and divided the finding by the total number of sent messages, this ratio coincides with the arithmetic mean. Thus, we assume that a higher value is associated with more user interest in the message and, consequently, higher user engagement. As this metric measured the engagement of the user with the system and not the engagement of the system with the user, we considered the set date to stop smoking for each user as the point to start counting each month. After this calculation, we offset the start date for each user to the first day the user rated the first message. This allowed us to compare the average engagement of each user across time. Therefore, the results are presented according the relative months to the user set date to stop smoking (eg, M3 results are the results of the third month after a participant stopped smoking). However, for one participant, M3 could be April if the participant stopped smoking in January, but for another participant, it could be December if the participant stopped smoking in September. The measurement of this follow-up was performed for the 12-month period that each participant in the IG was enrolled in the study, as participants in the CG did not use the mobile app.

Subjective Quality of the System

System quality was determined by the answers in an anonymized five-level Likert-type appreciation questionnaire [24], making it impossible to link the answers to the individual participants. The questionnaire was completed by each participant at the end of the 1-year follow-up period. The questions concerned the HRS according to the Recommender Systems Questionnaire of User Experience (ResQue) model [33] and the content of the messages according to the I-Change model [34]. Of the 15 constructs of the ResQue model, 12 were represented in the questions, as the remaining three did not apply to our HRS (ie, purchase intention). All the answers reflected the level of agreement of the participant with the proposed topics (1 indicating "totally disagree" and 5 indicating "totally agree"). We assessed how participants perceived the quality of the system after 12 months of usage by observing their questionnaire response distributions, means, and SDs. Group comparisons were not feasible owing to the complete anonymity of the questionnaire. The aim was to measure the degree of possible improvement of the system in the following areas included in the questionnaire: quality of recommended items, interaction adequacy, interface adequacy, perceived ease of use, perceived

usefulness, control/transparency, attitudes, behavioral intentions, and message content to influence smoking risk perception, confidence, social support, and coping action support.

Statistical Analysis

Descriptive analyses of participant characteristics according to absolute and relative frequencies for qualitative variables and mean (SD) for quantitative variables were conducted. For the primary endpoint and secondary outcomes (BMI, physical activity, and HRQoL), analyses were conducted on a per-protocol basis. These analyses included all participants who adhered to the schedule of eight consultations (basal through 365 days after the basal consultation).

Regarding incomplete data analyses, homoscedasticity tests were conducted among groups of cases with identical missing data patterns to evaluate whether data were missing completely at random [35]. A multinomial regression analysis was performed to assess the variables influencing the probability of a participant dropping out as compared with the probability of the treatment being effective. A “no efficacy” category was set as the reference category. As multinomial model effects are relative to the reference category, to assess whether the explicative variable effects were different between the “dropout” and “efficacy” categories, the same model was adjusted using the “dropout” category as a reference. Consequently, whether the variable effects were different between the “dropout” and “efficacy” categories was assessed. The relative risk ratio (RRR) with 95% CI was used for each model.

Lost data analysis and multinomial logistic models showed different patterns in participants who dropped out as compared with the patterns in those who completed the study, regardless of treatment efficacy. Therefore, smoking abstinence at 1 year was analyzed using logistic regression models on a per-protocol basis and intention-to-treat basis, and the effect measures were the OR and 95% CI.

For the selection of logistic regression and multinomial logistic models in per-protocol (n=94) analysis and intention-to-treat (n=240) analysis, a two-stage strategy was adopted. In the first step, the variable importance was quantified using Random Forest [36] with the mean decrease in accuracy as the score. Variables with a score above 0.5 were included in an Akaike information criterion–based stepwise selection strategy. The following two restrictions were applied to the final models selected: (1) absence of a pattern in model residuals and (2) variables with a generalized variance inflation factor [37] above 5 were not allowed in order to avoid collinearity.

To determine whether the HRS metrics had an impact on the clinical outcomes, we divided IG participants in cessation and no cessation subgroups at 12 months of the intervention. Thereafter, we conducted *t* tests for the 12-month results of precision, time to rate messages, and engagement, analyzing each of these two subgroups.

Results

Characteristics

Both groups had similar baseline characteristics, except for the maximum abstinence time and smoking cessation attempts (Table 1).

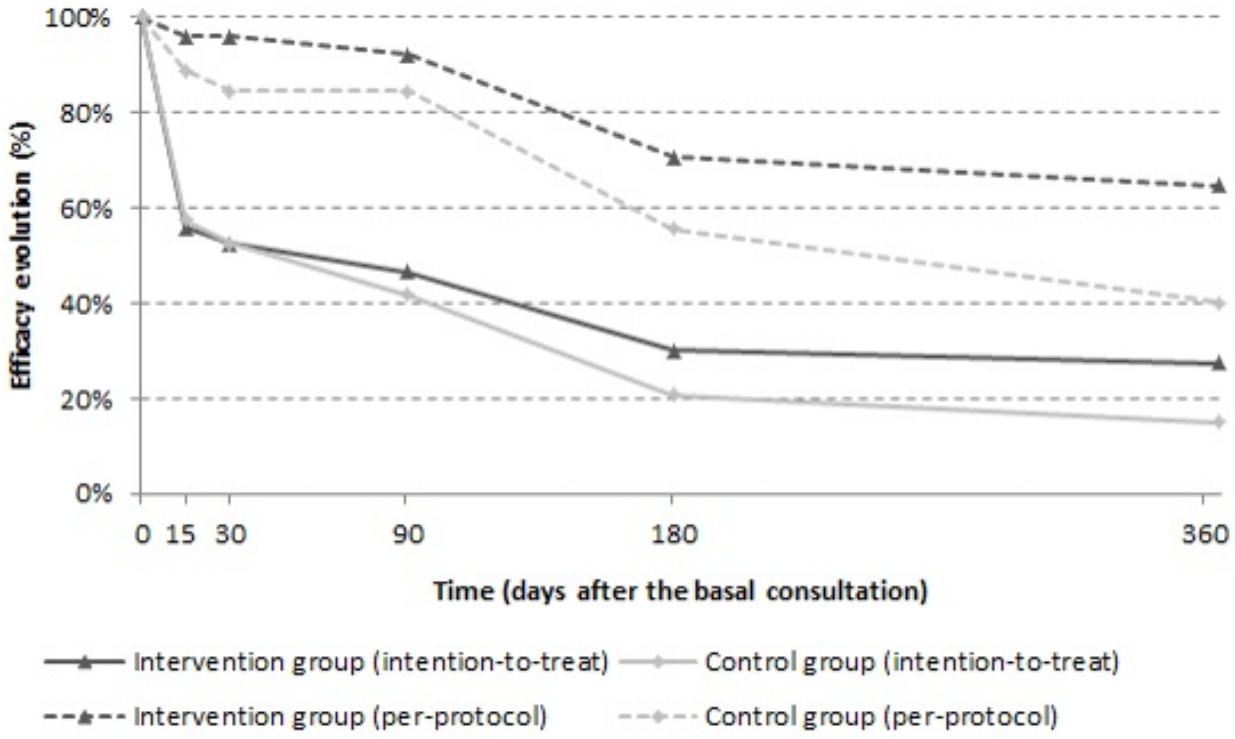
Table 1. Participant baseline characteristics.

Characteristic	Intervention group (n=120)	Control group (n=120)	*P* value
Female, n (%)	65 (54.2)	52 (43.3)	.09
Age (years), mean (SD)	48.38 (9.49)	50.93 (10.85)	.05
Age at smoking onset (years), mean (SD)	16.94 (4.07)	16.67 (3.60)	.54
Daily cigarettes, mean (SD)	21.45 (8.97)	20.75 (9.39)	.44
Lives with smokers, n (%)	45 (37.5)	49 (40.8)	.60
Partner smokes, n (%)	71 (59.2)	65 (54.2)	.43
Smoking cessation attempts, mean (SD)	0.88 (1.08)	1.14 (1.07)	.045
Maximum abstinence time, mean (SD)	10.45 (25.49)	13.68 (25.08)	.003
Body mass index, mean (SD)	27.02 (4.91)	27.03 (6.46)	.73
Unemployed, n (%)	33 (27.5)	38 (31.7)	.48
Previous treatments, n (%)	46 (38.3)	44 (36.7)	.79
Varenicline	13 (10.8)	15 (12.5)	.69
Bupropion	14 (11.7)	16 (13.3)	.70
Nicotine	13 (10.8)	15 (12.5)	.84
Others	16 (13.3)	11 (9.2)	.31
Comorbidities, n (%)	46 (38.3)	59 (49.2)	.12
Charlson index, mean (SD)	0.83 (1.25)	1.08 (1.34)	.06
Fagerström score, mean (SD)	5.89 (1.91)	5.60 (1.99)	.38
Richmond, mean (SD)	9.32 (0.80)	9.28 (0.88)	.88

Efficacy (Unadjusted): 1-Year Smoking Abstinence Rate

IG and CG participants who completed the study (per-protocol analysis) achieved efficacy rates of 64.7% (CI 51.6%-77.8%) and 40.0% (CI 25.7%-54.3%), respectively (*P*=.02; OR 2.75, CI 1.20-6.29) and a number needed to treat of 4 (CI 2.0-19.0). In the intention-to-treat analysis, IG and CG participants achieved efficacy rates of 27.5% (CI 19.5%-35.5%) and 15.0% (CI 8.6%-21.4%), respectively (*P*=.02; OR 2.15, CI 1.13-4.08) and a number needed to treat of 8 (CI 4.0-43.0). Figure 3 shows the efficacy evolution during the 1-year follow-up in each study group according to both analytic approaches.

Figure 3. Efficacy evolution during the 1-year follow-up.



Body Mass Index and Physical Activity

BMI variances were similar for both groups. In the IG and CG, the mean BMI variances were 1.01 (CI 045-1.57) and 1.10 (CI 0.72-1.48), respectively, at 6 months and were 1.47 (CI 0.90-2.03) and 1.22 (CI 0.67-1.75), respectively, at 12 months. The between-group mean BMI differences at 6 and 12 months were −0.09 (CI −0.77 to 0.60; *P*=.80) and 0.25 (CI −0.53 to 1.03; *P*=.52), respectively.

According to the observed IPAQ scores, physical activity evolution patterns were similar in both study groups. Moreover, in the IG, 12.8% (6/47) and 25.5% (13/51) of participants increased their physical activity at 6 and 12 months, respectively, whereas 14.9% (7/47) and 13.7% (7/51) of participants reduced it at the corresponding points. On the other hand, in the CG, 14.6% (6/41) and 24.4% (11/45) of participants increased their physical activity at 6 and 12 months, respectively, whereas 24.4% (10/41) and 8.9% (4/45) of participants reduced it at the corresponding points. Table 2 summarizes the BMI and IPAQ results.

Table 2. Body mass index, physical activity and health-related quality of life variances in each group.

Variable	Intervention group[a] (n=51)	Control group[a] (n=45)	*P* value
BMI[b] changes (kg/m^2)			
At 6 months[c]	1.01 (0.45 to 1.57)	1.10 (0.72 to 1.48)	.80
At 12 months	1.47 (0.90 to 2.03)	1.22 (0.67 to 1.75)	.52
IPAQ[d] changes at 6 months[c]			
Improvement	6 (12.8)	6 (14.6)	.47
No change	34 (72.3)	25 (61.0)	
Worsening	7 (14.9)	10 (24.4)	
IPAQ changes at 12 months			
Improvement	13 (25.5)	11 (24.4)	.73
No change	31 (60.8)	30 (66.7)	
Worsening	7 (13.7)	4 (8.9)	
VAS[e] changes			
At 6 months[c]	4.04 (−0.76 to 8.84)	3.88 (−0.57 to 8.32)	.96
At 12 months	5.78 (1.60 to 9.97)	2.78 (−1.86 to 7.41)	.33
SF-36[f] changes at 12 months			
Physical function	5.39 (2.18 to 8.61)	5.33 (1.43 to 9.24)	.98
Physical role	2.82 (−0.47 to 6.10)	5.69 (0.43 to 10.96)	.34
Body pain	3.78 (−1.52 to 9.09)	5.71 (−1.07 to 12.49)	.65
General health	6.35 (1.83 to 10.87)	6.38 (0.73 to 12.02)	.99
Vital	7.23 (1.92 to 12.54)	6.94 (1.75 to 12.14)	.94
Social	0.74 (−4.86 to 6.33)	7.78 (2.82 to 12.73)	.06
Emotional	4.90 (−0.22 to 10.02)	3.70 (−0.78 to 8.19)	.73
Mental	3.33 (−2.12 to 8.79)	4.78 (−1.06 to 10.61)	.72

[a]Data are expressed as mean (95% CI) or n (%).

[b]BMI: body mass index.

[c]Data are missing.

[d]IPAQ: International Physical Activity Questionnaire.

[e]VAS: visual analog scale.

[f]SF-36: 36-item Short-Form Health Survey.

Health-Related Quality of Life

Based on the EuroQoL-5D-5L questionnaire results, both groups had improved HRQoL scores at 6 months. Although the IG showed a greater improvement at 12 months, the difference was not statistically significant (Table 2). The mean between-group visual analog scale differences at 6 and 12 months were 0.17 (95% CI −6.36 to 6.70) and 3.01 (95% CI −3.14 to 9.16), respectively. Regarding the SF-36 dimensions, both groups showed improvement at 12 months, although there were no significant between-group differences. Table 2 summarizes the HRQoL results.

Safety: Adverse Events

Nineteen adverse events (11/19 [58%] associated with bupropion and 8/19 [42%] with varenicline) were identified in 14 IG participants (11.7%), whereas 23 adverse events (10/23 [43%] associated with bupropion and 13/23 [57%] with varenicline) were identified in 13 CG participants (10.8%) (*P*=.84).

The most frequent events associated with bupropion were headache (6/21, 28.6%), insomnia (4/21, 19.0%), vertigo (4/21, 19.0%), acute abdominal pain (2/21, 9.5%), and others (5/21, 23.8%), and those associated with varenicline were nausea (6/21, 28.6%), acute abdominal pain (4/21, 19.0%), insomnia (3/21, 14.3%), vomiting (2/21, 9.5%), and others (6/21, 28.6%).

Lost Data Pattern Analysis

The pattern of missing efficacy variable values was not totally random (*P*<.001). Multimedia Appendix 1 shows the variables whose distributions differed depending on the group; the values were not equally distributed between the participants who dropped out and those who completed the study. Table 3 shows the parameters estimated by the multinomial logistic regression

model for each of the variable effectiveness levels (no efficacy, efficacy, and dropout) with respect to the reference category.

Table 3. Multinomial logistic regression.

Variable	RRR[a]	95% CI	*P* value
Efficacy vs no efficacy			
Intercept	0.03	0.00-19.57	
Body mass index	0.96	0.88-1.06	.42
Maximum abstinence time	1.14	0.97-1.35	.11
Fagerström score	0.99	0.74-1.31	.93
Age at smoking onset	1.05	0.93-1.19	.43
Daily cigarettes	0.99	0.93-1.06	.82
Charlson index	1.00	0.72-1.39	.99
Low physical activity	0.63	0.18-2.18	.47
Medium physical activity	0.32	0.09-1.11	.07
Richmond score	1.43	0.82-2.49	.21
Group (intervention)	3.25	1.33-7.95	.01
Drug (varenicline)	0.81	0.33-1.96	.64
Dropout vs no efficacy			
Intercept	0.17	0.001-28.00	
Body mass index	1.01	0.93-1.09	.88
Maximum abstinence time	0.81	0.68-0.97	.03
Fagerström score	1.31	1.02-1.68	.03
Age at smoking onset	1.07	0.96-1.19	.23
Daily cigarettes	1.00	0.96-1.06	.87
Charlson index	0.86	0.65-1.13	.27
Low physical activity	0.78	0.29-2.09	.62
Medium physical activity	0.48	0.17-1.32	.15
Richmond score	1.13	0.73-1.75	.59
Group (intervention)	0.93	0.43-2.03	.85
Drug (varenicline)	0.36	0.17-0.77	.01
Efficacy vs dropout			
Intercept	0.18	0.001-61.91	
Body mass index	0.96	0.88-1.04	.29
Maximum abstinence time	1.41	1.18-1.67	<.001
Fagerström score	0.75	0.58-0.97	.03
Age at smoking onset	0.98	0.89-1.08	.73
Daily cigarettes	0.99	0.94-1.05	.69
Charlson index	1.17	0.87-1.57	.31
Low physical activity	0.81	0.27-2.43	.71
Medium physical activity	0.67	0.19-2.34	.53
Richmond score	1.27	0.77-2.09	.36
Group (intervention)	3.50	1.56-7.83	.002
Drug (varenicline)	2.26	1.05-4.86	.04

[a]RRR: relative risk ratio.

Importantly, the probability of being in the "efficacy" category was higher for IG participants as compared with the "no efficacy" category (RRR=3.25; *P*=.01). Furthermore, increasing the baseline Fagerström score (higher nicotine dependence) in a unit rendered a smoker more likely to belong to the "dropout" category as compared with the "no efficacy" category (RRR=1.31; *P*=.03). Increasing the maximum smoking abstinence time reduced the probability of a smoker being in the "dropout" category as compared with the "no efficacy" category (RRR=0.81; *P*=.03). Furthermore, the probability of being in the "dropout" category increased for participants using bupropion as compared with the "no efficacy" category (RRR=0.36; *P*=.01).

Efficacy (Adjusted): 1-Year Smoking Abstinence Rate

Table 4 shows the logistic regression analysis results for participants who completed the study (per-protocol basis). Notably, for the IG participants, the efficacy was 3.45 times higher with adjustment for age, motivation to stop smoking (Richmond scale), and comorbidity level. The efficacy probability decreased by 28.9% when a participant was not in the IG, with adjustment for the rest of the variables.

The results of the lost data pattern analysis and those obtained from the multinomial logistic models showed different patterns in participants who dropped out as compared with those who completed the study, regardless of treatment efficacy. However, as the variables associated with the dropout probability indicated a lack of treatment efficacy (belonging to the CG, short withdrawal period duration, fewer attempts, greater number of daily cigarettes, and baseline Fagerström score), a value of 0 (no efficacy) was assigned to participants who dropped out of the study. Therefore, the worst possible scenario models efficacy by intention-to-treat using a logistic regression model (Table 5).

These findings demonstrate that intervention efficacy was 3.13 higher with adjustment for the rest of the variables. Hence, the probability of long-term smoking cessation using the digital therapeutic solution was over three times higher than the probability of cessation with usual care alone.

Table 4. Logistic regression analysis of the 1-year smoking abstinence rate (per-protocol analysis).

Variable	OR	95% CI	*P* value
Intercept	0.00	0.00-0.02	—
Intervention group	3.45	1.39-9.13	.01
Age	1.09	1.03-1.17	.01
Motivation to stop smoking (Richmond score)	1.90	1.07-3.52	.03
Charlson index, medium comorbidity	0.76	0.16-3.60	.72
Charlson index, high comorbidity	0.07	0.01-0.45	.01

Table 5. Logistic regression analysis of the 1-year smoking abstinence rate (intention-to-treat analysis).

Variable	OR	95% CI	*P* value
Intercept	0	0.00-0.07	—
Group (intervention)	3.13	1.53-6.71	.002
Age	1.04	1.00-1.08	.04
Drug (varenicline)	1.49	0.74-3.04	.27
Maximum abstinence time	1.22	1.07-1.39	.004
Nicotine dependence (Fagerström score)	0.82	0.68-0.98	.03
Motivation for smoking cessation (Richmond score)	1.62	1.02-2.69	.049
Low physical activity (IPAQ[a])	0.93	0.30-2.55	.89
Medium physical activity (IPAQ)	0.66	0.19-1.95	.47

[a]IPAQ: International Physical Activity Questionnaire.

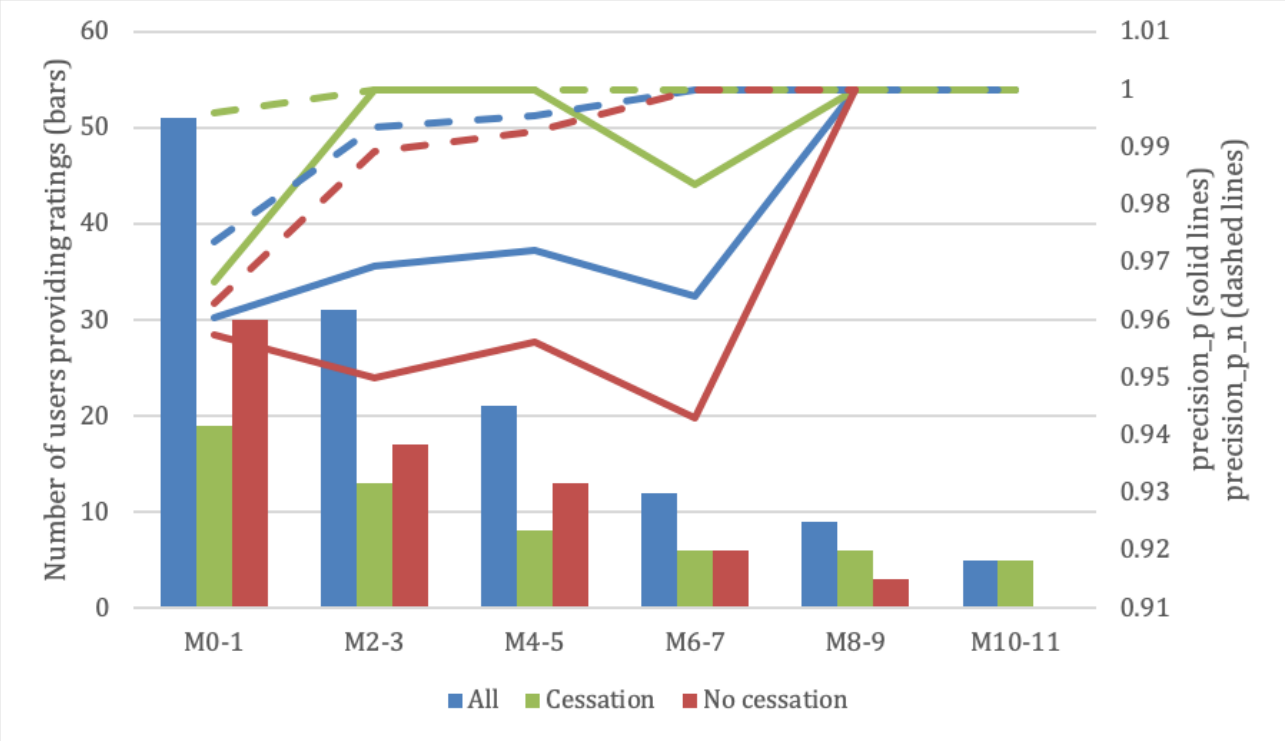
Health Recommender System Impact on Smoking Cessation Outcomes

A total of 17,111 messages and 2617 ratings were provided in the study, and there were 2311 ratings during the 12 months of follow up, 261 in the presmoking cessation phase, and 45 at 13 months or later, which were not part of this analysis.

Precision

The *t* test results for each month indicated that there were no significant differences for system precision_p or precision_p_n in the cessation and no cessation subgroups. The detailed results are presented in Multimedia Appendix 2. A bimonthly graphical representation of the evolution of precision_p and precision_p_n over time is shown in Figure 4.

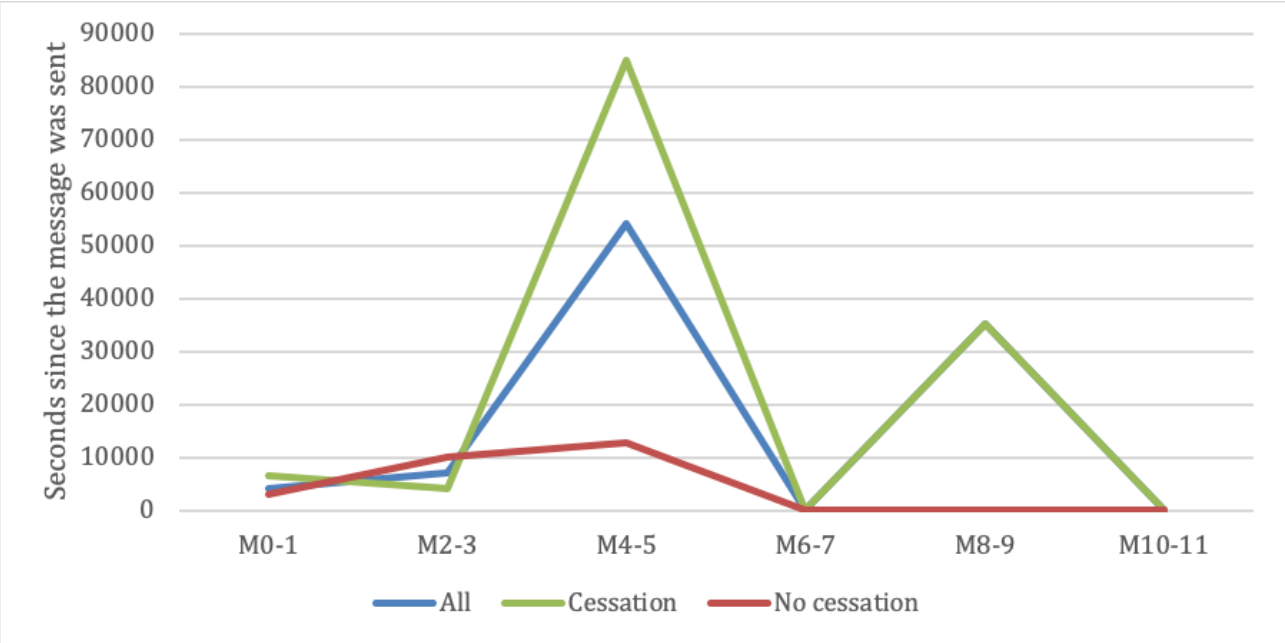
Figure 4. Evolution of health recommender system precision and number of users providing ratings by subgroup. M: month; precision_p: precision calculated by dividing the number of hits (messages rated as "like") by the total number of sent messages; precision_p_n: precision calculated by dividing the number of hits (messages rated as "like" and "neutral") by the total number of sent messages.



Time to Open Messages

We performed *t* tests (results are shown in Multimedia Appendix 2) and found no significant difference between the cessation and no cessation subgroups at any time point. A bimonthly graphical representation of the evolution of the time-to-open metric is shown in Figure 5.

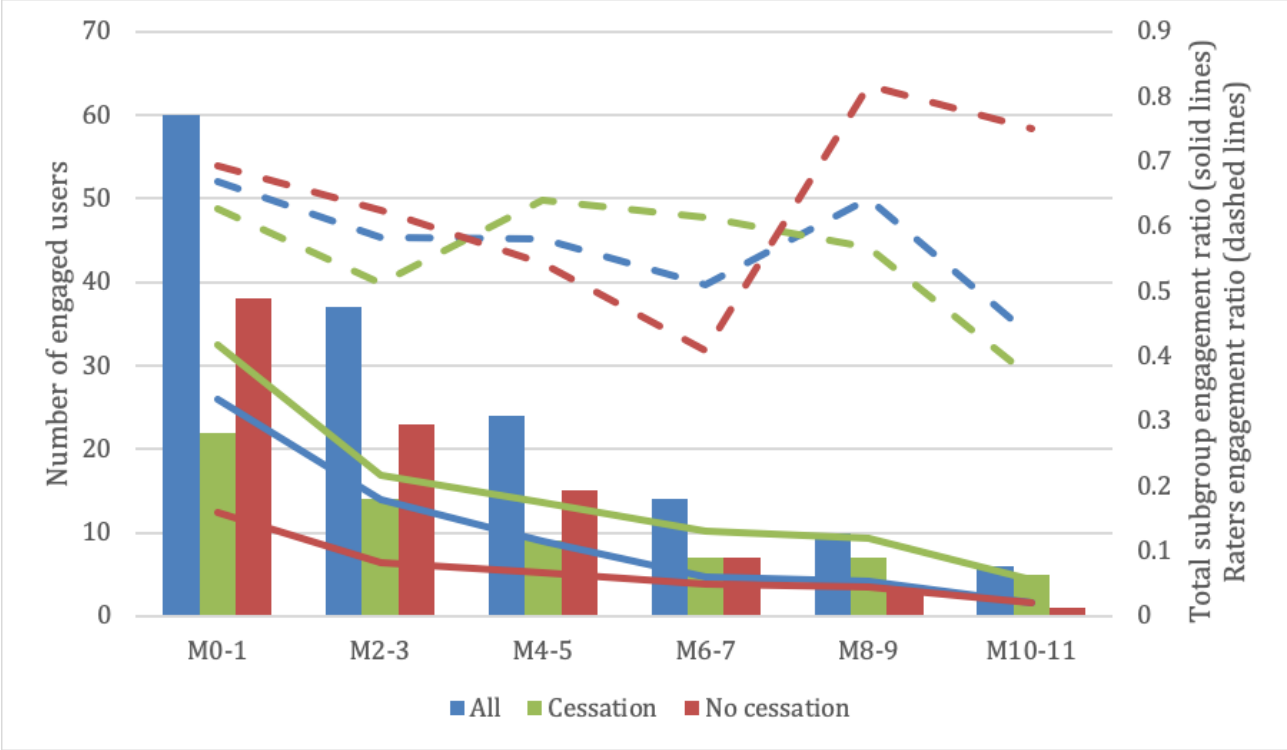
Figure 5. Mean time between a message being sent and opened. M: month.



Engagement

The *t* tests shown in Multimedia Appendix 2 indicate that there were significant differences between the cessation and no cessation subgroups in only the first 7 (P=.04), 8 (P=.03), and 10 (P=.04) months, showing a trend toward higher engagement in the cessation subgroup. A bimonthly graphical representation of the evolution of the IG participants' engagement with the system over time is shown in Figure 6.

Figure 6. Evolution of user engagement. M: month.



Perceived Quality

Only 32 participants volunteered to respond to the questionnaire. However, not all participants who responded to the questionnaire completed it. The results are presented in Table 6.

Table 6. Answers to the perceived quality questionnaire.

Question	N	Minimum score	Maximum score	Mean	SD
The messages recommended to me matched my interests.	32	4	5	4.19	0.40
The messages recommended to me were novel.	32	3	5	4.06	0.50
The messages recommended to me were diverse.	32	4	5	4.47	0.51
The layout of the message interface was adequate	32	3	5	4.03	0.60
I found it easy to tell the system what I like/dislike.	32	4	5	4.34	0.48
I felt in control of modifying my interest profile.	32	3	5	3.63	0.66
I became familiar with the messaging system very quickly.	30	3	5	4.00	0.37
I understood why the messages were recommended to me.	31	3	5	4.03	0.48
The messages gave me good suggestions.	31	4	5	4.23	0.43
Overall, I am satisfied with the messages.	31	4	5	4.35	0.49
The messages can be trusted.	31	3	5	3.97	0.55
I would recommend the use of the message recommendations to my friends who smoke.	31	3	5	3.61	0.72
The messages convinced me that I am at risk for health problems if I do not quit smoking.	31	3	5	3.71	0.64
The messages convinced me that my smoking is a risky habit.	31	3	5	4.06	0.57
The messages convinced me of the advantages of smoking cessation.	30	3	5	3.67	0.66
The messages showed me how to get social support from others during cessation.	31	3	5	3.68	0.54
The messages helped me feel confident that I could successfully quit smoking.	31	3	5	4.00	0.26
The messages helped prepare actions to cope with difficult situations.	31	4	5	4.19	0.40

Discussion

Principal Findings

Clinical Outcomes

The combination of our digital therapeutic solution with psychopharmacological treatment was more effective in achieving tobacco abstinence at 12 months as compared with psychopharmacological treatment alone. According to the per-protocol analysis findings, abstinence was 2.75 times higher among IG participants than among CG participants (adjusted OR of 3.45 for age, motivation, and comorbidity level). According to the intention-to-treat analysis, abstinence was 2.15 times higher among IG participants than among CG participants (adjusted OR of 3.13 for age, drug used, maximum abstinence time, nicotine dependence, motivation, and physical activity). Moreover, our findings suggest that the probability of dropout increased for participants who used bupropion and had high nicotine dependence, whereas a high maximum smoking abstinence time reduced this probability.

Despite mobile apps offering several advantages over traditional smoking cessation methods, few studies have examined the content quality or the effectiveness of apps promoting smoking cessation [38-40]. To the best of our knowledge, this study provides the first scientific evidence of the efficacy of a behavioral change intervention via a mobile app powered by AI for smoking cessation at a 1-year follow-up. Furthermore, no other study has provided statistically significant results for a mobile intervention used in combination with smoking cessation drugs. Previous studies published their cessation results with 6 months of follow-up at the most, as described in the comparison with prior work subsection below. We provide evidence for the long-term impact of our digital therapeutic solution, with a follow-up period that is double of the follow-up period in previous studies. Further, this study is the first to analyze physical activity via the IPAQ after smoking cessation. Physical activity evolution patterns were similar in the IG and CG, with an increase in the percentage of participants showing improved physical activity and no significant difference between the IG and CG participants.

Weight gain during the first year of abstinence concerns smokers [41], and it may be a potential barrier to pursue smoking cessation. In this study, weight gain, measured by an increase in BMI, occurred in both groups. The app did not affect weight gain as there was no difference between the IG and CG.

A previous review showed that a higher number of cigarettes smoked was associated with a lower QoL. Additionally, a low QoL and depression were related with lower odds of successfully stopping smoking [42]. Participants who received treatments, including varenicline and bupropion, reported QoL improvements and increased abstinence duration as compared with those who did not receive pharmacotherapy [42,43]. The present results are comparable to these findings. There was a positive trend in the HRQoL scores reported by the IG, but with no relevant difference as compared with the scores in the CG.

Technical Outcomes

The complementary technical results showed that the HRS is able to learn the participants' interests regarding the support message topics for smoking cessation. The system was more precise at the end of the intervention than at the beginning, as was expected for a recommender system. However, the minimum precision reached was very high (over 96%). We believe that there may be different reasons for this extremely high value. For instance, it could be due to the implemented hybrid HRS algorithm that mitigates the cold start problem and accurately recommends messages from the beginning according to a weighting formula described in a previous publication [20]. Additionally, it could be due to participants believing that their ratings would be viewed by health care professionals; hence, they modified their ratings toward a higher value than they otherwise would. This potential Hawthorne effect would be a limitation for digital health not identified in a recent study [44], as HRSs are only starting to be used in health care. Another reason may be that the system did not allow sufficient granularity for the users to vote; the participants were only afforded three rating options. A wider spread of rating options may have contributed to the differentiation of message relevance to a greater extent (for instance, separating fair, good, and entirely accurate messages). The statistical analysis did not find a significant difference in the precision achieved between participants who were successful in smoking cessation and those who were not successful. Hence, the achieved precision over time cannot be considered a predictor of smoking cessation, as the HRS recommended motivational messages equally well for all participants.

The time-to-open metric results showed that the HRS was not able to predict the best time to send a message for a decrease in the time between the message being sent and the user opening the message. This may be due to several factors influencing participant behavior other than the time they receive a message, which were not considered in the HRS. To improve the time, variables, such as the position of the phone, location, last activity, and cessation day, should be considered for inclusion as parameters in the HRS.

The generated engagement by the system showed statistical significance (P=.04) between the cessation and no cessation subgroups after the first 6 months of the intervention, favoring a higher engagement for those who managed to stop smoking. This is in line with the intuitive assumption that participants who engaged with the system to a greater extent received greater benefit and consequently were less likely to relapse. However, these differences were not significant at 9, 11, and 12 months owing to a decrease in the number of participants to such an extent at these points that the statistical power was limited.

As not all participants in the IG completed the perceived quality questionnaire and the questionnaire was anonymous, it is impossible to ensure that any conclusion derived from the collected results is not biased. However, the results showed a clearly positive perception of the quality of the system. Reasonably, only those who managed to stop smoking and presumably benefited from the system provided answers to the questionnaire, as they would have been more motivated to respond. On the other hand, participants who dropped out, probably due to relapse, would have not been interested in completing the questionnaire. However, it could also be argued that those who did not manage to stop smoking would have been willing to provide negative feedback on the system in the anonymized questionnaire. Nevertheless, the positive trend of the answers favors the first scenario. For similar research practices in the future, we suggest following a design that forces participants to respond to the questionnaire but still preserves their anonymity. This could be achieved by including the questionnaire in a previous stage of the trial to avoid an effect by the high dropout rate associated with digital interventions for smoking cessation.

The features of the presented AI-based digital therapeutic solution enabled the system to effectively support smoking cessation by providing support and advice for facilitating abstinence, enhance motivation, and clearly show a benefit [45]. The HRS was well perceived by the participants. Over time, it identified the most relevant motivational messages to send to each participant, and those who were engaged with the system to a great extent managed to successfully stop smoking by the end of the intervention. Therefore, this digital therapeutic solution may alleviate the known drawbacks of intensive complementary behavioral interventions for smoking cessation, which, despite their benefits, require extensive availability of well-trained professionals, leading to limited scalability, poor accessibility, and smoker resignation owing to long waiting times [46].

Limitations

In previous similar studies, the dropout rate was usually very high, reflecting the difficulty of smoking cessation and high relapse rate [3]. In this study, the dropout rates were 57.5% and 62.5% in the IG and CG, respectively. A high dropout rate may bias the results; however, the dropout rate was similar between the groups, minimizing the possibility of bias. The efficacy of the intervention was verified by per-protocol and intention-to-treat analyses, providing a more realistic picture while reducing the importance of dropout or full treatment compliance. Moreover, an analysis of the characteristics of the dropout and no dropout categories within each group was carry out, and the same pattern was obtained between both groups (Multimedia Appendix 3). These findings suggest the potential benefit of the intervention, as participants did not experience abnormal adverse effects and were more likely to be abstinent after 1 year. Nevertheless, further exploration of its efficacy is needed for validation in other cultures, and studies should include a larger sample size and real-world data.

We found some random inconsistent values in the entries of the registered time, including duplicate entries and negative time values. The presented results correspond to filtered data (removing such values). Removal of these values reduced data quality for the metric, and the findings may not accurately represent the entire participant group. These inconsistent values may have been registered owing to glitches in communication between the smartphone app and the tracking system on the server.

Metrics for engagement at the aggregated level could not be provided as anticipated in the technical protocol because the software service used to track user data is not able to retrieve information from data older than 2 years. This information was not listed in the features of the service when it was selected for the purposes of the study. We encourage future researchers to use in-house user data tracking services to avoid relying on third-party software.

Comparison With Prior Work

Previous studies have explored the use of mobile apps supporting smoking cessation. The SmartQuit study tested the feasibility, acceptability, preliminary efficacy, and mechanism of behavioral change of an innovative smartphone-delivered acceptance and commitment therapy app for smoking cessation in 196 adult participants and reported a successful cessation rate of 13% after 2 months of follow-up [47].

In another study, the authors assessed the efficacy of an interactive smoking cessation decision–aid app in 684 adults and reported a successful cessation rate of 10.2% [48]. Additionally, an app-based mindfulness training program for smoking cessation was assessed in 143 participants, and a successful cessation rate of 11.1% was achieved [49]. A parallel, double-blind, randomized, controlled, two-arm trial compared the efficacy of an evidence-informed smartphone app for smoking cessation (Crush the Crave) to that of an evidence-informed self-help guide (On the Road to Quitting) in 1599 subjects and reported successful cessation rates of 7.8% and 9.2%, respectively [50]. In all these studies, the abstinence rate was reported at 6 months from baseline. However, they did not include pharmacological treatment as part of the intervention, which is a clear difference from the approach in our study. It is remarkable that the dropout rates for both the IG and CG identified in this study are consistent with those previously reported in the literature [50].

Research has been performed on intervention efficacy for the use of an app specifically designed to support the smoking cessation process when added to pharmacological therapy. The preliminary efficacy of an app designed to prompt smokers to engage in physical activity was assessed at 6 months of follow-up in 44 regular smokers who received the app in addition to behavioral smoking cessation counselling, physical activity promotion, and pharmacological support, and the overall abstinence rates were 36% in an intention-to-treat analysis and 53% in a complete-case analysis [51]. In this previous study, the allocation of pharmacological treatment (varenicline and nicotine replacement therapy) was not controlled by the study protocol, which hinders the comparability of the results with those of our study. Additionally, in a randomized pilot study, a 12-week course of varenicline was prescribed to both arms to assess the feasibility and acceptability of the My Mobile Advice Program (MyMAP) smoking cessation app and estimate its effects on smoking cessation and medication adherence [52]; however, its efficacy could not be evidenced owing to the small sample size (n=33).

As stated in the conclusion section of a recent systematic review [53] that included 26 studies (n=33,849) involving text messaging and app-based smoking cessation interventions, there is moderate certainty evidence that automated text message–based smoking cessation interventions result in greater smoking cessation rates as compared with minimal smoking cessation support and there is moderate certainty evidence of the benefit of text messaging interventions in addition to other types of smoking cessation support as compared with smoking cessation support alone. The evidence of the comparison of smartphone apps with less intensive support had very low certainty, and further randomized controlled trials are needed to test these interventions. In this sense, the results of this study aim to contribute to an increase in the certainty level of the available evidence in this domain.

When focusing on the assessment of recommender systems, such as the one used in this study, the scientific community has extensively concentrated on improving the performance of recommender system algorithms with different metrics [54-58], mainly in nonhealth contexts, such as leisure [59] and e-commerce [60,61]. Among these approaches, common assessments for prediction accuracy are as follows: (1) mean absolute error and root mean squared error, as well as their normalized and averaged variants to predict ratings that users would give to items when the actual recommendation ratings of the items are known for the whole test set; (2) precision, recall, and false positive rate [58,62] for prediction of the usage of the recommendations; and (3) normalized discounted cumulative gain when the system presents a large list of elements (similar to a search in Google), where we expect that the most relevant elements are shown at the top of the list; and (4) coverage of the recommendation set [63]. Other authors have proposed the use of surveys to assess recommender systems, such as ResQue [33].

Regarding the evaluation of mobile Health (mHealth) apps for behavioral change, such as the one included in this study, some authors have proposed evaluation procedures, such as the National Institute for Health and Care Excellence adaptation [64], the Mobile App Rating Scale framework [65], and the Application Usage Factor, which is defined as the logarithm of the product of the number of active users of a mobile app and the median number of daily uses of the app [66]. Despite these attempts to set a methodological framework to assess mHealth apps, they still lack standardization and comprehensiveness [67]. In a recent study, McKay et al identified that there was no available related best practice [68]. Instead, they proposed the following generic guidelines for what these types of evaluations should include: (1) assessment of the quality of health-related content; (2) review of the usability and functionality of the app; and (3) critique of the app potential with regard to behavioral change promotion.

Despite the lack of consensus on specific evaluation methods, it is well accepted that engagement is a key element for mHealth solutions to be successful [69]. Engagement has been considered a positive early indicator for behavioral change [70], as the process of behavioral change requires time. Previous studies showed that technologies supporting reaching a specific behavior health goal help people stick to the desired goal [71,72]. Further, Scherer et al [73] found a significant positive correlation between engagement and fewer dropouts in an mHealth intervention. A recent approach to provide an engagement score

for a mHealth app is the Engagement Index proposed by Taki et al [74]. To calculate the score, the authors combined the number of pages in the app that a participant visited each day, number of app accesses during the program, number of push notifications opened, elapsed time between app accesses, and subjective answers to a questionnaire. However, achieving user engagement is difficult [75], and its level could be improved [76-78]. Therefore, it seems reasonable to consider engagement as a key metric for digital health solutions. However, the definition of engagement varies among studies. For instance, Iacoviello et al [79] considered the number of times users opened an app, number of interactions they had, and number of weeks in which users had at least one interaction with the program as indicators of engagement. Owen et al [80] determined engagement by calculating several variables, such as the number of downloads and number of sessions (a session being basically opening of the app). Yet, it is common to just use opening the app as an indicator of engagement, as adopted in previous studies [40,81]. All these engagement interpretations fall under the categorization of "system usage data" proposed by Short et al [82], which is most frequently used. However, other engagement measurement approaches have been proposed, such as ecological momentary assessments, psychophysiological measurements, and qualitive methods.

In our So-Lo-Mo study, the HRS was intuitively related to user engagement, as some studies have shown that good and timely recommendations to stop smoking motivate users to read more future recommendations [74,83]. Hence, we proposed a combination of qualitative and quantitative metrics, which measure the intended purpose of the system (quality of the recommendations, engagement of the participants with the system, and participants' perception of the system). To our knowledge, this is the first assessment involving a HRS. However, in the case of smoking cessation, low engagement may not necessarily result in low impact on behavioral change, as some studies have previously shown the "gateway effect" [84] and "happy abandonment" [85]. There are several reasons for this, including the burden of the required interactions as compared with perceived outcomes or the internalization of health habits denoting that support is perceived as no longer necessary [86]. Consequently, we needed to carefully analyze the engagement of users across time points keeping in mind all these factors to extract any conclusions. Thus, our system's intended usage [82] is expected to be higher during the first weeks of the smoking cessation process and to progressively decrease until there is no engagement with the system, which will be reflected in engagement measurements.

Conclusions

The So-Lo-Mo intervention offers a promising strategy for smoking cessation. The use of this digital therapeutic solution alongside pharmacological treatment was much more efficacious in achieving tobacco abstinence as compared with pharmacological treatment alone at the 1-year follow-up. However, this intervention did not improve participant HRQoL and physical activity levels.

Analysis of the impact of the HRS showed that participants benefited from the recommendations, and those who engaged with the system were more likely to succeed in smoking cessation. Therefore, the proposed HRS had a positive impact on the participants and offered them personalized and relevant messages, although it could not determine when supportive messages should be sent to minimize the time until they are read.

Health care providers should consider incorporating this digital therapeutic solution in their usual care, as it can facilitate positive outcomes for participants willing to stop smoking.

Acknowledgments

The digital therapeutic solution used in the So-Lo-Mo study was jointly developed by Salumedia Tecnologías SLU, the Aristotle University of Thessaloniki, Servicio Andaluz de Salud, and Universidad de Sevilla based on the SmokeFreeBrain project. We acknowledge Prof Hein de Vries and Dr Francine Schneider who provided guidance in the data analysis of the technical outcomes. This research was funded by the H2020 European Commission research and innovation program (grant agreement 681120) as part of the SmokeFreeBrain project (www.smokefreebrain.eu).

Conflicts of Interest

SHF is a product manager at Salumedia Tecnologías SLU, the company with exploitation rights to the digital therapeutic solution used in the So-Lo-Mo study (DigiQuit). He contributed to technical data extraction for the precision, time-to-open the messages, and perceived quality metrics and their analysis. He was not involved in the clinical trial design, execution, or analysis of the resulting clinical data. Some institutions of all the other authors (University of Seville, the Aristotle University of Thessaloniki, and the Servicio Andaluz de Salud as a legal representative institution for the Virgen del Rocío University Hospital) signed an exploitation agreement with Salumedia Tecnologías SLU to benefit from DigiQuit commercialization. This agreement was elaborated and signed before publishing the present results but after the trial was finished and its results were generated.

Multimedia Appendix 1
Incomplete data analysis.
[DOC File , 81 KB - mhealth_v8i4e17530_app1.doc]

Multimedia Appendix 2

Health recommender system performance analysis, Precision_p.
[DOCX File , 42 KB - mhealth_v8i4e17530_app2.docx]

Multimedia Appendix 3
Characteristics of the dropout and non-dropout within each group.
[DOC File , 126 KB - mhealth_v8i4e17530_app3.doc]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3306 KB - mhealth_v8i4e17530_app4.pdf]

References

1. WHO. WHO Global Report On Trends In Prevalence Of Tobacco Smoking, 2015. Geneva: World Health Organization; 2015.
2. Goodchild M, Nargis N, Tursan d'Espaignet E. Global economic cost of smoking-attributable diseases. Tob Control 2017 Jan 30;27(1):58-64. [doi: 10.1136/tobaccocontrol-2016-053305]
3. Cahill K, Stevens S, Perera R, Lancaster T. Pharmacological interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2013 May 31(5):CD009329. [doi: 10.1002/14651858.CD009329.pub2] [Medline: 23728690]
4. Reid RD, Pritchard G, Walker K, Aitken D, Mullen K, Pipe AL. Managing smoking cessation. CMAJ 2016 Dec 06;188(17-18):E484-E492 [FREE Full text] [doi: 10.1503/cmaj.151510] [Medline: 27698200]
5. Cahill K, Lindson-Hawley N, Thomas K, Fanshawe T, Lancaster T. Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst Rev 2016 May 09(5):CD006103 [FREE Full text] [doi: 10.1002/14651858.CD006103.pub7] [Medline: 27158893]
6. Condon LA, Coulson NS. Chapter 2 - Designing and Delivering Interventions for Health Behavior Change in Adolescents Using Multitechnology Systems: From Identification of Target Behaviors to Implementation. In: Behavior Change Research and Theory. Cambirdge, Massachusetts: Elsevier; 2017:27-45.
7. Elbert SP, Dijkstra A, Oenema A. A Mobile Phone App Intervention Targeting Fruit and Vegetable Consumption: The Efficacy of Textual and Auditory Tailored Health Information Tested in a Randomized Controlled Trial. J Med Internet Res 2016 Jun 10;18(6):e147 [FREE Full text] [doi: 10.2196/jmir.5056] [Medline: 27287823]
8. 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]
9. Teasdale N, Elhussein A, Butcher F, Piernas C, Cowburn G, Hartmann-Boyce J, et al. Systematic review and meta-analysis of remotely delivered interventions using self-monitoring or tailored feedback to change dietary behavior. Am J Clin Nutr 2018 Feb 01;107(2):247-256 [FREE Full text] [doi: 10.1093/ajcn/nqx048] [Medline: 29529158]
10. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
11. Brug J, Glanz K, Van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug 30;25(4):517-531. [doi: 10.1177/109019819802500409] [Medline: 9690108]
12. Dijkstra A, De Vries H, Roijackers J, van Breukelen G. Tailoring information to enhance quitting in smokers with low motivation to quit: Three basic efficacy questions. Health Psychology 1998;17(6):513-519. [doi: 10.1037/0278-6133.17.6.513]
13. Dijkstra A. The Psychology of Tailoring-Ingredients in Computer-Tailored Persuasion. Social Pers Psych Compass 2008 Mar;2(2):765-784. [doi: 10.1111/j.1751-9004.2008.00081.x]
14. Enwald HP, Huotari MA. Preventing the obesity epidemic by second generation tailored health communication: an interdisciplinary review. J Med Internet Res 2010 Jun 28;12(2):e24 [FREE Full text] [doi: 10.2196/jmir.1409] [Medline: 20584698]
15. Kamel Ghalibaf A, Nazari E, Gholian-Aval M, Tara M. Comprehensive overview of computer-based health information tailoring: a systematic scoping review. BMJ Open 2019 Jan 09;9(1):e021022 [FREE Full text] [doi: 10.1136/bmjopen-2017-021022] [Medline: 30782671]
16. Kreuter M, Farrell D, Olevitch L, Brennan L. Customizing communication with computer technology. In: Routledge , editor. Tailored health messages. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc; 2000.
17. Cupertino AP, Cartujano-Barrera F, Ramírez M, Rodríguez-Bolaños R, Thrasher JF, Pérez-Rubio G, et al. A Mobile Smoking Cessation Intervention for Mexico (Vive sin Tabaco... ¡Decídete!): Single-Arm Pilot Study. JMIR Mhealth Uhealth 2019 Apr 25;7(4):e12482 [FREE Full text] [doi: 10.2196/12482] [Medline: 31021326]
18. Bamidis P, Paraskevopoulos E, Konstantinidis E, Spachos D, Billis A. Multimodal e-health services for smoking cessation and public health: the smokefreebrain project approach. In: Proceedings of the 16th World Congress on Medical and Health

Informatics. 2017 Presented at: MEDINFO 2017: Precision Healthcare through Informatics; August 2017; Hangzhou p. 5-9. [doi: 10.3233/978-1-61499-830-3-5]
19. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Aug 26;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
20. Hors-Fraile S, Nuñes Benjumea FJ, Carrasco Hernández L, Ruiz FO, Fernandez-Luque L. Design of two combined health recommender systems for tailoring messages in a smoking cessation app. In: arXiv preprint. 2016 Presented at: International Workshop on Engendering Health with RecSys co-located with ACM RecSys; 2016; Boston, Massachusetts p. 160807192 URL: https://arxiv.org/abs/1608.07192
21. Hors-Fraile S, de Vries H, Malwade S, Luna-Perejon F, Amaya C, Civit A, et al. Opening the Black Box: Explaining the Process of Basing a Health Recommender System on the I-Change Behavioral Change Model. IEEE Access 2019;7:176525-176540. [doi: 10.1109/access.2019.2957696]
22. Hors-Fraile S, Rivera-Romero O, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit-Balcells A, et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: A scoping review. Int J Med Inform 2018 Jun;114:143-155. [doi: 10.1016/j.ijmedinf.2017.12.018] [Medline: 29331276]
23. Jódar-Sánchez F, Carrasco Hernández L, Núñez-Benjumea FJ, Mesa González MA, Moreno Conde J, Parra Calderón CL, et al. Using the Social-Local-Mobile App for Smoking Cessation in the SmokeFreeBrain Project: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 Dec 06;7(12):e12464 [FREE Full text] [doi: 10.2196/12464] [Medline: 30522992]
24. Hors-Fraile S, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit A, Spachos D, et al. Tailoring motivational health messages for smoking cessation using an mHealth recommender system integrated with an electronic health record: a study protocol. BMC Public Health 2018 Jun 05;18(1):698 [FREE Full text] [doi: 10.1186/s12889-018-5612-5] [Medline: 29871595]
25. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
26. Richmond RL, Kehoe LA, Webster IW. Multivariate models for predicting abstention following intervention to stop smoking by general practitioners. Addiction 1993 Aug;88(8):1127-1135. [doi: 10.1111/j.1360-0443.1993.tb02132.x] [Medline: 8401167]
27. Pearce MS, Hayes L, Newcastle Heart Project, Newcastle Thousand Families Study. Self-reported smoking status and exhaled carbon monoxide: results from two population-based epidemiologic studies in the North of England. Chest 2005 Sep;128(3):1233-1238. [doi: 10.1378/chest.128.3.1233] [Medline: 16162711]
28. Pérez-Trullén A, Bartolomé CB, Banegas JR. [New perspectives in the diagnosis and evolution of the consumption of tobacco: markers of susceptibility and damage]. Med Clin (Barc) 2006 Apr 29;126(16):628-631. [doi: 10.1157/13087722] [Medline: 16759556]
29. Vilagut G, Ferrer M, Rajmil L, Rebollo P, Permanyer-Miralda G, Quintana JM, et al. [The Spanish version of the Short Form 36 Health Survey: a decade of experience and new developments]. Gac Sanit 2005 Apr;19(2):135-150 [FREE Full text] [doi: 10.1157/13074369] [Medline: 15860162]
30. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling Data Quality Issues to Estimate the Spanish EQ-5D-5L Value Set Using a Hybrid Interval Regression Approach. Value Health 2018 May;21(5):596-604 [FREE Full text] [doi: 10.1016/j.jval.2017.10.023] [Medline: 29753358]
31. Booth M. Assessment of physical activity: an international perspective. Res Q Exerc Sport 2000 Jun 13;71 Suppl 2(sup2):114-120. [doi: 10.1080/02701367.2000.11082794] [Medline: 25680021]
32. Cavelaars M, Rousseau J, Parlayan C, de Ridder S, Verburg A, Ross R, et al. OpenClinica. J Clin Bioinformatics 2015;5(Suppl 1):S2. [doi: 10.1186/2043-9113-5-s1-s2]
33. Pu P, Chen L, Hu R, editors. A user-centric evaluation framework for recommender systems. : ACM Digital Library; 2011 Oct Presented at: RecSys '11: Fifth ACM Conference on Recommender Systems; October 2011; Chicago, Illinois p. 157-164 URL: https://dl.acm.org/doi/10.1145/2043932.2043962 [doi: 10.1145/2043932.2043962]
34. de Vries H. An Integrated Approach for Understanding Health Behavior; The I-Change Model as an Example. PBSIJ 2017 Mar 9;2(2). [doi: 10.19080/pbsij.2017.02.555585]
35. Jamshidian M, Jalal S. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. Psychometrika 2010 Dec 3;75(4):649-674 [FREE Full text] [doi: 10.1007/s11336-010-9175-3] [Medline: 21720450]
36. Breiman L. Random forests. Machine learning 2001;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
37. Fox J, Monette G. Generalized Collinearity Diagnostics. Journal of the American Statistical Association 1992 Mar;87(417):178-183. [doi: 10.1080/01621459.1992.10475190]
38. Garrison KA, Pal P, Rojiani R, Dallery J, O'Malley SS, Brewer JA. A randomized controlled trial of smartphone-based mindfulness training for smoking cessation: a study protocol. BMC Psychiatry 2015 Apr 14;15(1):83 [FREE Full text] [doi: 10.1186/s12888-015-0468-z] [Medline: 25884648]
39. 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 14;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]

40. 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 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
41. Germeroth LJ, Levine MD. Postcessation weight gain concern as a barrier to smoking cessation: Assessment considerations and future directions. Addict Behav 2018 Jan;76:250-257 [FREE Full text] [doi: 10.1016/j.addbeh.2017.08.022] [Medline: 28865363]
42. Goldenberg M, Danovitch I, IsHak WW. Quality of life and smoking. Am J Addict 2014 Sep 25;23(6):540-562. [doi: 10.1111/j.1521-0391.2014.12148.x] [Medline: 25255868]
43. Fu S, Dao M, Wong CK, Yu W. SF-6D utility scores of smokers and ex-smokers with or without respiratory symptoms attending primary care clinics. Health Qual Life Outcomes 2019 Mar 15;17(1):48 [FREE Full text] [doi: 10.1186/s12955-019-1115-z] [Medline: 30876466]
44. Almario CV. The Effect of Digital Health Technology on Patient Care and Research. Gastroenterol Hepatol (N Y) 2017 Jul;13(7):437-439 [FREE Full text] [Medline: 28867974]
45. Kathleen FH, Young-Il K, Meifang C, Rekha R, Maria P, Rajani SS, et al. Web-Based Intervention for Transitioning Smokers From Inpatient to Outpatient Care: An RCT. Am J Prev Med 2016 Oct;51(4):620-629. [doi: 10.1016/j.amepre.2016.04.008] [Medline: 27647062]
46. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
47. Baltrunas L, Ludwig B, Peer S, Ricci F. Context relevance assessment and exploitation in mobile recommender systems. Pers Ubiquit Comput 2011 Jun 21;16(5):507-526. [doi: 10.1007/s00779-011-0417-x]
48. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
49. Garrison K, Pal P, O'Malley SS, Pittman B, Gueorguieva R, Rojiani R, et al. Craving to Quit: A Randomized Controlled Trial of Smartphone App-Based Mindfulness Training for Smoking Cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331. [doi: 10.1093/ntr/nty126] [Medline: 29917096]
50. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a Mobile Phone Intervention on Quitting Smoking in a Young Adult Population of Smokers: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]
51. Hassandra M, Lintunen T, Hagger MS, Heikkinen R, Vanhala M, Kettunen T. An mHealth App for Supporting Quitters to Manage Cigarette Cravings With Short Bouts of Physical Activity: A Randomized Pilot Feasibility and Acceptability Study. JMIR Mhealth Uhealth 2017 May 26;5(5):e74 [FREE Full text] [doi: 10.2196/mhealth.6252] [Medline: 28550004]
52. McClure JB, Anderson ML, Bradley K, An LC, Catz SL. Evaluating an Adaptive and Interactive mHealth Smoking Cessation and Medication Adherence Program: A Randomized Pilot Feasibility Study. JMIR Mhealth Uhealth 2016 Aug 03;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]
53. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
54. Cremonesi P, Koren Y, Turrin R, editors. Performance of recommender algorithms on top-n recommendation tasks. : RecSys '10; 2010 Sep Presented at: Fourth ACM Conference on Recommender Systems; 2010; Barcelona, Spain p. 39-46 URL: https://dl.acm.org/doi/10.1145/1864708.1864721 [doi: 10.1145/1864708.1864721]
55. Hernández del Olmo F, Gaudioso E. Evaluation of recommender systems: A new approach. Expert Systems with Applications 2008 Oct;35(3):790-804. [doi: 10.1016/j.eswa.2007.07.047]
56. Ricci F, Rokach L, Shapira B, Kantor P. Introduction to recommender systems handbook. In: Recommender Systems Handbook. Boston, Massachusetts: Springer; 2011:1-35.
57. Schröder G, Thiele M, Lehner W. Setting goals and choosing metrics for recommender system evaluations. In: The RecSys 2011 Workshops - Decisions@RecSys 2011 and UCERSTI-2. 2011 Oct 23 Presented at: UCERSTI2 Workshop at the 5th ACM conference on recommender systems; 2011; Chicago p. 78-85 URL: http://ceur-ws.org/Vol-811/paper12.pdf
58. Shani G, Gunawardana A. Evaluating recommendation systems. In: Recommender systems handbook. Boston, Massachusetts: Springer; 2011:257-297.
59. Bennett J, Lanning S. The netflix prize. In: Proceedings of KDD Cup and Workshop 2007. 2007 Aug 12 Presented at: KDD Cup and Workshop; 2007; San Jose, California URL: https://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf
60. Li S, Karahanna E. Online Recommendation Systems in a B2C E-Commerce Context: A Review and Future Directions. JAIS 2015 Feb;16(2):72-107. [doi: 10.17705/1jais.00389]
61. Schafer J, Konstan J, Riedl J. E-commerce recommendation applications. Data mining and knowledge discovery 2001 Jan;5(1-2):115-153 [FREE Full text] [doi: 10.1007/978-1-4615-1627-9_6]
62. Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst 2004 Jan;22(1):5-53. [doi: 10.1145/963770.963772]

63. Ge M, Delgado-Battenfeld C, Jannach D. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: RecSys '10: Proceedings of the fourth ACM conference on Recommender systems. 2010 Sep Presented at: The fourth ACM conference on Recommender systems; 2010; Barcelona, Spain p. 257-260 URL: https://dl.acm.org/doi/10.1145/1864708.1864761 [doi: 10.1145/1864708.1864761]
64. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. Patient Educ Couns 2016 Mar;99(3):429-435. [doi: 10.1016/j.pec.2015.10.023] [Medline: 26607787]
65. 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 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
66. Lewis TL, Wyatt JC. App Usage Factor: A Simple Metric to Compare the Population Impact of Mobile Medical Apps. J Med Internet Res 2015 Aug 19;17(8):e200 [FREE Full text] [doi: 10.2196/jmir.4284] [Medline: 26290093]
67. Grundy QH, Wang Z, Bero LA. Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
68. 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;24(1):22-30. [doi: 10.1177/1357633x16673538]
69. Singh K, Drouin K, Newmark L, Rozenblum R, Lee J, Landman A. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief, Commonw Fund Pub 2016 Feb 18;5:11 [FREE Full text] [doi: 10.15868/socialsector.25066]
70. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
71. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 May Presented at: The SIGCHI conference on human factors in computing systems; 2011; Vancouver, Canada p. 3063-3072 URL: https://dl.acm.org/doi/10.1145/1978942.1979396 [doi: 10.1145/1978942.1979396]
72. Smith W, Ploderer B, Wadley G, Webber S, Borland R. Trajectories of engagementdisengagement with a story-based smoking cessation app. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 May Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May, 2017; Denver, Colorado, USA p. 3045-3056 URL: https://dl.acm.org/doi/10.1145/3025453.3026054 [doi: 10.1145/3025453.3026054]
73. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth Engagement: Joint Models for Intensively Collected User Engagement Data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]
74. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
75. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
76. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
77. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
78. Anscombe J, Bacardit A. AT Kearney Insights. Chicago, Illinois: AT Kearney; 2011. Is Mobile the Cure for the Rising Costs of Healthcare? URL: https://www.atkearney.com/documents/10192/422267/EAXIV_1_Is_Mobile_a_Cure_for_the_Rising_Costs_of_Healthcare.pdf/7337673b-bd46-4ae5-aa85-1906f83f6a05 [accessed 2016-02-20] [WebCite Cache ID 6fRLcACdh]
79. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, A Personalized Smartphone App for Smoking Cessation: Initial Evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
80. 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]
81. 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 2017 Jan 01;170:37-42 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]

82. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
83. Okoshi T, Tsubouchi K, Taji M, Ichikawa T, Tokuda H. Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications. 2017 Mar 17 Presented at: 2017 IEEE International Conference on Pervasive Computing and Communications, PerCom 2017; March, 2017; Big Island, United States p. 100-110 URL: https://ieeexplore.ieee.org/document/7917856 [doi: 10.1109/percom.2017.7917856]
84. Schwanda V, Ibara S, Reynolds L, Cosley D. Side effects and gateway tools: advocating a broader look at evaluating persuasive systems. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Side effectsgateway tools: advocating a broader look at evaluating persuasive systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2011 May Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May, 2011; Vancouver, Canada p. 345-348 URL: https://dl.acm.org/doi/10.1145/1978942.1978991 [doi: 10.1145/1978942.1978991]
85. Clawson J, Pater J, Miller A, Mynatt E, Mamykina L. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Sep Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September, 2015; Osaka, Japan p. 647-658 URL: https://dl.acm.org/doi/10.1145/2750858.2807554 [doi: 10.1145/2750858.2807554]
86. Epstein D, Caraway M, Johnston C, Ping A, Fogarty J, Munson S. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 2016; San Jose, California, USA p. 1109-1113 URL: https://dl.acm.org/doi/10.1145/2858036.2858045 [doi: 10.1145/2858036.2858045]

Abbreviations

AI: artificial intelligence
BMI: body mass index
CG: control group
EuroQoL-5D-5L: EuroQol 5-dimension 5-level
HRS: health recommender system
IG: intervention group
IPAQ: International Physical Activity Questionnaire
ResQue: Recommender Systems Questionnaire of User Experience
RRR: relative risk ratio
SF-36: 36-item Short Form

Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by C Hao, A Vallée; comments to author 13.01.20; revised version received 03.03.20; accepted 21.03.20; published 27.04.20.

Please cite as:
Carrasco-Hernandez L, Jódar-Sánchez F, Núñez-Benjumea F, Moreno Conde J, Mesa González M, Civit-Balcells A, Hors-Fraile S, Parra-Calderón CL, Bamidis PD, Ortega-Ruiz F
A Mobile Health Solution Complementing Psychopharmacology-Supported Smoking Cessation: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e17530
URL: http://mhealth.jmir.org/2020/4/e17530/
doi:10.2196/17530
PMID:32338624

Original Paper

Comparison of a Collective Intelligence Tailored Messaging System on Smoking Cessation Between African American and White People Who Smoke: Quasi-Experimental Design

Jamie M Faro[1], PhD; Catherine S Nagawa[1*], MS; Jeroan A Allison[1*], MS, MD; Stephenie C Lemon[1*], PhD; Kathleen M Mazor[2,3*], EdD; Thomas K Houston[4*], MD; Rajani S Sadasivam[1*], PhD

[1]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[2]School of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[3]Meyers Primary Care Institute, University of Massachusetts Medical School, Worcester, MA, United States

[4]Section on General Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

[*]these authors contributed equally

Corresponding Author:
Jamie M Faro, PhD
Department of Population and Quantitative Health Sciences
University of Massachusetts Medical School
368 Plantation St
The Albert Sherman Center
Worcester, MA
United States
Phone: 1 774 455 4672
Email: jamie.faro@umassmed.edu

Abstract

Background: The Patient Experience Recommender System for Persuasive Communication Tailoring (PERSPeCT) is a machine learning recommender system with a database of messages to motivate smoking cessation. PERSPeCT uses the collective intelligence of users (ie, preferences and feedback) and demographic and smoking profiles to select motivating messages. PERSPeCT may be more beneficial for tailoring content to minority groups influenced by complex, personally relevant factors.

Objective: The objective of this study was to describe and evaluate the use of PERSPeCT in African American people who smoke compared with white people who smoke.

Methods: Using a quasi-experimental design, we compared African American people who smoke with a historical cohort of white people who smoke, who both received up to 30 emailed tailored messages over 65 days. People who smoke rated the daily message in terms of perceived influence on quitting smoking for 30 days. Our primary analysis compared daily message ratings between the two groups using a *t* test. We used a logistic model to compare 30-day cessation between the two groups and adjusted for covariates.

Results: The study included 119 people who smoke (African Americans, 55/119; whites, 64/119). At baseline, African American people who smoke were significantly more likely to report allowing smoking in the home (P=.002); all other characteristics were not significantly different between groups. Daily mean ratings were higher for African American than white people who smoke on 26 of the 30 days (P<.001). Odds of quitting as measured by 30-day cessation were significantly higher for African Americans (odds ratio 2.3, 95% CI 1.04-5.53; P=.03) and did not change after adjusting for allowing smoking at home.

Conclusions: Our study highlighted the potential of using a recommender system to personalize for African American people who smoke.

Trial Registration: ClinicalTrials.gov NCT02200432; https://clinicaltrials.gov/ct2/show/NCT02200432

International Registered Report Identifier (IRRID): RR2-10.2196/jmir.6465

(JMIR Mhealth Uhealth 2020;8(4):e18064) doi:10.2196/18064

KEYWORDS

machine learning; computer-tailored health communication; smoking cessation; health disparities

Introduction

Computer-tailored health communication (CTHC) increases the personal relevance of health messaging by matching the messages to an individual or group's characteristics [1] and can be effective in motivating behavior change [2-8]. CTHC is traditionally accomplished using rule-based approaches in which selected variables from patients' baseline profiles are matched to if-then tailoring rules to select messages for specific subsets of patients [1,9]. Instead of using rule-based approaches, companies like Amazon use a special class of machine learning systems (recommender systems) to select messages. These systems combine the collective intelligence of their users (ie, the observed and inferred preferences of users as they interact with the system) and their user profiles [10-12]. Our prior work [13] developed and tested a recommender system (Patient Experience Recommender System for Persuasive Communication Tailoring [PERSPeCT]) for smoking cessation. Essentially, the system selects messages to send that are more likely to motivate you because "smokers like you were influenced by messages like this one." As reported in our prior randomized trial [13], the PERSPeCT system outperformed the standard rule-based system in terms of daily message ratings and 30-day cessation.

Recommender systems have several advantages over rule-based approaches in health communication interventions, including the ability to continuously learn from user feedback (eg, liked product, products purchased) and enhance personal relevance [9]. Consequently, an anticipated benefit is that the recommender system can be even more beneficial for tailoring content to members of minority groups who are influenced by complex factors [9], especially as developers often have difficulty selecting variables for tailoring content for these groups. In this study, we focused on testing the PERSPeCT system with African American people who smoke. African Americans are more likely to die from smoking-related diseases than whites [14]. They are also less likely to be successful at quitting because they are less likely to seek cessation support [15-18]. African American people who smoke report more risk factors for cessation difficulties including greater nicotine dependence, more depressive symptoms, and lower readiness to quit compared with other race/ethnic groups [19]. They are less likely to receive tobacco counseling from health care providers [7] or use nicotine replacement therapy (NRT) compared with white people who smoke [20]. Very few tailoring interventions have successfully targeted African American people who smoke [21]. Thus, there is a real need for interventions that motivate cessation and are personally relevant to African American people who smoke.

This study is a pilot evaluation of our original evaluation of our recommender system in a new population, African American people who smoke. In our original experiment [NCT02200432], we recruited a general population of people who smoke, and most participants (92%) who enrolled were white. In this study, we recruited only African American people who smoke (n=55), and then compared the results with white people who smoke. Understanding the differential response to PERSPeCT between white and African American people who smoke may lead to improving PERSPeCT for African American people who smoke [15]. We compared daily message ratings, intervention engagement, perceived intervention impact, and 30-day cessation.

Methods

Study Design

We recruited African American people who smoke (n=55) to the PERSPeCT intervention. Then, we conducted a quasi-experimental comparison of the effectiveness of PERSPeCT in the African American intervention with a nonconcurrent comparison group of white people who smoke (the historical cohort of 64 white people who smoke who had received PERSPeCT messages in a prior trial) [13]. The data collection procedure of this study mirrored the original trial. The African American person who smokes intervention was conducted between April 2017 and November 2017; whereas the historical control data were collected between October 2014 and January 2015. This study was approved by the University of Massachusetts medical school institutional review board.

Intervention: Patient Experience Recommender System for Persuasive Communication Tailoring

The description of the recommender system is detailed elsewhere [13,22,23]. Briefly, the system includes a messages database and machine learning algorithm. The message database includes 261 expert or peer written messages [24]. Expert-written messages were composed through an iterative group review process guided by theoretical frameworks and existing smoking cessation guidelines [25]. Peer-written messages were composed by current and former people who smoke responding to an online survey. Messages included motivational content such as reasons to quit and tips and strategies to support a quit attempt, such as substitution and distraction and use of NRT.

The system sent one message daily for 30 days. Messages were delivered to the person who smokes's email address. At enrollment, we explained to people who smoke how the system worked. Each day, the participant was asked to rate the message by replying with a rating from 1 (strongly disagree) to 5 (strongly agree). Note that ratings are not required for the system to send more messages, however, the more ratings the system receives, the more personalized the daily messages can become. The daily messaging system was supported by a website with additional information. The website included functions designed to support cessation induction and maintenance such as information on smoking risks, tips on communicating with family members, and a library of cessation materials.

To train the machine learning artificial intelligence of PERSPeCT, we used the demographic and smoking behavior characteristics of previous participants (current or former people

who smoke) and their message ratings. These participants generated 16,920 ratings of 261 messages. We comparatively tested the classical algorithms k-nearest neighbors, probabilistic matrix factorization, Bayesian probabilistic matrix factorization [BPMF], collective matrix factorization, and Bayesian collective matrix factorization to identify one that provided maximal prediction accuracy (ie, we evaluated the ability of the algorithms to generalize ratings to nontraining users). We used a strong generalization protocol that involved completely separating test users from train users, learning a model using all the train users' ratings, freezing all non–user-specific parameters, and finally training the user-specific parameters on a subset of each test user's observed ratings [13]. To implement this protocol, users were randomly divided into 5 folds, and we then generated 3 random training and validation sets for each test fold. We further divided each test user's ratings into 5 folds. To evaluate each method's performance given varying levels of information about a test user, we evaluated all methods with 5, 10, and 16 of each test user's ratings available for inference and learning of user-specific parameters. Each test user had a constant set of 4 test ratings per test fold. The validation sets were used to set the hyperparameters of each method (eg, k in k-nearest neighbors). An exhaustive grid search was used, and the hyperparameter ranges were iteratively extended to ensure that no selected hyperparameter values occurred at the end points of the search intervals. In evaluating rating prediction methods, we used a range of standard performance metrics including root mean squared error (RSME), Kendall's tau-b, and normalized discounted cumulative gain. In all tests, BPMF was identified as the best single model in our evaluation and used in the development of PERSPeCT. For example, comparing the RSME metric between the different algorithms, there was a small but statistically significant gap (P=.01) between the BPMF and other algorithms as determined by a paired t test with Bonferroni correction. The BPMF model estimates a probability distribution over a joint embedding of users and items into complementary latent spaces. The rating a given user supplies for a given item is approximated by the expected value of the product of the latent user and item factor vectors representing the user-item pair, with the expectation taken over the uncertainty in embeddings. The result of this message selection is that the PERSPeCT recommender system outperformed a standard comparison system using simple rules to tailor messages to level of person who smokes's motivation [13]. The proportion of days when people who smoke agreed or strongly agreed (daily rating ≥4) that the messages influenced them to quit was significantly higher in PERSPeCT (74%) than the standard comparison system (45%; P=.02).

Recruitment

Recruitment of the African American PERSPeCT intervention and historical control participants was different. As such, we have conducted a detailed comparison of demographic and smoking behavior covariates. Our historical cohort of white people who smoke was recruited from the university hospital (2014) and affiliated outpatient clinics [13]. To recruit African American people who smoke in the this study (2017), we used the ResearchMatch.org online database to find eligible people who smoke by filtering our search based on ethnicity (must be African American), smoking status (must be a current person who smokes), and age (18 years or older). ResearchMatch is a free and secure online tool developed by Vanderbilt University and used by academic institutions across the country [26]. Eligible participants received a brief summary describing our study via email. Participants chose whether they would like to receive more information by clicking "I accept" on the email. ResearchMatch created a list of those who wanted additional information and research staff contacted those responders via email describing the study in more detail and offering to set up the initial baseline phone interview.

For both the African American people who smoke and comparison historical group, our inclusion criteria were the same (current people who smoke who were aged 18 years or older, English speaking, and had internet access). To confirm participation, all people who smoke had to complete an intake telephone call with study staff and complete an online registration. We provided incentives of up to $100 in Amazon gift cards for participation in the data collection.

Experimental Procedure

As noted, all participants were required to complete an intake telephone call and log into the supportive website to complete an online consent form and a baseline questionnaire. Once registered, each participant was emailed messages selected by the PERSPeCT recommender system and asked to rate the influence of up to 30 messages within 65 days. At the end of this period, follow-up data collection was conducted with these people who smoke.

Data Collection

During registration, people who smoke provided information on their demographic characteristics (age, sex, race, and ethnicity), smoking behaviors, prior quit attempts, and readiness to quit (I am not thinking of quitting, I am thinking of quitting, I have set a quit date, I quit today, and I have already quit) [27,28].

During the intervention, we measured ratings of messages and engagement with the supportive website. For each message, people who smoke were asked to rate message's influence on their motivation to quit smoking. Ratings were on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Also, participants' visits to the supportive website was tracked using online scripts.

At 30 days of follow-up, people who smoke reported the perceived impact of the 30-day PERSPeCT experience (intervention impact) and self-reported their smoking status. Intervention impact was assessed using 7 questions. These questions included actions that are known to help a person who smokes prepare to quit (talk to a doctor about quitting smoking, get support from those around you to help quit smoking, make a list of reasons to quit smoking, and use behavioral strategies like distraction or substitution) and those that could help a person who smokes actively quitting (use NRT like the patch or gum, set a quit date, and quit smoking) [29]. For the primary dependent variable, 30-day cessation, we asked, "Since starting the Quit Smoking Messaging System study have you stopped smoking for one day or longer because you were trying to quit?"

Statistical Analyses

The analytic plan followed the flow of data collection. We first compared data collected during the active intervention and then analyzed 30-day follow-up in the African American PERSPeCT group and the comparison white PERSPeCT group.

For each day, we created a daily rating defined as the mean of the ratings provided by all people who smoke in that group that day. We graphed this data by day and reported the percentage of days where the African Americans rated the message higher than whites. We then compared mean of daily ratings between African American and white people who smoke using a *t* test. We measured engagement with the supportive website by comparing mean number of visits to the Web-assisted tobacco intervention between African American and white people who smoke using the *t* test statistic. We dichotomized (agree or strongly agree versus other) the responses to each of the seven questions that assessed intervention impact, comparing African American and white PERSPeCT groups using the chi-square statistic. We calculated the percentage of participants who reported 30-day cessation. In the logistic regression analyses, 30-day cessation (ie, quit for at least 1 day) was considered the dependent variable (yes or no). Participant's race was the independent variable, and we adjusted for covariates that were significantly different between the two groups. Finally, we conducted a formal mediation analysis to evaluate whether difference in ratings during the intervention (comparing African Americans and whites) mediated the difference in 30-day smoking cessation.

For both the during intervention and 30-day follow-up analyses, we used Stata statistical software (StataCorp LLC). For the mediation analysis, we used the Stata medeff command and sought to quantify the effect of PERSPeCT that operates through the path of differential experience with messages (as measured by daily ratings). For the mediation analysis, we first fit a linear regression model evaluating the association of ethnicity and the mediator (daily ratings) and then a second logistic regression model with the main outcome as 30-day smoking cessation. The independent variable for the main model was ethnicity, and the mediator was daily ratings. We report the percentage of the total effect mediated by daily ratings.

Results

Patient Characteristics

African American people who smoke were significantly more likely to allow smoking in the home compared with whites (P=.002); all other characteristics were balanced between the two groups (see Table 1).

Table 1. Demographic characteristics and smoking behavior at baseline.

Characteristic	White people who smoke (n=64), n (%)	African American people who smoke (n=55), n (%)	*P* value
Gender			**.80**
Male	24 (37)	22 (40)	
Female	40 (63)	33 (60)	
Age in years			**.76**[a]
19-34	21 (33)	17 (31)	
35-44	17(27)	14 (25)	
45+	26(40)	24 (44)	
Education			**.84**
Other[b]	51 (80)	43 (78)	
Advanced college degree	13 (20)	47 (22)	
Ethnicity			**.23**
Not Hispanic	59 (97)	52 (95)	
Hispanic or Latino	2 (3)	3 (5)	
Allow smoking at home			**.002**
No	40 (63)	19 (35)	
Yes	24 (37)	36 (65)	
Ever visited a smoking cessation website			**.37**
No	52(81)	48 (87)	
Yes	12 (19)	7 (13)	
During the past 12 months, stopped smoking for one day or longer because you were trying to quit smoking			**.85**
No	36 (56)	30 (55)	
Yes	28 (44)	25 (45)	
Current smoking status			**.87**
Not actively quitting	52 (81)	44 (80)	
Actively quitting	12 (19)	11 (20)	

[a]Tested for trend using the Mantel-Haenszel method (mHodds command in Stata).

[b]Some high school, high school diploma, some college or technical school.

During Intervention

More African American people who smoke rated all 30 messages than white people who smoke (Figure 1). Using daily message rating averages, African American people who smoke rated messages (daily message rating ≥4) on all days (100% of the 30 days of messages) compared with 77% (23/30) of days for the white people who smoke (*P*<.001; Figure 2). Daily mean ratings were higher for African American than white people who smoke on 26 of the 30 days (87% [26/30] African Americans higher vs 13% [4/30] of days where whites had higher daily rating; *P*<.001). Overall, daily message ratings were higher for African American people who smoke (mean 4.27 [SD .02]; range 4.23-4.31) than white people who smoke (mean 4.10 [SD .11]; range 4.05-4.15; *P*<.001).

African American people who smoke had significantly more visits to the Web-assisted tobacco intervention compared with white people who smoke (African American mean 5.5 [SD 1.3]; range 3.0-8.0 vs white mean 1.5 [SD 0.1]; range 1.2-1.7; *P*<.001).

Figure 1. Proportion of people who smoke completing ratings assessments by study time periods.

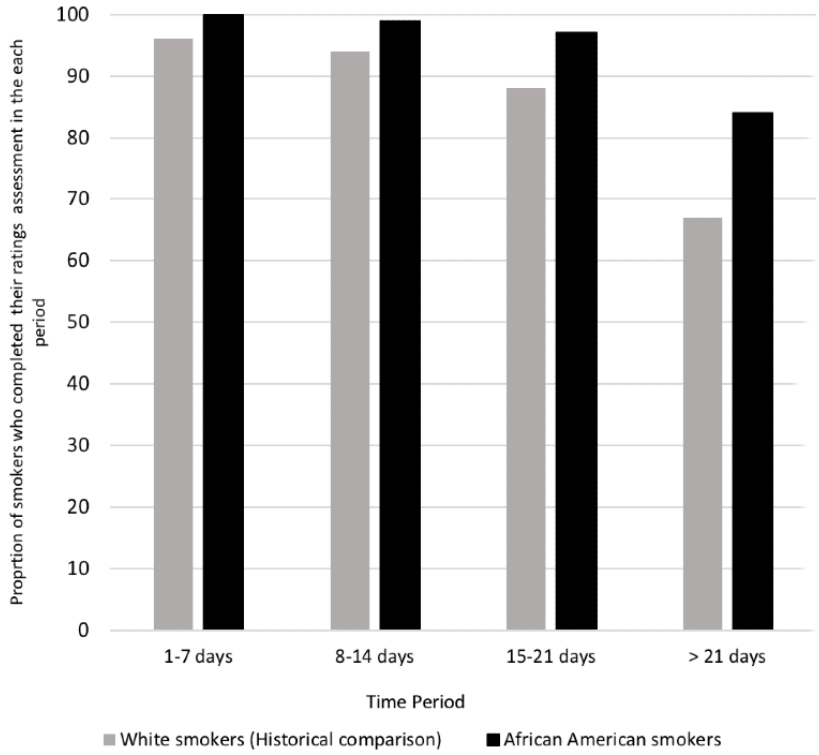
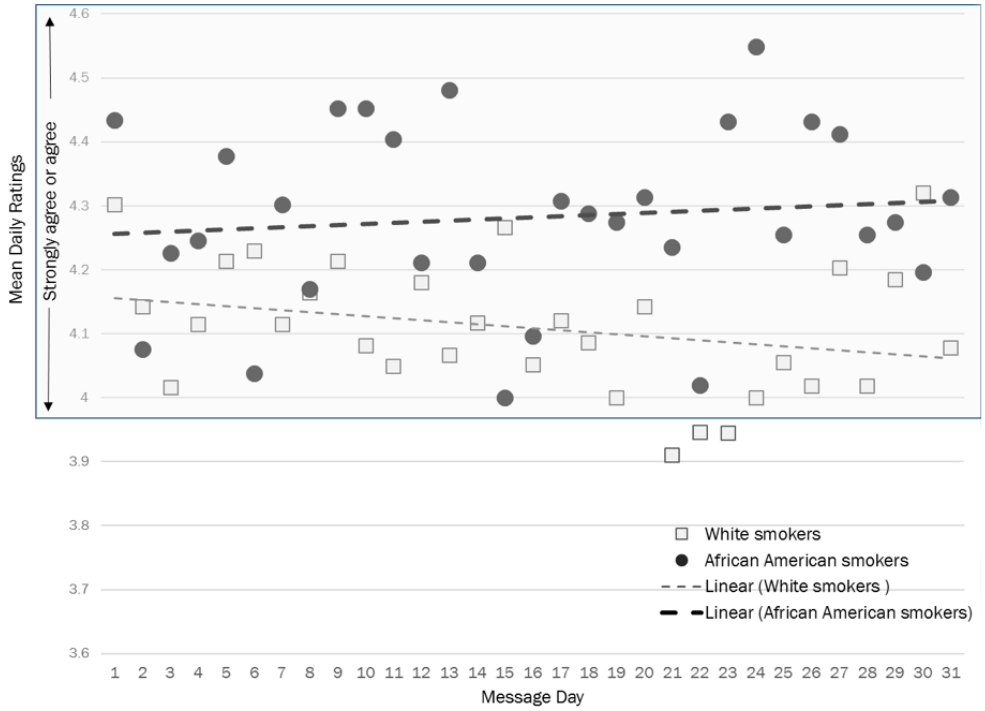


Figure 2. African American and white people who smoke mean message ratings over time.



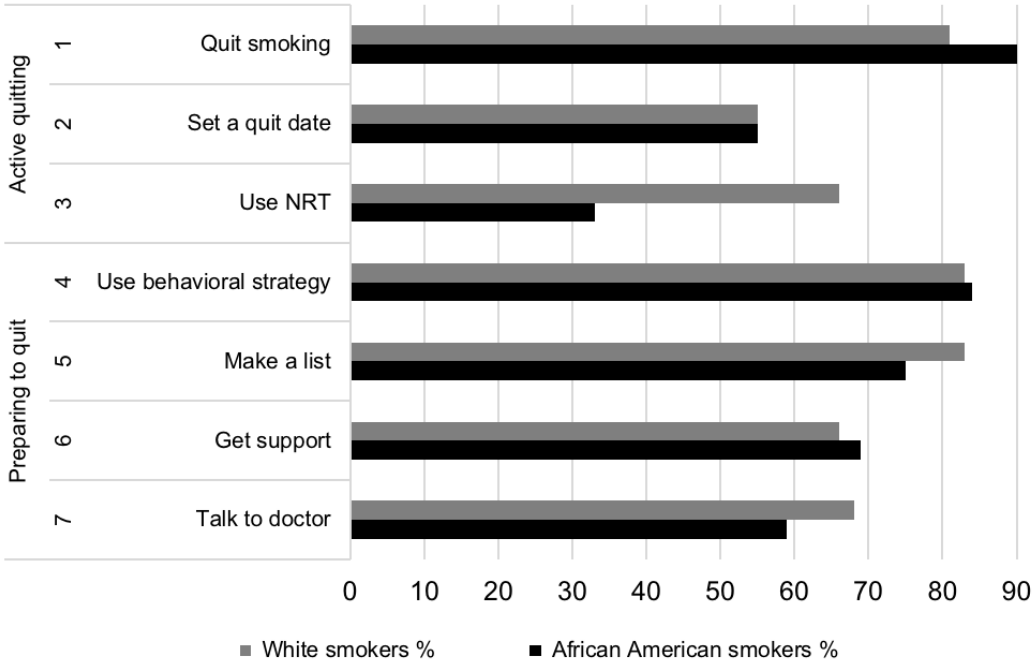
Follow-Up

Perceived influence of use of NRT was significantly lower among the African American people who smoke compared with the white poeple who smoke (66% [35/53] vs 33% [17/51], *P*=.001; Figure 3). More African American people who smoke reported a perceived influence to quit smoking than white people who smoke, but this was not significant (90% [46/51] vs 81% [43/53], *P*=.19).

African American people who smoke were significantly more likely to quit than white people who smoke (African American 59% [30/51], white 38% [19/50], *P*=.03). In the unadjusted logistic model, compared with white people who smoke, odds of quitting were significantly higher for African Americans (odds ratio [OR] 2.3; 95% CI 1.04-5.53). The result did not change after adjusting for allowing smoking at home (OR 2.4; 95% CI 1.04-5.53). In the secondary mediation analysis, we again found that daily ratings were higher among African

Americans versus whites (linear regression beta=0.49, *P*=.002), and experience with the PERSPeCT messages (measured by daily ratings) mediated 42% of the total effect of ethnicity on 30-day smoking cessation.

Figure 3. Perceived influence on quitting strategies between African-American and white people who smoke.



Discussion

Principal Findings

Consistently across the 30 days, message ratings of African American people who smoke were higher than those of white people who smoke. Further, mean African American people who smoke ratings started higher than white people who smoke, and ratings within this group increased over time. On the contrary, mean white people who smoke ratings started lower than African American people who smoke, and ratings within this group decreased over time. Self-reported perceived influence of the intervention on use of NRT was lower among African American people who smoke than white people who smoke. Intervention engagement was higher among African Americans people who smoke compared with white people who smoke. African American people who smoke were significantly more likely to report 30-day cessation as compared with white people who smoke, and this different was mediated by experience with PERSPeCT.

There may be several reasons for the higher ratings of African American people who smoke than white people who smoek in the study. The ability to influence the message a perosn who smokes receives (by rating the message) may have provided an enhanced feeling of control over the intervention. African American people who smoke may have been more attracted to this increased autonomy, the extent to which a behavior or course of action is personally endorsed and engaged [30], than white people who smoke. A potential advantage of using a recommender system is that the machine learning algorithms can learn from user feedback and improve the selection of messages over time [9]. In our study, only the ratings of African American people who smoke increased over time, whereas for white people who smoke the ratings decreased over time. There may have been a symbiotic relationship between intervention engagement and message selection. More African American people who smoke rated all messages than white people who smoek. The increased feedback may have resulted in better message selection, which then resulted in higher ratings and vice versa. African Americans may have benefited from increased engagement with the Web-assisted tobacco intervention, which made them more receptive to the messages from the system. Previous findings have shown that increasing the personalization of a message increases both the relevance and relatedness of the message to the user [1]. The increased engagement with the supportive website may also reflect higher motivation among the African American people who smoke than the white people who smoke resulting in higher ratings.

The lower perceived intervention impact for use of NRT among African American compared with white people who smoke may have highlighted a potential unintended consequence of using recommender systems. Evidence shows African American people who smoke are less likely to successfully quit than white people who smoke [15]. A primary reason for this is the low use of NRT by African American people who smoke [31]. Several reasons have been identified for the low use of NRT by African American people who smoke, including misinformation about the safety and addictive potential of NRT [32], as well as possible interactions with other medications. If the system targeting African American people who smoke is based primarily on user feedback (as PERSPeCT is), such a system may never select messages that address the use of NRT if the user provides low rating for these messages. This calls for the use of hybrid systems that combine a rule-based and recommender approach for CTHC, incorporating both their strengths. Optimal strategies for developing this hybrid approach

(eg, when the message selection should be rule-based versus a recommender approach) needs to be tested.

The 30-day cessation data highlights PERSPeCT's potential as a cessation intervention for African American people who smoke. Culturally tailored materials have been shown to enhance the effectiveness of previous trials [33]. Although we did not alter messages to the target population, we hoped that the use of individual feedback from African American users may have improved the system's ability to select messages that reflect values and practices specific to other African Americans people who smoke. As noted above, African American people who smoke were more engaged with the ratings assessment, which may have resulted in better message selection than in white people who smoke. The difference in cessation results was despite African American people who smokes' lower perceived influence for use of NRT, a known facilitator of cessation. Also, allowing smoking in the home, a known barrier to smoking cessation, was more prevalent in the African American sample, and this makes the success of PERSPeCT even more promising. Further studies are needed to test the long-term effect of the recommender system to promote cessation.

Limitations

This study has some limitations. The difference in recruitment approaches may have influenced the results of our study. As noted, we were able to recruit white people who smoke within the local area (Central Massachusetts and surrounding areas). To recruit African Americans, we had to recruit in multiple states using ResearchMatch. However, as Table 1 noted, allowing smoking at home was the only difference between the two study groups. In the 30-day cessation analysis, we adjusted for allowing smoking at home, and the results were the same as the unadjusted model. Second, the sample size may have been insufficient to detect cessation differences that may truly exist between white and African American people who smoke. Study results should therefore be interpreted with caution as a larger study is needed to confirm our findings. Third, as appropriate for a pilot, the study only assessed 30-day cessation and did not follow the people who smoke over 6 months or 1 year. The longer term is considered an appropriate time window to assess the impact of intervention on smoking cessation [34]. Fourth, the lack of random assignment creates uncertainties in inferring that the PERSPeCT recommender system resulted in the observed differences between white and African American people who smoke. As noted previously, we used a historical comparison (white people who smoke) to compare to the participants (African American people who smoke) enrolled in this study. Systematic differences in characteristics between the two study groups may have implications for selection bias. Future studies that can effectively account for dissimilarities between the two groups are needed to make causal inferences [35]. Fifth, use of a historical cohort of white people who smoke as a comparison group poses a temporal challenge; people who smoke recruited in 2014 may not accurately represent people who smokes' perspectives in 2017 or current perspectives [36]. Therefore, observed group differences among people who smoke may be a result of time-based differences (>2 years) rather than true group differences. Despite these limitations, the results of our exploratory study indicate potential for testing of the recommender system with African American people who smoke in a larger study.

Conclusions

Few systems have been able to select messages of higher influence, increasing engagement with the system and self-reported 30-day cessation rates. Additional innovations such as using a hybrid rule and recommender approach may be needed to effectively engage African American people who smoke while also motivating the use of NRT and other effective treatments. Also, since recommender systems can learn from user feedback and adapt over time, the system might be even more effective over a longer duration (6 months or a year). Future research is needed to test the long-term effectiveness of using a recommender system CTHC approach for smoking cessation in African American people who smoke to assess the true impact of PERSPeCT.

Acknowledgments

This work was supported by the National Cancer Institute under grants K07CA172677 and T32CA172009, in addition to the Patient-Centered Outcomes Research Institute award under grant PI12-001.

Conflicts of Interest

None declared.

References

1. Kreuter M. Tailoring Health Messages: Customizing Communication With Computer Technology. Mahwah: Lawrence Erlbaum Associates; 2000.
2. Brug J, Steenhuis I, van Assema P, de Vries H. The impact of a computer-tailored nutrition intervention. Prev Med 1996;25(3):236-242. [doi: 10.1006/pmed.1996.0052] [Medline: 8781000]
3. Prochaska JO, DiClemente CC, Velicer WF, Rossi JS. Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. Health Psychol 1993 Sep;12(5):399-405. [doi: 10.1037//0278-6133.12.5.399] [Medline: 8223364]

4. Campbell MK, DeVellis BM, Strecher VJ, Ammerman AS, DeVellis RF, Sandler RS. Improving dietary behavior: the effectiveness of tailored messages in primary care settings. Am J Public Health 1994 May;84(5):783-787. [doi: 10.2105/ajph.84.5.783] [Medline: 8179049]
5. Bull FC, Kreuter MW, Scharff DP. Effects of tailored, personalized and general health messages on physical activity. Patient Educ Couns 1999 Feb;36(2):181-192. [doi: 10.1016/s0738-3991(98)00134-7] [Medline: 10223022]
6. Kreuter MW, Strecher VJ. Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. Health Educ Res 1996 Mar;11(1):97-105. [doi: 10.1093/her/11.1.97] [Medline: 10160231]
7. Marcus BH, Emmons KM, Simkin-Silverman LR, Linnan LA, Taylor ER, Bock BC, et al. Evaluation of motivationally tailored vs. standard self-help physical activity interventions at the workplace. Am J Health Promot 1998;12(4):246-253. [doi: 10.4278/0890-1171-12.4.246] [Medline: 10178617]
8. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med 2008 May;34(5):373-381 [FREE Full text] [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
9. Sadasivam RS, Cutrona SL, Kinney RL, Marlin BM, Mazor KM, Lemon SC, et al. Collective-intelligence recommender systems: advancing computer tailoring for health behavior change into the 21st century. J Med Internet Res 2016 Mar 07;18(3):e42 [FREE Full text] [doi: 10.2196/jmir.4448] [Medline: 26952574]
10. Segaran T. Programming Collective Intelligence: Building Smart Web 2.0 Applications. 1st Edition. Sebastapol: O'Reilly; 2007.
11. Vanderbilt T. The science behind the Netflix algorithms that decide what you'll watch next. 2013. URL: http://www.wired.com/2013/08/qq_netflix-algorithm/ [accessed 2016-08-08] [WebCite Cache ID 6jbtsrYLv]
12. Joseph A. Deconstructing recommender systems: how Amazon and Netflix predict your preferences and prod you to purchase. 2012. URL: https://spectrum.ieee.org/computing/software/deconstructing-recommender-systems [accessed 2020-04-06]
13. Sadasivam RS, Borglund EM, Adams R, Marlin BM, Houston TK. Impact of a collective intelligence tailored messaging system on smoking cessation: the perspect randomized experiment. J Med Internet Res 2016 Nov 08;18(11):e285 [FREE Full text] [doi: 10.2196/jmir.6465] [Medline: 27826134]
14. Schoenborn CA, Adams PF, Peregoy JA. Health behaviors of adults: United States, 2008-2010. Vital Health Stat 2013 May:257 [FREE Full text] [Medline: 25116426]
15. African Americans and Tobacco Use. Atlanta: Centers for Disease Control and Prevention; 2014. URL: http://www.cdc.gov/tobacco/disparities/african-americans/index.htm [accessed 2020-04-06]
16. Kulak JA, Cornelius ME, Fong GT, Giovino GA. Differences in quit attempts and cigarette smoking abstinence between whites and African Americans in the United States: literature review and results from the international tobacco control US Survey. Nicotine Tob Res 2016 Apr;18 Suppl 1:S79-S87 [FREE Full text] [doi: 10.1093/ntr/ntv228] [Medline: 26980868]
17. Nollen NL, Cox LS, Yu Q, Ellerbeck EF, Scheuermann TS, Benowitz NL, et al. A clinical trial to examine disparities in quitting between African-American and White adult smokers: design, accrual, and baseline characteristics. Contemp Clin Trials 2016 Mar;47:12-21 [FREE Full text] [doi: 10.1016/j.cct.2015.12.001] [Medline: 26667382]
18. Webb MS, Francis J, Hines BC, Quarles FB. Health disparities and culturally specific treatment: perspectives and expectancies of African American smokers. J Clin Psychol 2007 Jun;63(6):567-583. [doi: 10.1002/jclp.20372] [Medline: 17457845]
19. Webb Hooper M, Baker EA, McNutt MD. Racial/ethnic differences among smokers: revisited and expanded to help seekers. Nicotine Tob Res 2014 May;16(5):621-625. [doi: 10.1093/ntr/ntt206] [Medline: 24336396]
20. Fu SS, Kodl MM, Joseph AM, Hatsukami DK, Johnson EO, Breslau N, et al. Racial/ethnic disparities in the use of nicotine replacement therapy and quit ratios in lifetime smokers ages 25 to 44 years. Cancer Epidemiol Biomarkers Prev 2008 Jul;17(7):1640-1647 [FREE Full text] [doi: 10.1158/1055-9965.EPI-07-2726] [Medline: 18583471]
21. Boland VC, Stockings EA, Mattick RP, McRobbie H, Brown J, Courtney RJ. The methodological quality and effectiveness of technology-based smoking cessation interventions for disadvantaged groups: a systematic review and meta-analysis. Nicotine Tob Res 2018 Feb 07;20(3):276-285. [doi: 10.1093/ntr/ntw391] [Medline: 28034998]
22. Marlin BM, Adams RJ, Sadasivam R, Houston TK. Towards collaborative filtering recommender systems for tailored health communications. AMIA Annu Symp Proc 2013;2013:1600-1607 [FREE Full text] [Medline: 24551430]
23. Adams R. PERSPeCT: collaborative filtering for tailored health communications. Proceedings of the 8th ACM Conference on Recommender Systems 2014:329-332. [doi: 10.1145/2645710.2645768]
24. Coley HL, Sadasivam RS, Williams JH, Volkman JE, Schoenberger Y, Kohler CL, National Dental PBRNQUITPRIMO Collaborative Group. Crowdsourced peer- versus expert-written smoking-cessation messages. Am J Prev Med 2013 Nov;45(5):543-550 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.004] [Medline: 24139766]
25. Fiore C, Jaen TB, Baker TB. Treating tobacco use and dependence: 2008 update. Clinical practice guideline. Rockville: U.S. Department of Health and Human Services; 2008. URL: https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/clinicians-providers/guidelines-recommendations/tobacco/clinicians/update/treating_tobacco_use08.pdf [accessed 2020-04-07]
26. Harris PA, Scott KW, Lebo L, Hassan N, Lightner C, Pulley J. ResearchMatch: a national registry to recruit volunteers for clinical research. Acad Med 2012 Jan;87(1):66-73 [FREE Full text] [doi: 10.1097/ACM.0b013e31823ab7d2] [Medline: 22104055]

27. Centers for Disease Control and Prevention. 2010 Behavioral Risk Factor Surveillance System Survey Questionnaire URL: https://www.cdc.gov/brfss/questionnaires/pdf-ques/2010brfss.pdf [accessed 2020-04-07]
28. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. J Consult Clin Psychol 1991 Apr;59(2):295-304. [doi: 10.1037//0022-006x.59.2.295] [Medline: 2030191]
29. Cherrington A, Williams JH, Foster PP, Coley HL, Kohler C, Allison JJ, et al. Narratives to enhance smoking cessation interventions among African-American smokers, the ACCE project. BMC Res Notes 2015;8:567 [FREE Full text] [doi: 10.1186/s13104-015-1513-1] [Medline: 26467316]
30. Patten CA, Clinic M, Goggin K, Harris KJ, Richter K, Williams K, et al. Relationship of autonomy social support to quitting motivation in diverse smokers. Addict Res Theory 2016;24(6):477-482 [FREE Full text] [doi: 10.3109/16066359.2016.1170815] [Medline: 27833474]
31. US Surgeon General. Tobacco use among U.S. racial/ethnic minority groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, Hispanics. A Report of the Surgeon General. Executive summary. MMWR Recomm Rep 1998 Oct 09;47(RR-18):v-xv [FREE Full text] [Medline: 9784089]
32. Carpenter MJ, Ford ME, Cartmell K, Alberg AJ. Misperceptions of nicotine replacement therapy within racially and ethnically diverse smokers. J Natl Med Assoc 2011;103(9-10):885-894 [FREE Full text] [doi: 10.1016/s0027-9684(15)30444-2] [Medline: 22364057]
33. Webb MS. Treating tobacco dependence among African Americans: a meta-analytic review. Health Psychol 2008 May;27(3S):S271-S282. [doi: 10.1037/0278-6133.27.3(suppl.).s271] [Medline: 18979980]
34. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicotine Tob Res 2003 Feb;5(1):13-25. [Medline: 12745503]
35. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23 [FREE Full text] [doi: 10.1197/jamia.M1749] [Medline: 16221933]
36. Papageorgiou SN, Koretsi V, Jäger A. Bias from historical control groups used in orthodontic research: a meta-epidemiological study. Eur J Orthod 2017 Feb;39(1):98-105. [doi: 10.1093/ejo/cjw035] [Medline: 27129869]

Abbreviations

BPMF: Bayesian probabilistic matrix factorization
CTHC: computer-tailored health communication
NRT: nicotine replacement therapy
OR: odds ratio
PERSPeCT: Patient Experience Recommender System for Persuasive Communication Tailoring
RMSE: root mean square error

Edited by G Eysenbach; submitted 31.01.20; peer-reviewed by Y Liao; comments to author 24.02.20; revised version received 16.03.20; accepted 23.03.20; published 27.04.20.

Please cite as:
Faro JM, Nagawa CS, Allison JA, Lemon SC, Mazor KM, Houston TK, Sadasivam RS
Comparison of a Collective Intelligence Tailored Messaging System on Smoking Cessation Between African American and White People Who Smoke: Quasi-Experimental Design
JMIR Mhealth Uhealth 2020;8(4):e18064
URL: http://mhealth.jmir.org/2020/4/e18064/
doi:10.2196/18064
PMID:32338619

Original Paper

Multidimensional Cognitive Behavioral Therapy for Obesity Applied by Psychologists Using a Digital Platform: Open-Label Randomized Controlled Trial

Meelim Kim[1,2], MA; Youngin Kim[3,4], MD; Yoonjeong Go[1]; Seokoh Lee[1], BS; Myeongjin Na[2], BA; Younghee Lee[1], BS; Sungwon Choi[2*], PhD; Hyung Jin Choi[1*], MD

[1]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea

[2]Department of Psychology, Duksung Women's University, Ssangmun-Dong, Dobong-Gu, Republic of Korea

[3]Noom Inc, New York City, NY, United States

[4]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyung Jin Choi, MD
Department of Biomedical Sciences
Seoul National University College of Medicine
28 Yungun-Dong, Chongno-Gu
Seoul
Republic of Korea
Phone: 82 2 740 8204
Email: hjchoi@snu.ac.kr

Abstract

Background: Developing effective, widely useful, weight management programs is a priority in health care because obesity is a major health problem.

Objective: This study developed and investigated a new, comprehensive, multifactorial, daily, intensive, psychologist coaching program based on cognitive behavioral therapy (CBT) modules. The program was delivered via the digital health care mobile services Noom Coach and InBody.

Methods: This was an open-label, active-comparator, randomized controlled trial. A total of 70 female participants with BMI scores above 24 kg/m^2 and no clinical problems besides obesity were randomized into experimental and control groups. The experimental (ie, digital CBT) group (n=45) was connected with a therapist intervention using a digital health care service that provided daily feedback and assignments for 8 weeks. The control group (n=25) also used the digital health care service, but practiced self-care without therapist intervention. The main outcomes of this study were measured objectively at baseline, 8 weeks, and 24 weeks and included weight (kg) as well as other body compositions. Differences between groups were evaluated using independent *t* tests and a per-protocol framework.

Results: Mean weight loss at 8 weeks in the digital CBT group was significantly higher than in the control group (–3.1%, SD 4.5, vs –0.7%, SD 3.4, *P*=.04). Additionally, the proportion of subjects who attained conventional 5% weight loss from baseline in the digital CBT group was significantly higher than in the control group at 8 weeks (32% [12/38] vs 4% [1/21], *P*=.02) but not at 24 weeks. Mean fat mass reduction in the digital CBT group at 8 weeks was also significantly greater than in the control group (–6.3%, SD 8.8, vs –0.8%, SD 8.1, *P*=.02). Mean leptin and insulin resistance in the digital CBT group at 8 weeks was significantly reduced compared to the control group (–15.8%, SD 29.9, vs 7.2%, SD 35.9, *P*=.01; and –7.1%, SD 35.1, vs 14.4%, SD 41.2, *P*=.04). Emotional eating behavior (ie, mean score) measured by questionnaire (ie, the Dutch Eating Behavior Questionnaire) at 8 weeks was significantly improved compared to the control group (–2.8%, SD 34.4, vs 21.6%, SD 56.9, *P*=.048). Mean snack calorie intake in the digital CBT group during the intervention period was significantly lower than in the control group (135.9 kcal, SD 86.4, vs 208.2 kcal, SD 166.3, *P*=.02). Lastly, baseline depression, anxiety, and self-esteem levels significantly predicted long-term clinical outcomes (24 weeks), while baseline motivation significantly predicted both short-term (8 weeks) and long-term clinical outcomes.

Conclusions: These findings confirm that technology-based interventions should be multidimensional and are most effective with human feedback and support. This study is innovative in successfully developing and verifying the effects of a new CBT approach with a multidisciplinary team based on digital technologies rather than standalone technology-based interventions.

Trial Registration: ClinicalTrials.gov NCT03465306; https://clinicaltrials.gov/ct2/show/NCT03465306

(JMIR Mhealth Uhealth 2020;8(4):e14817) doi:10.2196/14817

KEYWORDS

obesity; digital health care; cognitive behavioral therapy; mobile phone

Introduction

One of the major concerns of the health care industry is to find effective and widely practical solutions for weight management, given that obesity is one of the dominant public health problems of the 21st century. It is well known that weight reduction is highly correlated with reductions in the incidence of type 2 diabetes, as well as other medical weight-related comorbidities and psychosocial issues, and that it improves the quality of life [1].

Accordingly, various types of treatments for obesity have been developed. Several drugs have been proposed as pharmacotherapy for obesity since the 1990s, but most have demonstrated a lack of efficacy and unfavorable risks [2]. Bariatric surgery is another obesity treatment that has been used for over 50 years. Because the prevalence of obesity is rapidly rising, the number of patients who believe that bariatric surgery is an effective treatment to cure their obesity is also increasing [3]. Additionally, patients may believe that surgical intervention to overcome obesity will ultimately lead to behavioral changes sustaining weight loss [3], which may increase the risk of weight regains after the surgery. To date, the most effective standard obesity treatment is weight-loss lifestyle modification based on a combination of behavioral and cognitive approaches and nutrition and physical education.

Clinical psychological treatment approaches are pivotal and involve engaging patients in lifestyle modification and motivating them to successfully lose weight with the help of a multidisciplinary team [4]. Cognitive behavioral therapy (CBT) for obesity is aimed at not only losing weight but also preventing weight regain, thereby avoiding the dissatisfactory long-term results of earlier behavioral treatments. It firmly distinguishes between weight loss and weight maintenance, allowing patients to practice effective weight-maintenance strategies (eg, avoiding unrealistic weight goals and addressing obstacles to weight maintenance) [5]. One study applied a 12-week CBT program for obese people, resulting in a 6% reduction in body fat relative to the control group [6]. Moreover, a 20-week CBT intervention involving a 10-week main program followed by a 10-week less-intensive care program significantly improved body composition and improved soft drink consumption habits compared to the control group [7].

Although cognitive behavioral programs involving weekly clinic visits are known to be the most effective treatments for obesity, they place high demands due to time, cost, distance, status of endorsement, and difficulties securing child care [8]. A previous study found that people would prefer cost-effective and time-saving methods to lose weight [9]. Researchers have thus explored alternative methods for carrying out weight-loss programs, such as television, computers, and smartphone apps, to meet individual needs and to make obesity treatment more accessible. Among these, self-monitoring via smartphone apps has shown the greatest potential to make diet tracking easier and engaging because of its convenience and accessibility [10]. Despite the use of smartphone apps for self-monitoring, a *law of attrition* in digital health interventions still holds, whereby users stop using technology-based components over time. Because the effectiveness of treatments via digital tools is closely associated with the user's extent of engagement [11], a high attrition rate is a critical issue in the assessment of the efficacy of digital intervention programs. Therefore, based on behavioral modification principles, periodic prompts that encourage healthy behaviors are one method to remind and motivate people to change their health behaviors. A systematic review of the use of technology tools to send periodic notifications about users' behavior changes found them to be more effective than nontechnological notifications or no notifications [12]. However, this review only focused on the effectiveness of digital interventions for behavior change as a whole and did not investigate how to enhance engagement with the intervention.

The goal of this study was to test a novel approach to losing weight and maintaining the new weight after participation in an intensive and comprehensive human coaching program based on CBT modules via digital tools, such as the Noom Coach app and InBody Dial. The Noom Coach app is one of the most popular smartphone apps currently available; it has received higher quality assessment scores than other smartphone apps [13]. It allows participants to log their food intake, exercise activities, and weight, and to engage in in-app group activities, read in-app articles, and interact with a human coach via in-app messages. In-app group activity lets participants communicate with other participants and share their experience of healthy lifestyle trials. In-app articles deliver practical information about healthy lifestyles written by physicians, nutritionists, and clinical psychologists. In-app messages enable participants to receive individualized feedback from human coaches based on their own records presented on the Web-based dashboard. A Web-based dashboard is provided to the coaches to monitor participants' data. InBody Dial is a body composition analyzer for the home linked to a mobile app, allowing users to conveniently measure their body composition. Furthermore, we addressed the self-sustainability of the promoted lifestyle change after the intervention. We hypothesized that individuals randomized to the digital CBT group would lose weight and

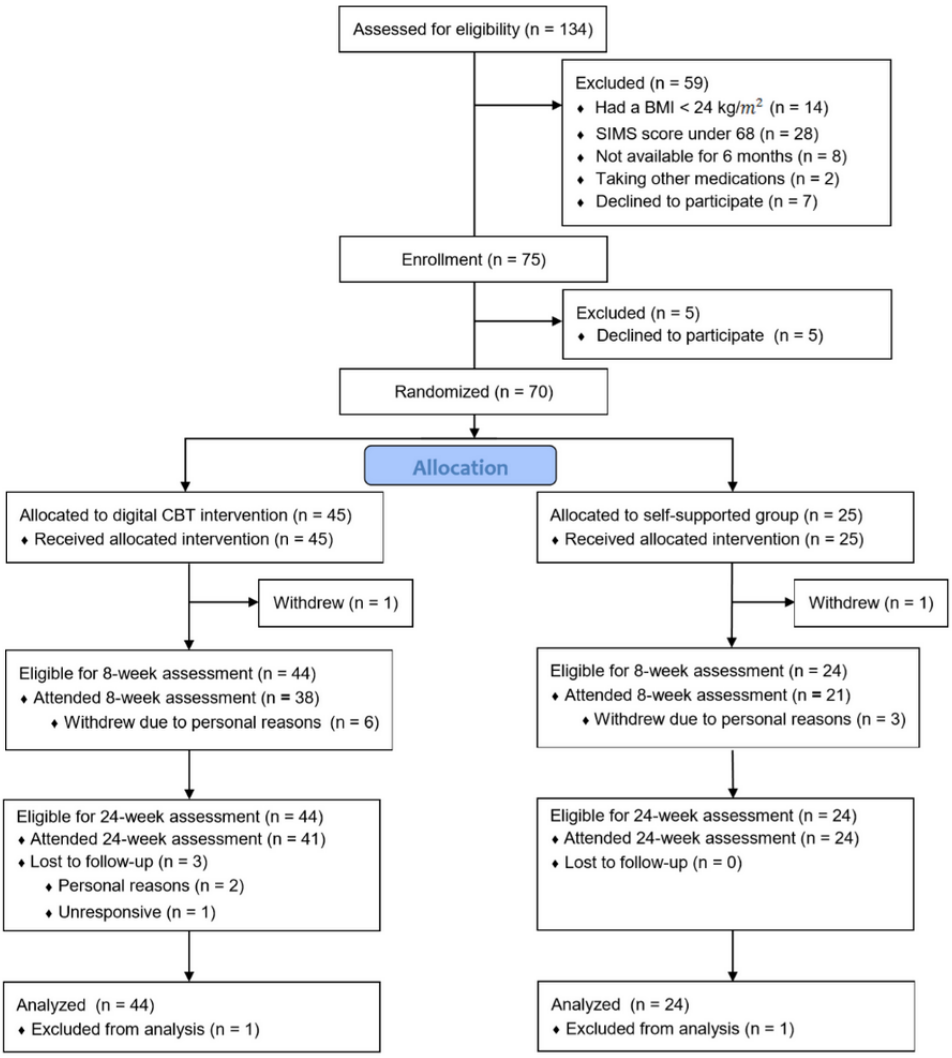
better maintain their weight loss than individuals in the control group.

Methods

Participants

A total of 70 female subjects were recruited between September and October 2017 through both online and offline boards of a university campus in Seoul, South Korea, and a social network service. Eligibility criteria included the following: 18-39 years of age, body mass index of 25-40 kg/m^2, smartphone usage, and scores in the highest 40% (ie, scores above 68 out of 112 total) on the Situational Motivation Scale (SIMS). Participants were ineligible if they had a history of major medical problems, such as diabetes, angina, or stroke; a major psychiatric disorder involving hospitalization or medication in the past; and a current or planned pregnancy within the next 6 months. The flow of participants from recruitment to final assessment at 24 weeks is shown in Figure 1.

Figure 1. Digital cognitive behavioral therapy (CBT) CONSORT (Consolidated Standards of Reporting Trials) flow diagram. SIMS: Situational Motivation Scale.



The Institutional Review Board of Seoul National University Hospital approved the study (approval number H-1707-122-872). All study participants provided written informed consent. This study was conducted to examine the clinical efficacy of the obesity digital CBT model and find factors predicting its efficacy. The study was registered with ClinicalTrials.gov (NCT03465306).

Study Design

This was an open-label, active-comparator, randomized controlled trial (RCT). Following initial screening, all participants were asked to attend an orientation session where the study was described in more detail. Written informed consent and baseline measurements were obtained in person. Blood samples were taken in the morning after overnight fasting to avoid daily variations in activities. The basics of the tutorial and log-in procedures for both the Noom app and the InBody H20B (InBody Co) body composition analyzer were demonstrated to all participants during the orientation session of the study. The Noom app was mainly used to keep a food diary and deliver messages between the therapist and participants, while the InBody H20B analyzer was used to monitor and collect the body composition data of the participants. The randomization was designed to randomly assign 75 participants in total to a control (app only) group or a digital CBT (app + human CBT) group at a ratio of 1:2 in

order to deliver a more powerful trial within resource constraints and to maximize the statistical power of predictor analysis (ie, within-group analysis) [14]. Randomization was performed by the project manager by drawing lots. The digital CBT group was given daily feedback and assignments from a psychologist, based on the CBT modules, for 8 weeks and could access the digital tools from the intervention period to the 24-week follow-up. The control group was instructed to use only a food diary without therapist intervention until the 24-week follow-up but was given the same digital tools and instruction as the digital CBT group. Thus, the control group underwent the same standard-of-care trial as the digital CBT group, except that it was asked to practice self-care. All participants were asked to visit at baseline, 8 weeks, and 24 weeks for objective measurements and completion of questionnaires, and they were each paid US $4 for attending each of the appointments. This study was conducted from September 2017 to April 2018.

Assessment

The primary outcome was change in body weight. Other measures, such as change in BMI and body fat mass, were secondary outcomes. Anthropometric measurements were assessed by the InBody H20B analyzer at baseline, 8 weeks, and 24 weeks in light street clothing and without socks and shoes. For secondary outcomes, blood samples were collected at baseline and 8 weeks after a 10-hour fast. We examined serum insulin, leptin, glucose concentrations, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transferase, total cholesterol, and triglyceride levels to assess the changes in these indices in relation to the change in body weight. The engagement criteria of the program were completing actions, such as responding to the daily assessment (responses per day), logging meals (meals per week), consuming green foods as defined by Noom [15] (logged per week), performing exercise (times per week), registering exercise time (minutes per week), recording steps taken (steps per week), logging weigh-ins (times per week), reading articles (articles per week), completing group posts (posts per week), posting group comments (comments per week), sending messages to the coach (messages per week), and liking group posts (likes per week). These criteria were used to assess the use of the app by each participant with objective measures.

Participants' situational motivation toward the weight-loss program was assessed using an adapted version of the SIMS. The SIMS typically measures four types of motivation—intrinsic motivation, identified regulation, external regulation, and amotivation—to engage in a task (ie, the weight-loss program) at a specific point in time, with four items per subscale. The SIMS has demonstrated acceptable levels of reliability and validity in past research. The Body Shape Questionnaire-8C (BSQ-8C) is a brief version of the Body Shape Questionnaire (BSQ) consisting of eight items extracted from the full version measuring the extent of psychopathology of concerns about body shape. Higher values on the BSQ indicated more body dissatisfaction. Depression was assessed using the Korean version of the Beck Depression Inventory-II (K-BDI-II) scoring system. A total score of 0-9 indicated no depression, 10-15 indicated mild depression, 16-23 indicated moderate depression, and 24-63 indicated severe depression. Anxiety was measured using the 20-item Trait Anxiety Inventory (TAI) of the State-Trait Anxiety Inventory, with higher scores indicating greater trait anxiety. The Rosenberg Self-Esteem Scale (RSES) measure of self-esteem was used in this research with a 10-item scale consisting entirely of negatively worded items. Thus, higher scores implied lower self-esteem. Eating behavior notions were measured with the Dutch Eating Behavior Questionnaire (DEBQ), which identifies three distinct psychologically based eating behaviors: restrained eating, emotional eating, and external eating. It contains 33 items, with higher scores indicating a greater tendency to present subscale behavior. The frequency of occurrence of automatic negative thoughts associated with depression was assessed by the Automatic Thoughts Questionnaire (ATQ-30). The scores ranged from 30 to 150, where higher scores indicated more frequent automatic negative thoughts. All the psychological questionnaires were in Korean.

Interventions

The intervention of this study was a multifactorial, daily-based personalized coaching program implemented by a psychologist using CBT modules via the digital platform. The digital CBT contents were based on programs proposed to clinicians [16] as a guide. We monitored and assessed various factors related to the behavior, cognition, mood, and motivation of each participant assigned to the digital CBT group.

The following were assessed every day using responses to questions and scores from the questionnaires: eating behaviors (eg, Where did you eat? What type of food did you have? How fast did you eat? and What time did you eat?), automatic thoughts (eg, What came to your mind when you were eating or thinking of food?), mood (eg, Score your mood from 0 to 100 regarding each type of negative mood: irritated, lonely, anxious, bored, and depressed), and motivation (eg, Score your status from 0 to 10 based on the following items: willingness to lose weight, importance of losing weight, assurance of losing weight, and helpfulness of this program to lose weight). Scores were used to individually track the daily patterns of the four factors—eating behaviors, automatic thoughts, mood, and motivation—and provide individualized interventions. As such, participants in the digital CBT group received daily self-report assessments in a Google survey form via text message on their phone. Participants were also instructed to log their dietary intake and physical exercise on a daily basis. Additionally, they were asked to measure their weight, BMI, and fat mass twice a week with the InBody H20B analyzer as soon as they woke up in the morning and were instructed to log their meals and physical activity by self-report on the Noom Coach app on a weekly basis.

After participants' responses to the components related to the four factors were collected, digital mobile tools collected the data to allow the therapist to securely monitor participants' progress through a Web-based dashboard. The participants received at least three individual messages from the coach every day, except on weekends and holidays, via the Noom Coach app. Furthermore, the therapist individually sent a daily report, a weekly report, and a midweek report (ie, Week 4) to the participants for the purpose of goal setting and to strengthen

their motivation. Weekly group missions were provided to the digital CBT group based on the expectation that social supports (eg, communicating needs and building positive support) would intensify the motivation. When the participants were inactive for more than 3 consecutive days or asked for thorough counseling, the therapist phoned them and conducted motivational interviews. The motivational interviews could be implemented only once a week per person. The duration of the phone call did not exceed 15 minutes.

All contents of the coaching messages, group missions, and articles were managed by a supervisor of the digital health care coach, who has a master-level degree in clinical psychology. She has trained as a behavioral therapist using CBT modules, such as self-monitoring, goal setting, problem solving, nutritional and physical activity education, stimulus control, challenging automatic thoughts, thought restructuring, and relapse prevention. Throughout the intervention, we expected the participants in the digital CBT group to experience a lifestyle change by finding a healthy pattern of living that fit each participant's context. The diagram of the digital CBT process and features of the digital platform are presented in Figure 2 and Figure 3, respectively.

Figure 2. Diagram of the digital cognitive behavioral therapy (CBT) process.

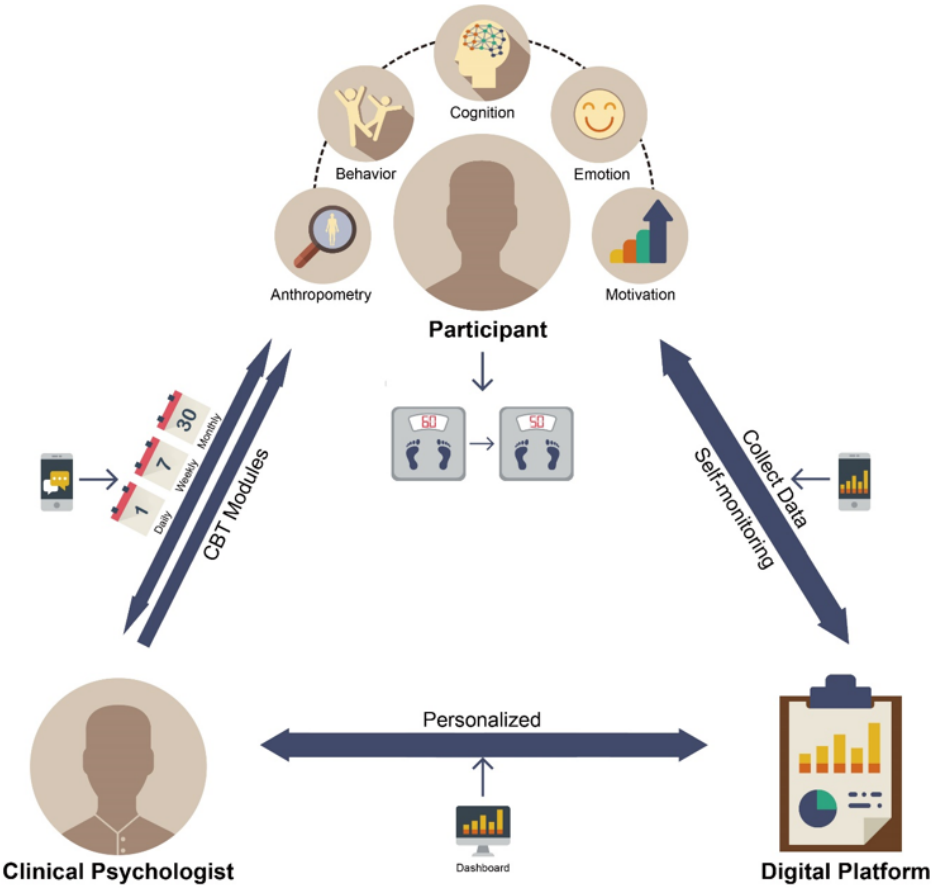
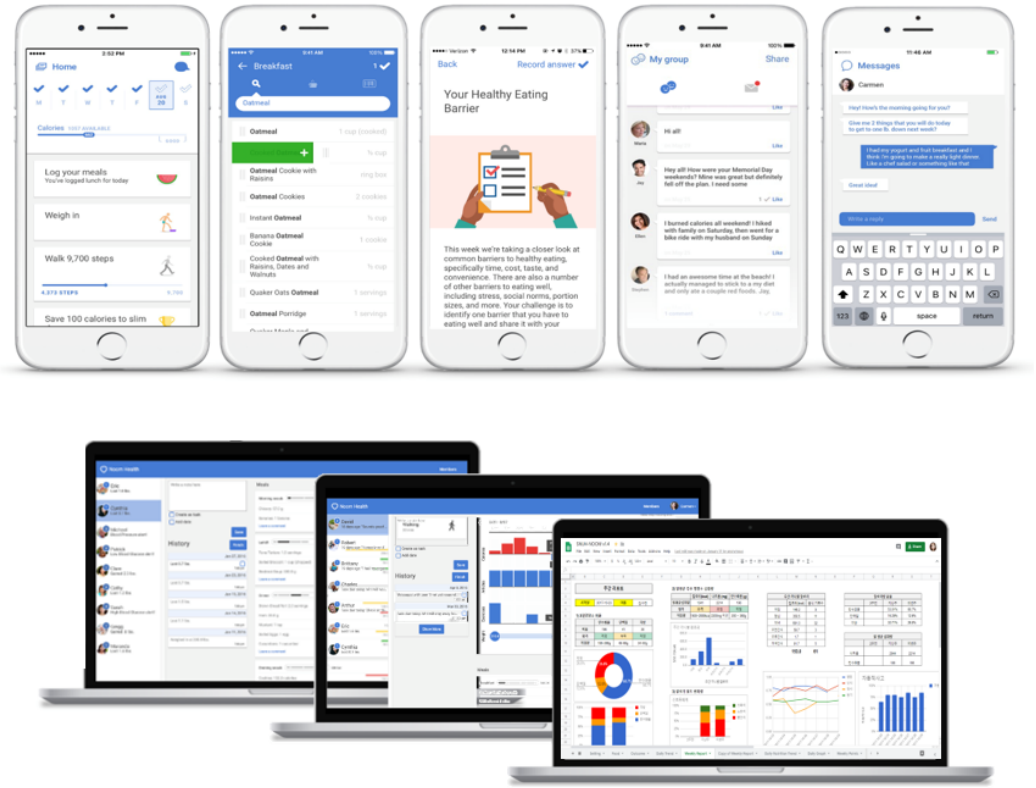


Figure 3. Screenshots of the digital platform (ie, mobile apps) for the participants (top) and screenshots of the digital platform (ie, dashboard) for the therapist (ie, clinical psychologist) (bottom).



Statistical Analysis

The sample size was selected to provide the study with a statistical power of 80% to detect clinically meaningful mean differences in weight loss of 5 kg with an SD of 7 kg in treatment effect, based on previous studies [17]. Assuming an average attrition rate of 10%, a sample of at least 70 subjects was selected. For differences in baseline characteristics, independent-sample *t* tests were used for continuous variables and a chi-square test of independence was used for categorical data assessing the demographic patterns of subjects.

We conducted the analysis following per-protocol principles. The participants who attended at either 8 or 24 weeks were included in the analysis of the applicable period without missing imputations. There were no outliers in the dataset. To investigate differences in the outcomes between the two groups, changes in the outcomes of weight, BMI, and fat mass were analyzed using an independent-sample *t* test. To investigate statistical differences between baseline and postintervention within a group, a paired *t* test was used. To detect statistical differences of the proportion within the thresholds and engagement rates between groups, a chi-square test was used. Correlation analysis using the Pearson correlation coefficient was used to investigate which variables at the baseline had a predictive role in changes in anthropometrics at 8 and 24 weeks. Receiver operating characteristic (ROC) curve analysis was undertaken to identify the optimum trade-off between sensitivity and specificity for cutoffs in weight-change distribution. For the ROC analysis in this study, we set a cutoff of 3% loss of initial body weight as a *good response* at 24 weeks for the digital CBT group data. The Youden index was used for the optimal cutoff. The results regarding the proportion of people who reached 5% weight-loss threshold are also reported to permit comparison with other previous studies. All analyses were conducted using SPSS Statistics for Windows, version 20 (IBM Corp), and statistical significance of two-tailed *P* values were set at .05. For multiple comparison correction, a threshold of $P<.001$ was used (ie, the *P* value threshold of .05 divided by 42, corresponding to two different time periods and 21 phenotypes).

Results

Overview

There were no significant differences between the randomization groups on key demographic characteristics (see Table 1). However, the DEBQ emotional eating scale (DEBQ-EM) ($P=.001$) and the DEBQ external eating scale (DEBQ-EX) ($P=.049$) scores of the two groups did differ at baseline. These differences between the groups were found after lots were drawn for the randomized control procedure. Participants had a mean age of 21.8 years (SD 3.3) and a mean BMI of 28.0 kg/m^2 (SD 3.2).

Table 1. Baseline characteristics of participants in both groups.

Characteristic	Control (ie, app only) (n=25)	Digital CBT[a] (ie, app + human CBT) (n=45)
Age (years), mean (SD)	21.0 (2.7)	22.3 (3.5)
Anthropometric measures, mean (SD)		
Weight (kg)	71.9 (7.7)	74.5 (9.0)
BMI (kg/m^2)	27.7 (2.9)	28.2 (3.4)
Fat mass (kg)	29.3 (6.0)	30.2 (6.8)
Fat percent (%)	40.5 (4.8)	40.4 (5.4)
Lean body mass (kg)	23.8 (3.3)	24.0 (2.6)
Blood measures, mean (SD)		
Fasting glucose (mg/dL)	87.0 (8.1)	87.3 (7.4)
Triglyceride (mg/dL)	92.2 (35.9)	93.2 (42.6)
Total cholesterol (mg/dL)	184.7 (24.9)	191.1 (30.4)
Alanine aminotransferase (U/L)	12.7 (6.9)	15.3 (11.9)
Aspartate aminotransferase (U/L)	17.0 (4.7)	16.9 (4.8)
Gamma-glutamyl transpeptidase (U/L)	15.3 (8.5)	21.3 (32.8)
Leptin (ng/mL)	37.5 (14.7)	42.5 (15.3)
Fasting insulin (μU/mL)	12.6 (6.1)	16.1 (9.1)
Homeostasis Model for Assessment of Insulin Resistance[b], mean (SD)	2.8 (1.5)	3.5 (2.1)
Scale or questionnaire (score), mean (SD)		
Situational Motivation Scale	77.0 (5.8)	76.1 (5.7)
Body Shape Questionnaire-8C	34.8 (8.9)	36.2 (7.5)
Beck Depression Inventory-II in Korean	14.7 (9.6)	13.6 (9.0)
Trait Anxiety Inventory	47.8 (11.0)	48.0 (10.4)
Rosenberg Self-Esteem Scale	21.9 (6.4)	19.8 (5.6)
DEBQ[c] restrained eating scale	30.6 (7.3)	29.9 (6.6)
DEBQ emotional eating scale[d]	29.1 (11.6)	38.0 (10.1)
DEBQ external eating scale[d]	32.0 (7.0)	34.9 (4.8)
Automatic Thoughts Questionnaire	57.6 (26.0)	57.2 (22.3)
Yale Food Addiction Scale	2.2 (1.7)	3.0 (1.7)
Residence status, n (%)		
Living with family	10 (40)	27 (60)
Living alone	8 (32)	8 (18)
Living with roommates	7 (28)	9 (20)
Others	0 (0)	1 (2)
Number of attempts to lose weight by different methods, n (%)		
None	0 (0)	1 (2)
Once	3 (12)	4 (9)
Twice	12 (48)	15 (33)
Three times	3 (12)	13 (29)
Four times	4 (16)	8 (18)
Five times	2 (8)	4 (9)

Characteristic	Control (ie, app only) (n=25)	Digital CBT[a] (ie, app + human CBT) (n=45)
Six times	1 (4)	0 (0)

[a]CBT: cognitive behavioral therapy.

[b]Insulin resistance = (insulin [μU/mL] × glucose [mg/dL]) / 405.

[c]DEBQ: Dutch Eating Behavior Questionnaire.

[d]There was a statistical difference between the two groups at baseline.

Primary Outcome of Weight Change and Anthropometric Outcomes

The primary outcome (ie, weight change) was assessed at two time points—immediately after lifestyle change with digital CBT (8 weeks) and at the long-term follow-up without digital CBT (24 weeks)—to investigate the self-sustaining effect of lifestyle change induced by 8 weeks of digital CBT. Of the 70 randomized participants, 65 (93%) were assessed for the primary outcome—body weight—at 24 weeks and 5 (7%) were lost to follow-up. Figures 4 and 5 represents the mean weight change along with other anthropometric measures—BMI, body fat mass, and body lean mass—at each study time point. Participants in the digital CBT group showed significant changes in mean body weight at 8 weeks compared to the control group (–3.1%, SD 4.5, vs –0.7%, SD 3.4, *P*=.04) but not at 24 weeks. The proportion of subjects who showed *good response* was 45% (17/38) in the digital CBT group and 29% (6/21) in the control group at 8 weeks (*P*=.22), while at 24 weeks it was 54% (22/41) in the digital CBT group and 42% (10/24) in the control group (*P*=.35). In addition, the number reaching the conventional 5% weight loss from the baseline in the digital CBT group was significantly higher than in the control group at 8 weeks (12/38, 32%, vs 1/21, 4%, *P*=.02) but not at 24 weeks (18/41, 44%, vs 7/24, 29%, *P*=.24). Changes in mean BMI (–3.1%, SD 4.6, vs –0.7%, SD 3.5, *P*=.04) and body fat mass (–6.3%, SD 8.8, vs –0.8%, SD 8.1, *P*=.02) of the digital CBT group were also significant compared to the control group at 8 weeks but not at 24 weeks (see Multimedia Appendix 1, Table MA1-1). Body lean mass did not significantly differ between the two groups at both 8 and 24 weeks. Examining within-group changes, only the digital CBT group achieved significant weight changes, as well as BMI and body fat mass, at both 8 and 24 weeks; the digital CBT group achieved significant changes in lean body mass at 24 weeks but not at 8 weeks (see Multimedia Appendix 1, Tables MA1-2 and MA1-3).

Figure 4. Patterns of changes in mean body weight (A), BMI (B), body fat mass (C), and lean body mass (LBM) (D). CBT: cognitive behavioral therapy. **P*<.05; ***P*<.01.

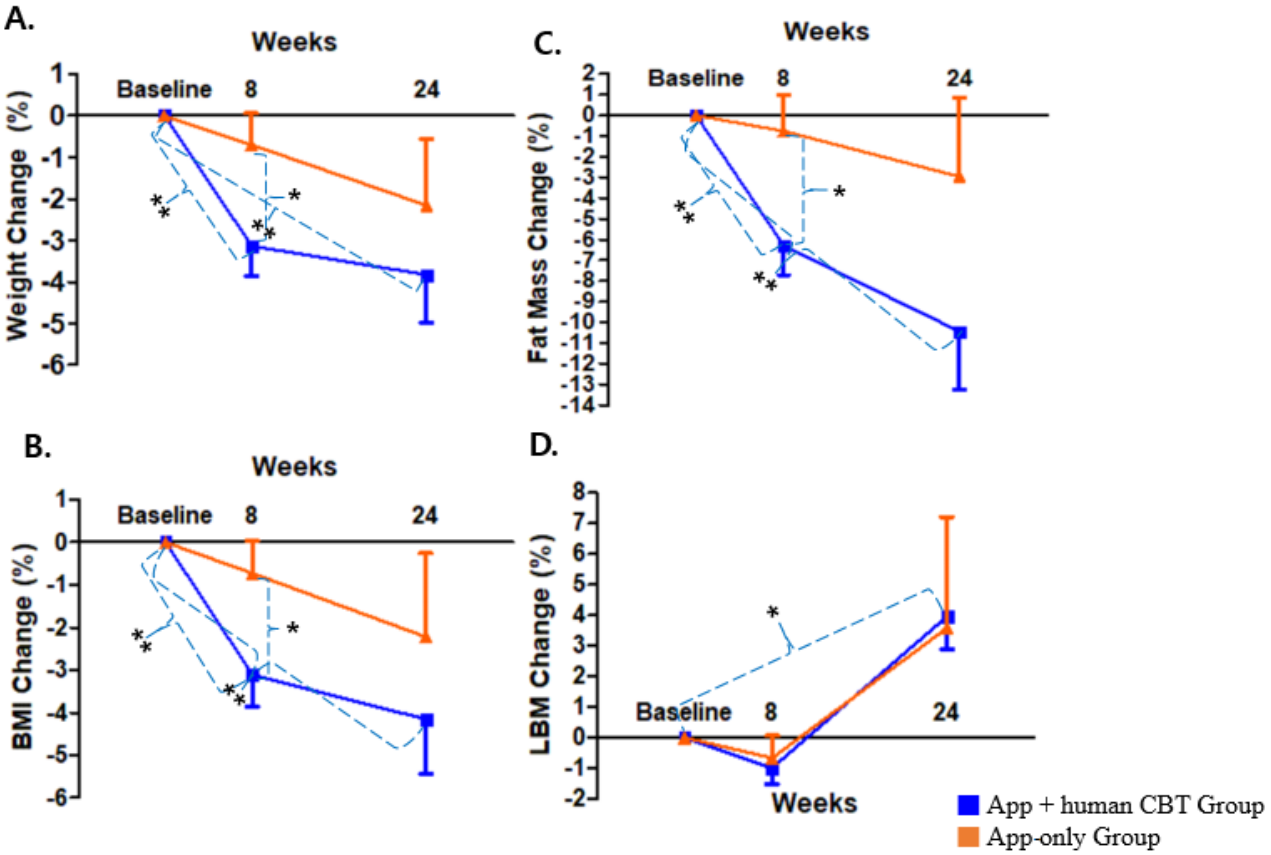


Figure 5. Weight change based on individual data from the experimental group at the 8-week follow-up (A), from the experimental group at the 24-week follow-up (B), from the control group at the 8-week follow-up (C), and from the control group at the 24-week follow-up (D). CBT: cognitive behavioral therapy.

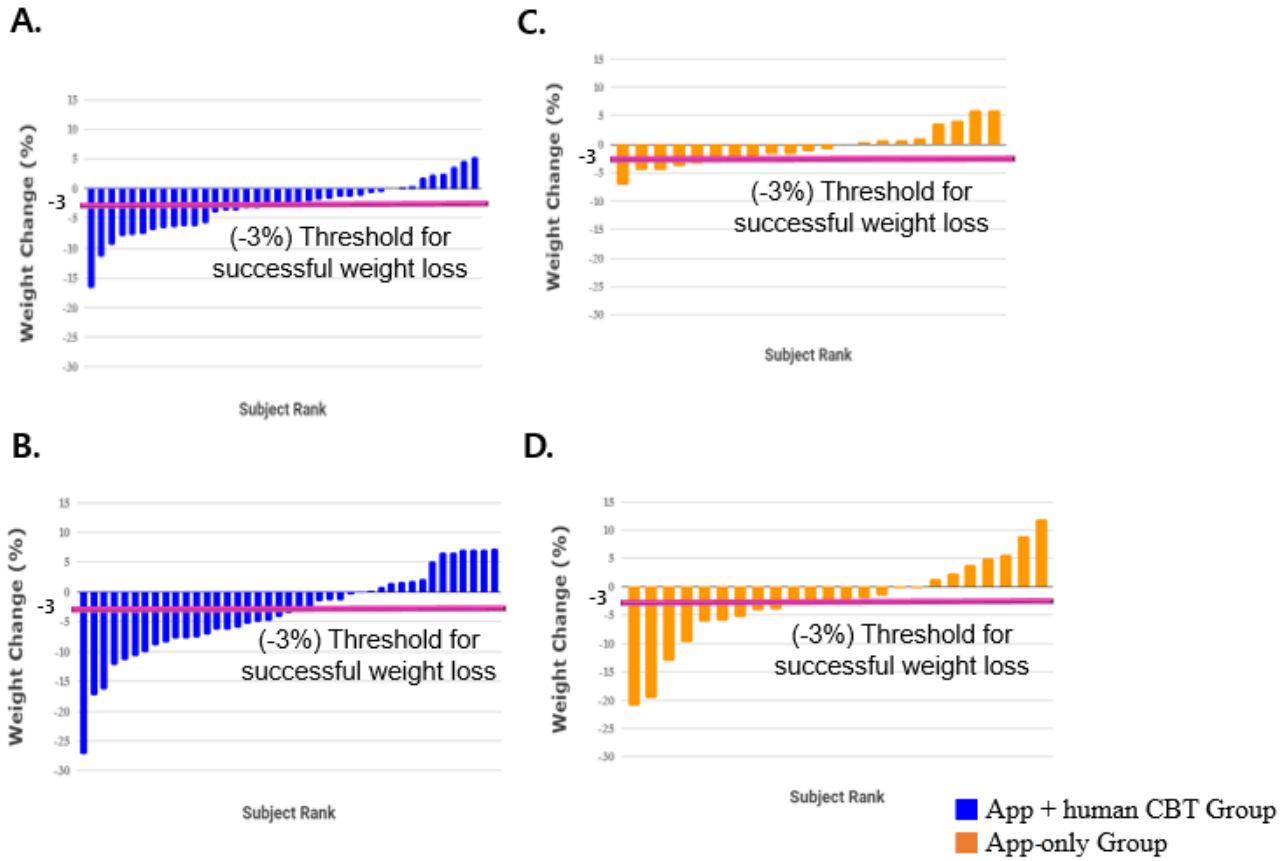
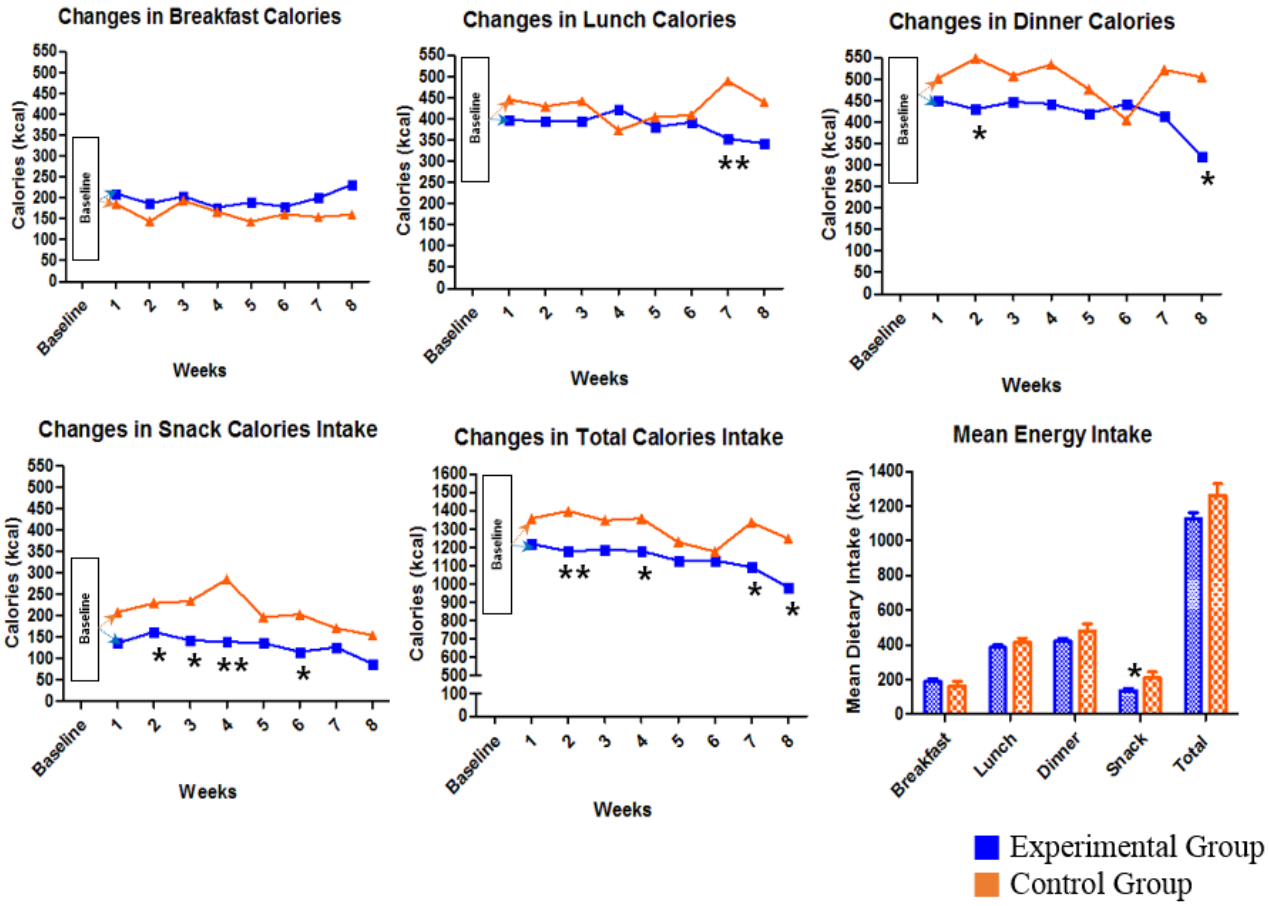


Figure 6. Changes in meal calories between experimental and control groups during the intervention period, as well as the contrast of mean energy intake between groups. *P<.05; ** P<.01.

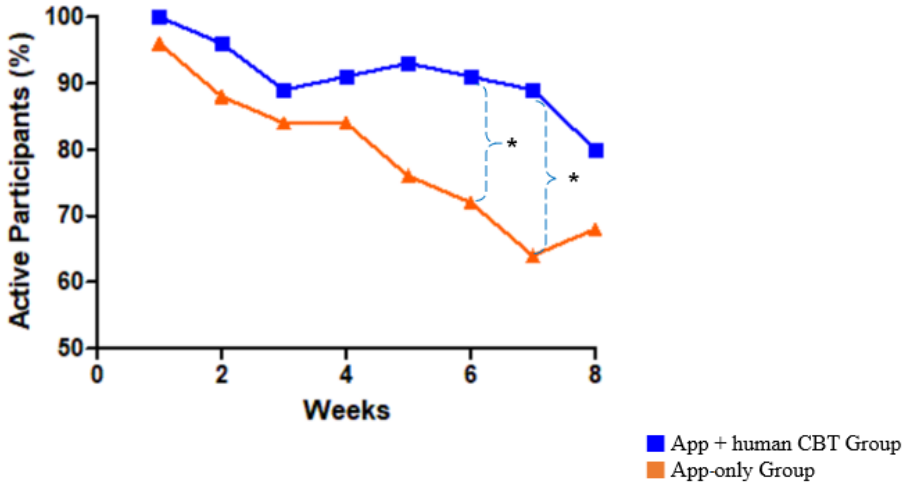


Secondary Outcomes: Metabolic and Psychological Outcomes

Multimedia Appendix 1, Table MA1-4, shows a comparison of the metabolic outcomes from baseline to 8 weeks in each group and by intervention condition. The mean decreases in leptin (–15.8%, SD 29.9, vs 7.2%, SD 35.9, P=.01), insulin (–4.4%, SD 35.2, vs 15.4%, SD 35.1, P=.048), and Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) (–7.1%, SD 35.1, vs 14.4%, SD 41.2, P=.04) were significantly greater in the digital CBT group than in the control group. For within-group analysis, the changes in glucose (–2.91%, P=.04) and leptin (–15.82%, P=.003) of the digital CBT group were significant. No significant outcome changes were found in the control group. The mean percentage changes in psychological outcomes are shown in Multimedia Appendix 1, Table MA1-5, by intervention condition. There was no significant difference between the groups regarding the number of changes in psychological outcomes except for the change in the DEBQ-EM from baseline to 8 weeks (P=.048). Paired t test analysis showed significant changes in the BSQ-8C and DEBQ-EX scores at 8 and 24 weeks in both groups. However, the changes in the scores of the DEBQ restrained eating scale (DEBQ-RE) (P<.001) at 8 weeks, and those of the K-BDI-II (P=.001), TAI (P=.04), RSES (P=.03), and ATQ-30 (P=.02) at 24 weeks, appeared to be significant only in the digital CBT group (see Multimedia Appendix 1, Tables MA1-6 and MA1-7). Behavioral outcomes, measured via the Noom app, are represented as the amount of calorie intake and the pattern of weekly changes between the groups and the average energy intake of each group, as presented in Figure 3. Mean snack calories (P=.02) significantly differed between the two groups, and total calories (P=.06) had a tendency toward critical difference by intervention condition (see Figure 6 and Multimedia Appendix 1, Table MA1-8).

Lastly, the digital CBT group had a higher engagement rate when using digital tools than the control group, though it declined over time in both groups (see Figure 7 and Multimedia Appendix 1, Table MA1-9).

Figure 7. Patterns of changes in engagement rate of the experimental and control groups during the intervention period. *P<.05.

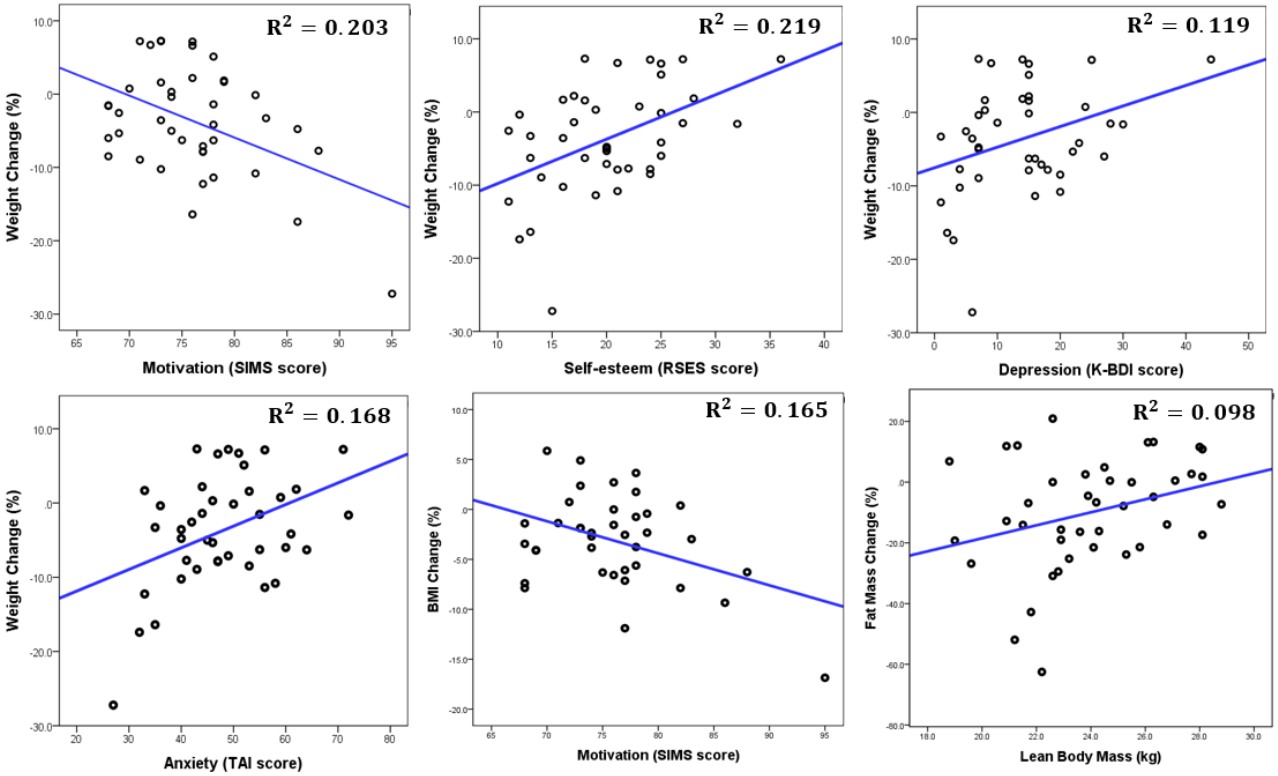


Predictors of the Primary Outcome, Weight Change

Correlations Between the Primary Outcome and the Baseline Characteristics

The baseline motivation, as measured by the SIMS, was significantly correlated with weight change at 8 weeks (P=.009) and 24 weeks (P=.003). Depression, as measured by the K-BDI-II (P=.03); anxiety, as measured by the TAI (P=.008); and self-esteem, as measured by the RSES (P=.002) at baseline also showed a significant correlation with weight change at 24 weeks but not at 8 weeks. Depression, anxiety, self-esteem, restrained eating behavior, external eating behavior, and automatic thoughts at baseline were significantly correlated with BMI change at 24 weeks. Lastly, lean body mass, anxiety, and self-esteem at baseline were significantly correlated with change in body fat mass at 24 weeks. Figure 8 illustrates the significant correlations between the predictive markers and the change of the anthropometric measures at 24 weeks. All the results of the correlation analysis are presented in detail in Multimedia Appendix 1, Table MA1-10. Multimedia Appendix 1, Figure MA1-1, also illustrates the correlations between predictive markers and the change of BMI.

Figure 8. The correlation between weight change at the long-term follow-up period (24 weeks) and the level of motivation, self-esteem, depression, and anxiety at baseline. Also shown are the correlation between BMI change at the long-term follow-up and the level of motivation at baseline, and the correlation between fat mass change at the long-term follow-up and lean body mass at baseline. K-BDI: Korean version of the Beck Depression Inventory; RSES: Rosenberg Self-Esteem Scale; SIMS: Situational Motivation Scale; TAI: Trait Anxiety Inventory.



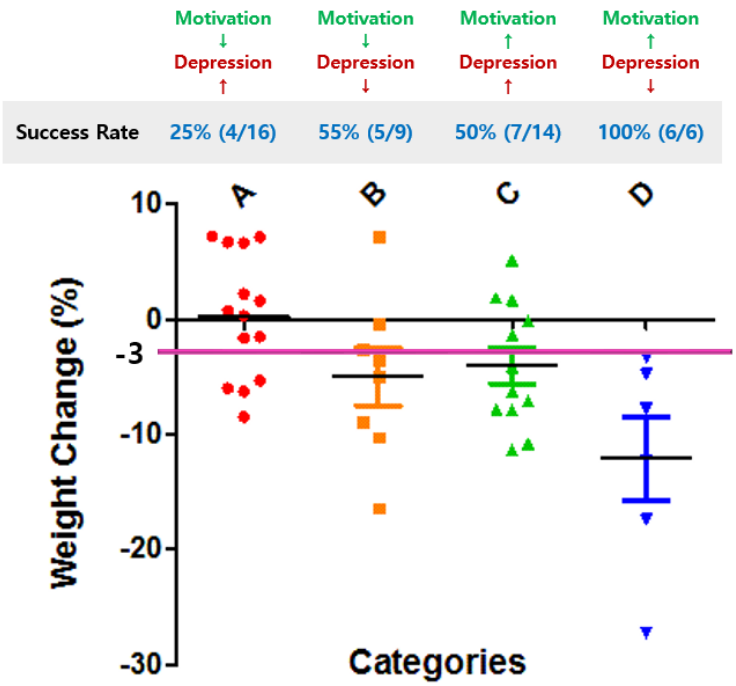
Receiver Operating Characteristic Analysis Determining the Optimal Cutoff Scores of the Predictive Markers of Success in Weight Loss by Digital Cognitive Behavioral Therapy

Multimedia Appendix 1, Table MA1-11, shows the sensitivity and specificity of the baseline psychological characteristics showing significant correlations with weight change, the primary outcome. The definition of optimal statistical prediction threshold is weight loss of more than 3% of the initial body weight. This is an important threshold because our treatment was CBT as a lifestyle modification without any biological intervention. Both motivation and self-esteem had the greatest area under the curve (AUC) (0.63). The AUCs of depression and anxiety were 0.61 and 0.62, respectively. To predict a good response, the cutoff for motivation (SIMS score=76.5) provided a good trade-off between sensitivity (59%) and specificity (74%). Additionally, the cutoff for depression (K-BDI-II score=7.5), anxiety (TAI score=41.5), and self-esteem (RSES score=24.5) provided optimal sensitivity and specificity to predict a good response. Overall, motivation showed the best predictive performance.

Clinical Efficacy of Digital Cognitive Behavioral Therapy Based on the Optimal Cutoff Scores of the Predictive Markers in the Clinical Setting

The high-motivation subgroup (SIMS scores >76.5) showed a 65% (13/20) probability of successful 3% weight loss, whereas the low-motivation subgroup (SIMS scores <76.5) showed a 36% (9/25) probability of successful 3% weight loss. Optimal predictive performance was achieved by combining both motivation and depression scores. The high-motivation plus low-depression subgroup (SIMS scores >76.5 and K-BDI-II scores <7.5) showed a 100% (6/6) probability of successful 3% weight loss. Other subgroups showed a lower probability of successful 3% weight loss: 55% (5/9) of the low-motivation and low-depression subgroup, 50% (7/14) of the high-motivation and high-depression subgroup, and 25% (4/16) of the low-motivation and high-depression subgroup (see Figure 9).

Figure 9. The clinical efficacy of digital cognitive behavioral therapy (CBT) by applying the optimal cutoff scores of the predictive markers in the clinical setting. The pink line represents the threshold for successful weight loss.



Even when the strict statistical threshold for multiple comparison corrections was applied, changes in weight, BMI, and fat mass from baseline to 8 weeks in the digital CBT group were considered significant (P<.001). The changes in the scores of the DEBQ-RE from baseline to 8 weeks, and in the K-BDI-II and DEBQ-EX scores from baseline to 24 weeks, in the digital CBT group were also significant after multiple corrections (P<.001). Furthermore, the changes in the scores of the BSQ-8C from baseline to 8 weeks and 24 weeks in both the digital CBT and control groups were considered significant after multiple corrections (P<.001).

Discussion

Principal Findings

This study successfully examined the efficacy of a newly developed, multifactorial, and daily-based personalized CBT model conducted by a psychologist via a digital platform for managing body weight, BMI, and body fat mass and showed a legacy effect even after the intervention terminated. This was performed by comparing this group to the active comparators using only the app as the control group. Furthermore, this study successfully explored the predictors for the efficacy of digital CBT from the baseline characteristics and recommended them

as precision medicine biomarkers, namely, depression, anxiety, self-esteem, and motivation.

Among mobile health (mHealth) RCTs for obesity, this study has unique implications regarding the application of CBT strategies by a human coach in the intervention. This study, therefore, contributes to the broader literature on weight-loss treatments that involve human factors. There have been widespread studies of mHealth approaches to weight-loss programs [18-33]. There are several studies of obesity treatments that did not investigate CBT settings; these include studies of human-based mHealth RCTs [18,21,24,29,34,35], human-based mHealth but without RCT design [17,26,30], and mHealth not based on human factors but with RCT design [19,20,22,25,27,28,32,33]. There are also several studies of human-based RCT designs, including CBT settings for obesity but not mHealth procedures (ie, telephone, website, face-to-face, and others) [16,36-40].

This study is comparable to other mHealth RCTs. The mean percentage weight loss of our study was 4% of initial body weight, and previous mHealth RCTs reported a mean percentage weight loss ranging from 1% to 3% [24,29,32,34]. Moreover, this study successfully showed weight maintenance. Most interventions for obesity have shown a tendency to regain weight after discontinuing the treatment [11,16,24,29,32,41-43], but our digital CBT intervention showed a sustained trend of further decrease even up to 16 weeks after cessation of the 8-week intervention. This affords solid support for the assumption that digital CBT promotes an overall healthy lifestyle. However, because we do not have data beyond a 1-year period, a direct comparison with previous studies is not feasible. Preventing weight regain at 24 weeks is closely related with a decrease in body fat mass and an increase in lean body mass at 8 weeks, which are relevant to physical activity rate and nutrition status [44,45]. Indeed, an improvement in both physical activity and diet, representing changes in lifestyle, leads to healthy body composition. Therefore, the patterns of changes in not only body weight but also body fat mass and lean body mass may imply that the participants in the digital CBT group experienced self-sustainable transitions in daily decision making for a healthy life.

With regard to the appropriate threshold, previous behavioral weight-loss studies often reported 5% weight loss in the majority of participants [16,26,30]. Conventionally, several studies adopted a 5% threshold as a clinically significant threshold [16,19,26,29]. However, in contrast to the conventional 5% threshold, we adopted a tempered 3% weight-loss threshold as the *good response* threshold for two main reasons. First, the duration of the active intervention period in this study was shorter than in other studies and only persisted transiently for the initial 2 months. The majority of previous behavioral studies had a full 6-month active intervention design [16,36,37]. However, the duration of the active intervention period in our study was only 8 weeks (2 months). There was no intervention delivered after 8 weeks (2 months) until the 6-month time point. Thus, the subjects did not receive the intervention during the remaining 4 months after the initial 2-month active intervention. Second, the components of the intervention in this study did not include extreme restrictions or requirements in either diet or exercise. The main goal of our intervention was to implement sustainable weight management skills by learning an appropriate behavioral process as well as establishing new cognitive processes. Therefore, the weight loss per se could be weaker than with the stringent diet restrictions and exercise requirements of a behavioral program during the intervention. In addition to the 3% threshold, we also reported the results based on the conventional 5% threshold to allow a direct comparison of clinical efficacy between studies.

Regarding personalization, our digital CBT was fully tailored to each participant's characteristics in multifactorial domains: the behavioral, cognitive, emotional, motivational, and physical domains. The therapist in our study altered the feedback styles based on data from five types of domains for every participant and conducted intensive daily monitoring. Most of the previous RCTs on mHealth interventions for obesity—those not based on human factors—considered one or two factors of individual symptoms that led to the implementation of homogeneous interventions [18,19,22,31]. Although there are some interventions that use custom algorithms to provide individualized feedback, they only focus on diet, physical activity, weight loss, or any two of these [27,28,32,46]. Furthermore, some earlier mHealth RCTs for obesity based on human factors only dealt with diet and physical activities [21,24]. One study managed three domains for the intervention: diet, physical activities, and eating behaviors [29]. However, instructions on behavior change strategies were not delivered by smartphone but by attending weekly group sessions for the first phase of the intervention. The study was deficient in other principal factors, such as emotional, cognitive, and motivational domains, implying insufficient potentiality for long-term lifestyle change. Because cognitive conceptualization and emotional regulation process are naturally associated with behavioral patterns, consideration of all these components can allow changes in one's lifestyle and ultimately solve problems related to obesity [47]. Therefore, it is important to address the respective multifactorial domains so as to conduct tailored treatment for individuals with fully integrated techniques. Our digital CBT strategies operate in a fully comprehensive system that deals with behavioral, cognitive, emotional, motivational, and physical factors and allows integrated mediation to successfully manage obesity.

After examining aspects of temporal strategies for intervention, we arranged three different time points (ie, daily, weekly, and monthly points) and initiated a daily human-agent intervention in an mHealth RCT for obesity. All previous face-to-face, electronic health (eHealth), and mHealth RCTs for obesity treatment have been either weekly- or monthly-based interventions delivered by therapists [7,16,21,24,25,29,36-38]. Temporal strategies can influence the engagement rate, which is closely related to treatment outcomes [48]. Unfortunately, according to a systematic review, mHealth RCTs related to weight-loss programs suffer from a high attrition rate of more than 30% [49]. Our digital CBT trial, however, showed high in-app activity rates as well as engagement in the intervention program. Only one participant in the digital CBT group had to withdraw for personal reasons; 80% of the participants were active until the end of the treatment session. One possible reason

for these outcomes is that our digital CBT intervention effectively managed participants' motivation to lose weight as well as participate in the program. Our intervention did this by delivering individualized messages every day, based on data from the in-app database and daily assessment of various psychological factors using CBT modules, as well as facilitating real-time access to the therapist. Additionally, the midreport, employed as a monthly intervention, allowed personalized precision treatment based on initial psychological conditions to keep participants motivated. Thus, the engagement rate of the digital CBT group improved from 91% to 93%, whereas the engagement rate in the control group dropped from 84% to 76% between Week 4 and Week 5.

Through our digital CBT, changes in biological indexes, leptin, insulin, and HOMA-IR indicated that factors related to physical health can be successfully improved. Moreover, we also successfully managed motivation, emotion, cognition, and behavior. The level of self-body-image satisfaction and external eating behaviors was improved in both groups. This indicates that simply including the standard mHealth treatment in the control group in our study was practical for improving body image perception and external eating habits. Digital CBT improved the level of depression, anxiety, self-esteem, and automatic thoughts related to depression. In fact, the DEBQ-EM and DEBQ-EX scores showed a significant difference between the two groups at baseline but were not notably correlated with the primary measures at baseline. This may be considered a random circumstance of randomization. Therefore, these differences can be interpreted as not affecting the main outcomes of our study. Furthermore, a significant difference in reported snack calorie intake between the two groups suggests that our digital CBT intervention had an impact on managing snack calories compared to other meals. Stress is highly correlated with the frequency of snacks [50]. Thus, it is possible that our digital CBT intervention affected snack calorie intake by finding individualized stress coping strategies, restructuring cognitive structures of automatic negative eating or weight-related thoughts, and developing regular and balanced eating behaviors. Therefore, this provides evidence that the participants in the digital CBT group changed their lifestyle to constantly manage their weight.

This study can be considered a practical one because it explored clinical markers that predict the effect of digital CBT and suggested plausible criteria that can be applied to clinical settings. The follow-up results at 24 weeks in this study showed that the levels of motivation, depression, anxiety, and self-esteem were the predictive markers of weight loss based on the digital CBT intervention. Some of our results regarding the predictors of weight control conflict with the findings of previous research [51], but they are consistent with recent findings that the level of motivation is the strongest predictive trait for weight control [52,53]. We defined people who lost less than 3% of their baseline weight as poor responders to the treatment. Thus, people with a SIMS score lower than 76.5 are recommended to find and pursue their own way of enhancing their motivation to lose weight before they undertake digital CBT. Furthermore, a person whose score is higher than 7.5 on the K-BDI-II, 41.5 on the TAI, or 24.5 on the RSES is encouraged to handle the relevant issue before, or at the same time as, digital CBT. This will prevent further distress from repeated failure to control weight, save limited resources, and allow better concentration in individuals with a higher chance of success in weight control.

Considering the comparator of this study as the best active comparator without human coaching, digital CBT is a competent intervention for obesity in the current situation in the digital health care industry. We provided education on how to log meals and exercise as well as how to use InBody Dial and the mobile app, not only to the digital CBT group but also to the control group during the orientation. Thus, the control group in this study can be defined as an active group as in previous studies [19,22,31]. As expected, the control group in this study showed favorable results. Therefore, the results of this study are superior and significant compared to those of previous studies of digital health care interventions.

Limitations

While the results are highly promising, the study is not without limitations. First of all, the participants were limited to those in their 20s and 30s, resulting in limited generalizability. Second, since this is not a blinded study, an observer bias could have been generated. Thus, an implication of this study that should be noted is that it tested the digital CBT and did not validate it. Third, the sample size was relatively small (N=70). Therefore, most of the results did not pass the strict multiple-comparison-corrected *P* threshold. Fourth, the follow-up period needs to be extended to increase the reliability and validity of our results. Accordingly, we recommend that future studies examine more information on personal characteristics, such as single nucleotide polymorphisms (SNPs) and daily patterns of digital phenotypes for individuals within in-app data, in order to enhance the interpretation of the efficacy of digital-based interventions. Fifth, the total amount of food calories in the app might have been underestimated because the amount per serving for diverse types of food was not precise and people may have miscalculated their food intake. The primary reason for errors in food records is that most people have difficulties in estimating food portions [54]. The discrepancy in food choice between the food diary and actual meals (ie, recording similar but not exact menus, skipping reports of foods eaten, or logging foods not offered) could explain the remainder of the total miscalculation [55]. Thus, we suggest that a direct assessment of food choice and intake, such as buffet tests, should be performed in parallel with logging intake in the food diary on the app for future research. In addition, it should be noted that it is necessary to involve dietitians on multidisciplinary health care teams for obesity CBT, as their evaluations of dietary assessment and nutritional advice would greatly strengthen the efficacy of the intervention. Lastly, there is a feasibility issue regarding the digital CBT of this study since it is intensive and costly, requiring daily intervention by therapists trained in both physical and mental health care. Therefore, more research involving human factors in technology-based treatments should be conducted to collect enough data to create automatic functions, thereby decreasing the burdens of therapists in the future.

Conclusions

For the first time, we discovered that human-based digital CBT is capable of treating obesity using digital tools. Anthropometric measures, such as body weight and body compositions, were comparably improved by the digital CBT model as well as physiological indices and obesity-related psychological factors. There was no relapse in weight change after the end of the intervention. We also found predictable psychological markers to estimate the efficacy of the digital CBT treatment for obesity. This will open up new aspects of digital precision remedies for obesity in the digital health care industry.

Acknowledgments

We would like to thank Hyunjae Kim, the chief executive manager of LookinBody, for supporting this study by sharing his company's advanced digital technologies. Also, I would like to thank Ms Younghyun Yun from the Department of Anatomy and Cell Biology, Seoul National University College of Medicine, for the preparation of the excellent illustrations and graphic design. This study was supported by a grant from the National Research Foundation of Korea (NRF), funded by the Korean Government, Ministry of Science and ICT (MSIT) (No. NRF-2018R1A5A2025964), and was supported by the Creative-Pioneering Researchers Program through Seoul National University (SNU).

Noom provided the funding to conduct this research and InBody provided body composition analyzer devices for this research. Representatives of InBody had no role in the management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. YK, an employee of Noom, participated in the generation of the study design and in data collection.

Authors' Contributions

MK conceptualized and designed the clinical infrastructure for the digital CBT intervention during the implementation phase. HJC and SC gave valuable research insights when designing the digital CBT intervention. MK, HJC, YK, YG, MN, SL, and YL contributed to the study design and data collection. MK, HJC, YG, and SC analyzed and interpreted the data. MK wrote the manuscript and edited the contents of the manuscript. HJC and SC reviewed the manuscript. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

YK is an employee of Noom.

Multimedia Appendix 1
Changes in outcomes from baseline, correlations, and predicting efficacy of digital cognitive behavioral therapy.
[DOCX File , 200 KB - mhealth_v8i4e14817_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1466 KB - mhealth_v8i4e14817_app2.pdf]

References

1. Gardner CD, Trepanowski JF, Del Gobbo LC, Hauser ME, Rigdon J, Ioannidis JP, et al. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: The DIETFITS randomized clinical trial. JAMA 2018 Feb 20;319(7):667-679 [FREE Full text] [doi: 10.1001/jama.2018.0245] [Medline: 29466592]
2. Wyatt HR. Update on treatment strategies for obesity. J Clin Endocrinol Metab 2013 Apr;98(4):1299-1306 [FREE Full text] [doi: 10.1210/jc.2012-3115] [Medline: 23443815]
3. Sjöström L, Narbro K, Sjöström CD, Karason K, Larsson B, Wedel H, Swedish Obese Subjects Study. Effects of bariatric surgery on mortality in Swedish obese subjects. N Engl J Med 2007 Aug 23;357(8):741-752. [doi: 10.1056/NEJMoa066254] [Medline: 17715408]
4. Castelnuovo G, Pietrabissa G, Manzoni GM, Cattivelli R, Rossi A, Novelli M, et al. Cognitive behavioral therapy to aid weight loss in obese patients: Current perspectives. Psychol Res Behav Manag 2017;10:165-173 [FREE Full text] [doi: 10.2147/PRBM.S113278] [Medline: 28652832]
5. Cooper MJ, Wells A, Todd G. A cognitive model of bulimia nervosa. Br J Clin Psychol 2004 Mar;43(Pt 1):1-16. [doi: 10.1348/014466504772812931] [Medline: 15005903]
6. Brennan L. Cognitive Behavioural Evaluation and Treatment of Adolescent Overweight and Obesity [doctoral thesis]. Melbourne, Australia: RMIT University; 2006 Aug 31. URL: http://researchbank.rmit.edu.au/eserv/rmit:6333/Brennan.pdf [accessed 2006-08-31]

7. Tsiros MD, Sinn N, Brennan L, Coates AM, Walkley JW, Petkov J, et al. Cognitive behavioral therapy improves diet and body composition in overweight and obese adolescents. Am J Clin Nutr 2008 May;87(5):1134-1140. [doi: 10.1093/ajcn/87.5.1134] [Medline: 18469231]
8. Steele RM, Mummery WK, Dwyer T. A comparison of face-to-face or internet-delivered physical activity intervention on targeted determinants. Health Educ Behav 2009 Dec;36(6):1051-1064. [doi: 10.1177/1090198109335802] [Medline: 19502534]
9. Sherwood NE, Morton N, Jeffery RW, French SA, Neumark-Sztainer D, Falkner NH. Consumer preferences in format and type of community-based weight control programs. Am J Health Promot 1998;13(1):12-18. [doi: 10.4278/0890-1171-13.1.12] [Medline: 10186930]
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. Yon BA, Johnson RK, Harvey-Berino J, Gold BC. The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake. J Am Diet Assoc 2006 Aug;106(8):1256-1259. [doi: 10.1016/j.jada.2006.05.004] [Medline: 16863723]
12. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: Systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
13. 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]
14. Hey SP, Kimmelman J. The questionable use of unequal allocation in confirmatory trials. Neurology 2014 Jan 07;82(1):77-79 [FREE Full text] [doi: 10.1212/01.wnl.0000438226.10353.1c] [Medline: 24306005]
15. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
16. Cooper Z, Doll HA, Hawker DM, Byrne S, Bonner G, Eeley E, et al. Testing a new cognitive behavioural treatment for obesity: A randomized controlled trial with three-year follow-up. Behav Res Ther 2010 Aug;48(8):706-713 [FREE Full text] [doi: 10.1016/j.brat.2010.03.008] [Medline: 20691328]
17. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A mobile phone-based health coaching intervention for weight loss and blood pressure reduction in a national payer population: A retrospective study. JMIR Mhealth Uhealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
18. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 01;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
19. 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]
20. Spring B, Duncan JM, Janke EA, Kozak AT, McFadden HG, DeMott A, et al. Integrating technology into standard weight loss treatment: A randomized controlled trial. JAMA Intern Med 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1221] [Medline: 23229890]
21. Allen JK, Stephens J, Dennison Himmelfarb CR, 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]
22. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: A randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-S12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
23. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
24. Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): A 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755 [FREE Full text] [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
25. Jun MK, Ha JY. Effect of smartphone apps applying BodyThink program on obesity in adolescent girls [Article in Korean]. J Korean Acad Nurs 2016 Jun;46(3):390-399. [doi: 10.4040/jkan.2016.46.3.390] [Medline: 27411766]
26. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
27. Mummah SA, Mathur M, King AC, Gardner CD, Sutton S. Mobile technology for vegetable consumption: A randomized controlled pilot study in overweight adults. JMIR Mhealth Uhealth 2016 May 18;4(2):e51 [FREE Full text] [doi: 10.2196/mhealth.5146] [Medline: 27193036]

28. Turner-McGrievy GM, Wilcox S, Boutté A, Hutto BE, Singletary C, Muth ER, et al. The Dietary Intervention to Enhance Tracking with Mobile Devices (DIET Mobile) Study: A 6-month randomized weight loss trial. Obesity (Silver Spring) 2017 Aug;25(8):1336-1342 [FREE Full text] [doi: 10.1002/oby.21889] [Medline: 28600833]
29. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: The ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]
30. Toro-Ramos T, Lee D, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a smartphone application for the management of metabolic syndrome components focusing on weight loss: A preliminary study. Metab Syndr Relat Disord 2017 Nov;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
31. Turner-McGrievy G, Tate D. Tweets, apps, and pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec 20;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
32. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: A randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
33. Shaw RJ, Bosworth HB, Silva SS, Lipkus IM, Davis LL, Sha RS, et al. Mobile health messages help sustain recent weight loss. Am J Med 2013 Nov;126(11):1002-1009 [FREE Full text] [doi: 10.1016/j.amjmed.2013.07.001] [Medline: 24050486]
34. Nezami BT, Ward DS, Lytle LA, Ennett ST, Tate DF. A mHealth randomized controlled trial to reduce sugar-sweetened beverage intake in preschool-aged children. Pediatr Obes 2018 Nov;13(11):668-676. [doi: 10.1111/ijpo.12258] [Medline: 29119719]
35. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mHealth intervention for large, sustained change in multiple diet and activity risk behaviors: The Make Better Choices 2 randomized controlled trial. J Med Internet Res 2018 Jun 19;20(6):e10528 [FREE Full text] [doi: 10.2196/10528] [Medline: 29921561]
36. Tate DF, Wing RR, Winett RA. Using internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 07;285(9):1172-1177. [doi: 10.1001/jama.285.9.1172] [Medline: 11231746]
37. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329 [FREE Full text] [doi: 10.1038/oby.2004.40] [Medline: 14981225]
38. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
39. Stapleton P, Bannatyne A, Chatwin H, Urzi K, Porter B, Sheldon T. Secondary psychological outcomes in a controlled trial of Emotional Freedom Techniques and cognitive behaviour therapy in the treatment of food cravings. Complement Ther Clin Pract 2017 Aug;28:136-145. [doi: 10.1016/j.ctcp.2017.06.004] [Medline: 28779921]
40. Zerwas SC, Watson HJ, Hofmeier SM, Levine MD, Hamer RM, Crosby RD, et al. CBT4BN: A randomized controlled trial of online chat and face-to-face group therapy for bulimia nervosa. Psychother Psychosom 2017;86(1):47-53 [FREE Full text] [doi: 10.1159/000449025] [Medline: 27883997]
41. Wadden TA, Hollander P, Klein S, Niswender K, Woo V, Hale PM, NN8022-1923 Investigators. Weight maintenance and additional weight loss with liraglutide after low-calorie-diet-induced weight loss: The SCALE Maintenance randomized study. Int J Obes (Lond) 2013 Nov;37(11):1443-1451. [doi: 10.1038/ijo.2013.120] [Medline: 23812094]
42. Srivastava G, Buffington C. A specialized medical management program to address post-operative weight regain in bariatric patients. Obes Surg 2018 Aug;28(8):2241-2246. [doi: 10.1007/s11695-018-3141-z] [Medline: 29464536]
43. Pareek M, Schauer PR, Kaplan LM, Leiter LA, Rubino F, Bhatt DL. Metabolic surgery: Weight loss, diabetes, and beyond. J Am Coll Cardiol 2018 Feb 13;71(6):670-687 [FREE Full text] [doi: 10.1016/j.jacc.2017.12.014] [Medline: 29420964]
44. Vehrs PR, Hager R. Assessment and interpretation of body composition in physical education. J Phys Educ Recreat Dance 2013 Jan 26;77(7):46-51 [FREE Full text] [doi: 10.1080/07303084.2006.10597907]
45. Kyle UG, Schutz Y, Dupertuis YM, Pichard C. Body composition interpretation. Contributions of the fat-free mass index and the body fat mass index. Nutrition 2003;19(7-8):597-604. [doi: 10.1016/s0899-9007(03)00061-3] [Medline: 12831945]
46. Kim EK, Kwak SH, Jung HS, Koo BK, Moon MK, Lim S, et al. The effect of a smartphone-based, patient-centered diabetes care system in patients with type 2 diabetes: A randomized, controlled trial for 24 weeks. Diabetes Care 2019 Jan;42(1):3-9. [doi: 10.2337/dc17-2197] [Medline: 30377185]
47. Fabricatore AN. Behavior therapy and cognitive-behavioral therapy of obesity: Is there a difference? J Am Diet Assoc 2007 Jan;107(1):92-99. [doi: 10.1016/j.jada.2006.10.005] [Medline: 17197276]
48. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: A systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
49. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
50. Jeong EK, Kim KN. Influence of stress on snack consumption in middle school girls. Nutr Res Pract 2007;1(4):349-355 [FREE Full text] [doi: 10.4162/nrp.2007.1.4.349] [Medline: 20368961]

51. Teixeira PJ, Going SB, Sardinha LB, Lohman TG. A review of psychosocial pre-treatment predictors of weight control. Obes Rev 2005 Feb;6(1):43-65. [doi: 10.1111/j.1467-789X.2005.00166.x] [Medline: 15655038]
52. Teixeira PJ, Going SB, Houtkooper LB, Cussler EC, Martin CJ, Metcalfe LL, et al. Weight loss readiness in middle-aged women: Psychosocial predictors of success for behavioral weight reduction. J Behav Med 2002 Dec;25(6):499-523. [doi: 10.1023/a:1020687832448] [Medline: 12462956]
53. Teixeira PJ, Going SB, Houtkooper LB, Cussler EC, Metcalfe LL, Blew RM, et al. Pretreatment predictors of attrition and successful weight management in women. Int J Obes Relat Metab Disord 2004 Sep;28(9):1124-1133. [doi: 10.1038/sj.ijo.0802727] [Medline: 15263921]
54. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet 2018 Apr;75(2):219-225. [doi: 10.1111/1747-0080.12401] [Medline: 29280547]
55. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. Nutrition 2005 Jun;21(6):672-677. [doi: 10.1016/j.nut.2004.11.006] [Medline: 15925290]

Abbreviations

ATQ-30: Automatic Thoughts Questionnaire
AUC: area under the curve
BSQ: Body Shape Questionnaire
BSQ-8C: Body Shape Questionnaire-8C
CBT: cognitive behavioral therapy
DEBQ: Dutch Eating Behavior Questionnaire
DEBQ-EM: Dutch Eating Behavior Questionnaire emotional eating scale
DEBQ-EX: Dutch Eating Behavior Questionnaire external eating scale
DEBQ-RE: Dutch Eating Behavior Questionnaire restrained eating scale
eHealth: electronic health
HOMA-IR: Homeostatic Model for Assessment of Insulin Resistance
K-BDI-II: Korean version of the Beck Depression Inventory-II
mHealth: mobile health
NRF: National Research Foundation of Korea
RCT: randomized controlled trial
ROC: receiver operating characteristic
RSES: Rosenberg Self-Esteem Scale
SIMS: Situational Motivation Scale
SNP: single nucleotide polymorphism
SNU: Seoul National University
TAI: Trait Anxiety Inventory

Edited by G Eysenbach; submitted 25.05.19; peer-reviewed by S Smith, J Chen, A Pfammatter, M Gaman; comments to author 30.09.19; revised version received 13.10.19; accepted 09.02.20; published 30.04.20.

Please cite as:
Kim M, Kim Y, Go Y, Lee S, Na M, Lee Y, Choi S, Choi HJ
Multidimensional Cognitive Behavioral Therapy for Obesity Applied by Psychologists Using a Digital Platform: Open-Label Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e14817
URL: http://mhealth.jmir.org/2020/4/e14817/
doi:10.2196/14817
PMID:32352391

Original Paper

A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial

Sandra Hauser-Ulrich[1], MSc; Hansjörg Künzli[1], lic phil; Danielle Meier-Peterhans[2], MSc; Tobias Kowatsch[3,4], PhD

[1]Department of Applied Psychology, University of Applied Sciences Zurich, Zurich, Switzerland

[2]Department Primary School, Division Education, Culture and Sports, Aarau, Switzerland

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

[4]Center for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

Corresponding Author:
Sandra Hauser-Ulrich, MSc
Department of Applied Psychology
University of Applied Sciences Zurich
Pfingstweidstrasse 96
Zurich, 8005
Switzerland
Phone: 41 58 934 84 ext 51
Email: sandra.hauser-ulrich@zhaw.ch

Abstract

Background: Ongoing pain is one of the most common diseases and has major physical, psychological, social, and economic impacts. A mobile health intervention utilizing a fully automated text-based health care chatbot (TBHC) may offer an innovative way not only to deliver coping strategies and psychoeducation for pain management but also to build a working alliance between a participant and the TBHC.

Objective: The objectives of this study are twofold: (1) to describe the design and implementation to promote the chatbot painSELfMAnagement (SELMA), a 2-month smartphone-based cognitive behavior therapy (CBT) TBHC intervention for pain self-management in patients with ongoing or cyclic pain, and (2) to present findings from a pilot randomized controlled trial, in which effectiveness, influence of intention to change behavior, pain duration, working alliance, acceptance, and adherence were evaluated.

Methods: Participants were recruited online and in collaboration with pain experts, and were randomized to interact with SELMA for 8 weeks either every day or every other day concerning CBT-based pain management (n=59), or weekly concerning content not related to pain management (n=43). Pain-related impairment (primary outcome), general well-being, pain intensity, and the bond scale of working alliance were measured at baseline and postintervention. Intention to change behavior and pain duration were measured at baseline only, and acceptance postintervention was assessed via self-reporting instruments. Adherence was assessed via usage data.

Results: From May 2018 to August 2018, 311 adults downloaded the SELMA app, 102 of whom consented to participate and met the inclusion criteria. The average age of the women (88/102, 86.4%) and men (14/102, 13.6%) participating was 43.7 (SD 12.7) years. Baseline group comparison did not differ with respect to any demographic or clinical variable. The intervention group reported no significant change in pain-related impairment (P=.68) compared to the control group postintervention. The intention to change behavior was positively related to pain-related impairment (P=.01) and pain intensity (P=.01). Working alliance with the TBHC SELMA was comparable to that obtained in guided internet therapies with human coaches. Participants enjoyed using the app, perceiving it as useful and easy to use. Participants of the intervention group replied with an average answer ratio of 0.71 (SD 0.20) to 200 (SD 58.45) conversations initiated by SELMA. Participants' comments revealed an appreciation of the empathic and responsible interaction with the TBHC SELMA. A main criticism was that there was no option to enter free text for the patients' own comments.

Conclusions: SELMA is feasible, as revealed mainly by positive feedback and valuable suggestions for future revisions. For example, the participants' intention to change behavior or a more homogenous sample (eg, with a specific type of chronic pain) should be considered in further tailoring of SELMA.

Trial Registration: German Clinical Trials Register DRKS00017147; https://tinyurl.com/vx6n6sx, Swiss National Clinical Trial Portal: SNCTP000002712; https://www.kofam.ch/de/studienportal/suche/70582/studie/46326.

(JMIR Mhealth Uhealth 2020;8(4):e15806) doi:10.2196/15806

KEYWORDS

conversational agent; chatbot; digital health; pain self-management; cognitive behavior therapy; smartphone; psychoeducation; text-based; health care; chronic pain

Introduction

Chronic pain is a widespread medical condition associated with significant negative social, physical, mental, and economic impacts [1]. The prevalence of chronic pain in Europe is estimated between 10% and 30% [2] and approximately 30% for the US population [3]. In terms of economic impact, costs associated with chronic pain in Germany were estimated to reach 29 billion EUR in 2003, with absence from work due to pain estimated at a total of 14.5 million days per year [4].

Current psychological approaches to the management of chronic pain include interventions that aim to achieve increased self-management, behavior change, and cognitive change rather than cure of the pain itself [5]. Cognitive behavioral therapy (CBT) has proven to be effective and is considered the standard therapy in chronic pain treatment [6-9]. The main focus of CBT is to establish new or enhance existing coping strategies in pain management to change negative and dysfunctional cognitions, emotions, and behavior [5,10]. As part of CBT, psychoeducation shows positive therapeutic effects and can effectively counter the process of chronification [11,12]. Psychoeducation increases patients' understanding of their condition, thereby improving compliance and helping patients to better cope with the disease.

Despite agreement on a multimodal therapy approach, according to a World Health Organization (WHO) study [13], CBT is typically not sufficiently applied and many patients do not receive CBT as part of multimodal therapy. These gaps are mainly attributed to the lack of therapists' familiarity with elements of CBT, the lack of qualified providers, or that the only available providers are located too far away for on-site consultation hours. Further barriers include lack of time, insufficient financial resources, and misunderstanding of patients' roles in their pain management [14]. Research has shown that some patients are not successful in learning new coping strategies [15,16] and remain skeptical about implementing psychological strategies.

Technology-based interventions for the management of chronic pain are becoming more popular [7,14,17], offering a potential solution to overcome these barriers to treatment because they are easily accessible and cost-effective. Further advantages of technology-based interventions include less waiting time, anonymity, and flexibility in terms of time and location of use [18,19]. These interventions are ubiquitous, with increasingly powerful technical abilities and wide acceptability [20].

In this context, so-called conversational agents (ie, computer programs that imitate communication with humans) with a health focus have recently gained interest in academia and industry, with promising results related to acceptance [21] and working alliance [22]. Working alliance refers to the extent to which a therapist and client build an attachment bond and interact with each other to achieve a shared understanding of therapeutic goals and tasks [23], which has been robustly linked to treatment success in both face-to-face and guided web-based programs [24-27]. However, computers, provided that they offer basic human cues (eg, speech output or a human-like embodiment), are perceived as social actors [28]. The concept of working alliance can be adopted to the patterns of interaction between the participant and conversational agent (eg, quality and length of messages, frequency of interaction). If the conversational agent takes the role of a communication partner and embodies a digital coach, the communication style will affect the process of relationship building and hence treatment success [22,29]. Indeed, evidence from longitudinal studies suggests that a working alliance can also be established with conversational agents [22,30]. With the help of conversational agents, interventions can be applied in a more natural and interactive way [31,32]. Digital health interventions delivered via text messages on smartphones have the potential to support patients in their everyday life, as they are cheap, fast, democratic, and popular [33]. A recent meta-analysis showed the effectiveness of conversational agents in 19 clinical and nonclinical randomized controlled trials [34].

However, there is only limited evidence on chronic pain management interventions delivered by smartphone-based conversational agents. Shamekhi et al [35] demonstrated that a home-based conversational agent can be effective when used in conjunction with medical group visits in promoting stress management techniques. However, to the best of our knowledge, no study has investigated the effectiveness of a smartphone-based conversational agent as a stand-alone intervention.

Consequently, we developed painSELfMAnagement (SELMA) as a text-based health care chatbot (TBHC) for the self-management of chronic pain. A TBHC is a conversational agent that supports health professionals in the delivery of evidence-based interventions in a ubiquitous and fully automated fashion with simple text-based messages and media objects (eg, videos, podcasts). A TBHC aims to deliver the treatment and to increase working alliance by communicating therapeutic goals and tasks in an empathetic way [36,37].

Against this background, we here describe the design and implementation of SELMA, an 8-week smartphone-based TBHC intervention for self-management of pain by patients with ongoing or cyclic pain, and present findings from a pilot randomized controlled trial, in which effectiveness, acceptance, and adherence were evaluated.

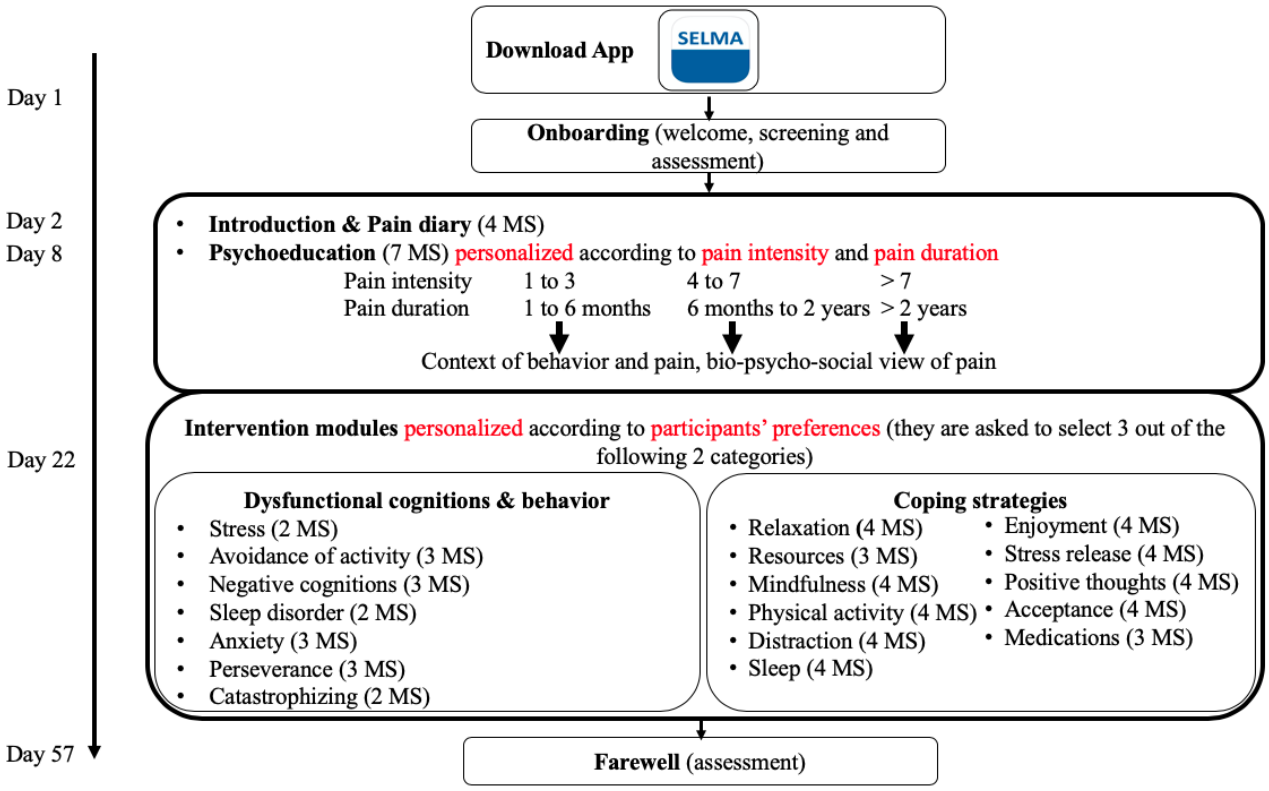
Methods

Intervention

Overview

The pain self-management coach is implemented as a TBHC named SELMA, which is represented by a drawn image of a face and acts as a guide through the CBT lesson materials. SELMA is displayed via text messages, and complemented introductions and elaborations about the material based on previous answer options and themes of particular interest to the participant. First, SELMA delivers psychoeducation from day 1 to 21 in 7 daily or other daily text message sequences that follow the same structure: initial greeting, psychoeducational input, main lesson, and goodbye. Psychoeducation enhances peoples' understanding of their disease while reducing insecurity and feelings of fear. This approach further creates transparency and reduces stereotypes against a biopsychosocial pain therapy, while strengthening therapy motivation and self-help potential [38]. From the period of day 22 until the end, SELMA delivers CBT intervention modules addressing coping strategies, along with dysfunctional cognitions and emotions that follow the same structure: initial greeting, introduction, main lesson, guidance through an exercise, goodbye. Coping strategies empower participants to take over an active role in their pain management, enabling them to control pain by analyzing and adapting their behavior, emotions, and cognitions [39]. Depending on whether or not a message sequence involves an exercise, it takes about 5 to 30 minutes to complete. Each of the 18 intervention modules consists of 2 to 4 message sequences (see Figure 1).

Figure 1. Overview of the intervention schedule. Note: One message sequence (MS) contains approximately 15 conversational turns and takes about 5 to 20 minutes to complete, depending on whether or not an exercise is involved.



Content Design and Schedule

The program starts with the onboarding process, an introductory process in which the TBHC SELMA introduces herself as a digital coach for pain self-management. SELMA provides an overview of the intervention schedule, mentions the importance of applying psychological therapy, and the importance of diagnostics in pain management. Participants are also informed that the program should not be used as a replacement for standard therapy and urges users to make an emergency call if necessary. During the first week, participants learn about the utility of a pain diary and are instructed to apply it over the following 2 weeks. In parallel, and until the end of week 3, SELMA provides information on psychoeducation based on the participants' input with regard to pain intensity and duration. SELMA explains the link between chronic pain, cognition, emotions, social impacts, and the process of chronification. During weeks 4 to 8, SELMA offers various intervention modules. Based on the participants' interest, they are able to select among 6 modules that either address dysfunctional cognitions, behavior, and emotions (eg, stress, fear of pain, anxiety) or various coping strategies (eg, activity, resources, mindfulness, acceptance). For example, the module about dysfunctional behavior covers *avoidance of activity*. It explains the link between avoiding activity and pain, and SELMA motivates participants to start physical activity and reminds them to keep their own level of proficiency in mind. By contrast, the module *mindfulness*, a coping strategy module, briefly explains the concept of mindfulness and provides users with a mindfulness exercise. Specifically, SELMA instructs participants on how to integrate mindfulness into their daily routine and provides users with a relaxation exercise. Figure 1 shows an

overview of the intervention schedule, and Multimedia Appendix 1 provides an overview of the weekly core themes and tasks to complete (including 19 references [10,38-55]). The program closes with a brief summary and farewell. All content is delivered via text messages, video clips, audio clips, figures, and PDF worksheets.

Technical Concept

Technically, SELMA was developed with MobileCoach [56], an open-source software platform for the design and evaluation of mobile TBHCs [31,55]. This platform allows the intervention authors to design (fully automated and script-based) digital health interventions consistent with the talk-and-tools paradigm [57]. That is, SELMA offers a simple chat-based interface with predefined answer options that can be used to communicate with participants in (the "talk"). To build up a social relationship [28] and working alliance [22] with participants, the linguistic style of SELMA was based on the assumption that interpersonal closeness is positively related to the attachment bond between the patient and chatbot [25,58,59]. The conversational style of SELMA is likely to affect relationship-building processes [30,60], and imitates a real human chat-based conversation by using emojis and some sense of humor. Supportive computer agents have been perceived positively [61-63]. SELMA expresses sympathy and affective empathy [64], and places emphasis on the participants' achieved tasks. This approach is based on social support [65,66] and aims for a supportive style of coaching. To engage participants further, SELMA sends out personalized text messages every day or every other day to initiate a conversation.

Message Structure and Reminders

Daily conversations are structured as a sequence of approximately 10 to 15 messages sent by SELMA, and approximately 5 to 10 replies are expected from the participants. SELMA is always initiating a conversational turn. Every text message sent by SELMA is also a notification. If the SELMA app is closed and SELMA sends a message, then a notification will always be triggered. These notifications are sticky, meaning that they are displayed in the notification dashboard and thus act as reminders. If the app is already opened, then no additional notification is triggered. To reduce the burden of the intervention, no additional reminders were used (eg, text message). In addition, if the app is closed and the push notifications are not clicked or if the app is not running in the foreground but no answer is provided (usually for 48 hours), then SELMA starts with the next message sequence. Even if participants forget to open the app or read messages (not clicking the push notification or not clicking answer options), SELMA displays the missed messages, together with the next message sequence. This mechanism ensures that participants can read all messages from SELMA, including missed messages, by simply scrolling back. Moreover, this approach is consistent with prior work [37,67,68]. To ensure completion of the questionnaire, every question displayed on one screen had to be completed; otherwise, participants were unable to proceed. Participants were excluded from the analyses if questionnaires were not completed.

Implementation

Each message sequence begins with a warm greeting, in which the chatbot enquires about the participant's mood and replies in an empathic way (eg, "Welcome back dear [$nickname]. It's nice to have you with me today.") [69]. Multimedia Appendix 2 shows examples of the sympathetic and affective empathy elements of conversation. To support establishment of a working alliance, user engagement, and motivation, SELMA addresses participants' accountability by referring to earlier tasks and activities (eg, "Welcome back to the coaching! Were you able to relax yesterday?"); she supports the completion of activities and tasks (eg, "Hi [$nickname]. How is it going with practicing your exercises?"), and motivates participants to repeat them (eg, "How did you manage the exercise, perhaps you can repeat it before the next time we meet?").

In addition to this chat-based messaging interface, "tools" are provided that support the delivery of the intervention content. For example, SELMA offers audio clips or video clips from within the chat-based interface. Figure 2 provides a representative example of an interaction with SELMA and further examples are shown in Multimedia Appendix 3.

Based on intervention rules, SELMA sends out automated messages containing intervention content or questions to the mobile (iOS or Android) apps of the participants. Based on answers given to predefined answer options (eg, from a multiple-choice question on coping strategies), free text input (eg, a question asking for the participant's nickname), or number input fields (eg, a question asking the participant's age), participants are guided through the program as outlined in Multimedia Appendix 1 and Figure 1. Some screening questions are embedded as specific focal points to determine routing to different conversational paths. The path parameters therefore change over the course of the program, depending on the user's input about prior knowledge and interest (pain intensity, pain duration, and preferred intervention modules); see Table 1 for an overview. Predefined answer options were predominantly used to assure reproducibility of the intervention. That is, automated interpretation of free-text input would result in uncertainty by triggering follow-up intervention content. From an ethical viewpoint, this was not intended. To ensure the continuance of the program and to guide those who do not want to actively choose a module (coping strategies or dysfunctional cognitions are focal points), the system displays default modules. SELMA informs participants when a module is finished, and the following day or the day after, she sends out an overview on completed modules and displays all modules for selection. The video clip in Multimedia Appendix 4 demonstrates a sample conversation.

Figure 2. Chat-based interaction with predefined answer options (bright blue, left) and multimedia content (bright blue, right); for further examples, see Multimedia Appendix 3.

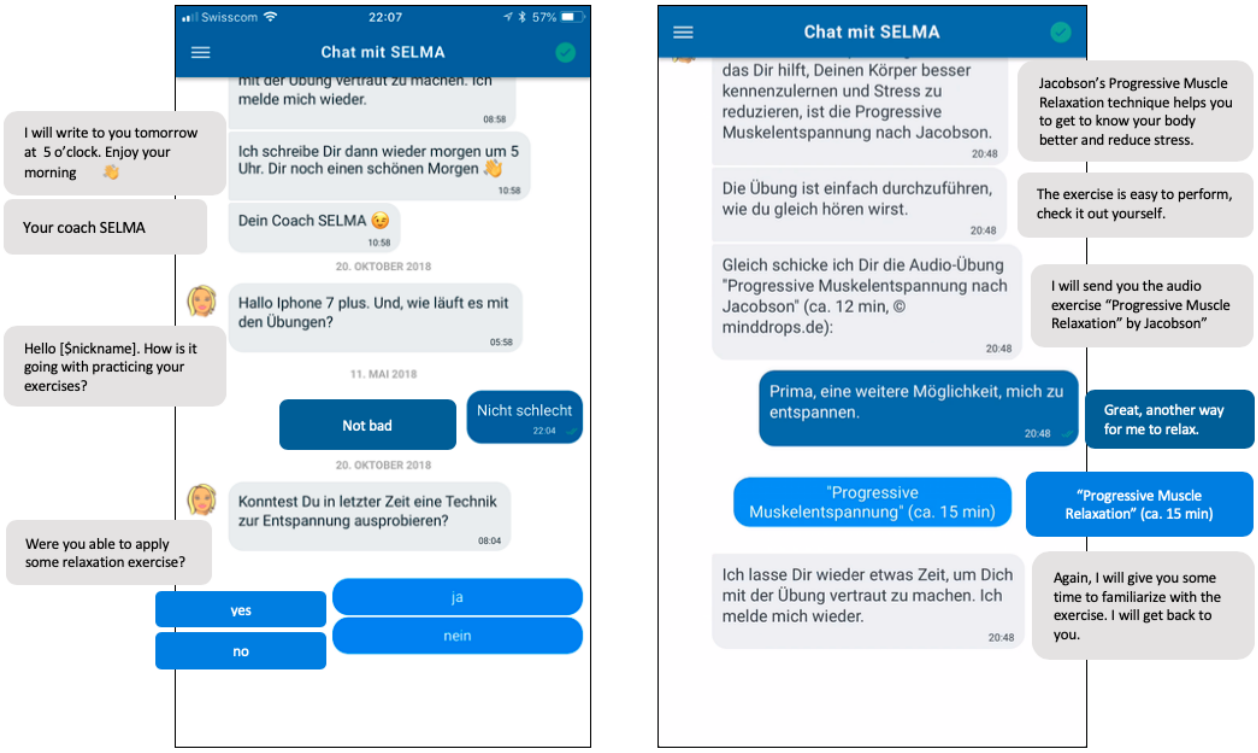


Table 1. Information gathered during the intervention.

Information gathered	Input option
Onboarding	
Name	Free text input
Gender	Male/female
Age	Free number input
Time of contact	5.00-20.00 h, every full hour
Day 1 to 20	
Pain intensity (1 to 10)	1 to 3, 4 to 7, more than 7 (3 groups)
Pain duration (2 months to >5 years)	Up to 2 years, more than 2 years (2 groups)
Day 20 to 57	
Interest in dysfunctional cognitions	Choice of 7 modules (to choose a total of 3)
Interest in coping strategies	Choice of 11 modules (to choose a total of 3)
Day 57	
Appealing aspects	Free-text input
Aspects for improvement	Free-text input

Assessment of the SELMA Intervention

We conducted a randomized controlled pilot trial to assess the SELMA intervention. The clinical trial was approved by the Cantonal Ethics Committee of Zurich (KEK-ZH study protocol identifier Nr. 2017-02136) and was registered at the Swiss National Clinical Trial Portal (SNCTP000002712) and the WHO-accredited German Clinical Trials Register (DRKS00017147). Data protection requirements were fulfilled according to the KEK-ZH.

Sample Size Calculation

We estimated the sample size based on the primary outcome (pain-related impairment) for a linear mixed model (LMM) and on the basis of a repeated-measures analysis of variance (ANOVA). Consistent with previous studies [70-72], we assumed a small to medium effect size for the primary outcome,

pain-related impairment. Statistical power calculation using G*Power[3] software [73] revealed that a sample size of 68 would be sufficient with a power of .80 to detect a small to medium time-by-group interaction effect (Cohen *d*=0.35) for pain-related impairment with an alpha of .05 by applying a repeated-measures ANOVA (within-between interaction). To ensure a sufficient number of participants, we aimed for approximately 115 individuals with ongoing or cyclic pain.

Recruitment

Recruitment was carried out from March 2018 to August 2018 in the German-speaking part of Switzerland. We recruited potential participants by flyers posted on social media websites (Facebook, Twitter), pain-related websites and forums (SchmerzLOS e.V. [74], MyHandicap [75], Paincompanion [76]) or via the project webpage (Multimedia Appendix 5). We further asked some pain therapists, physiotherapists, and osteopathy therapists in the agglomeration of Zug (Switzerland) to publish a flyer in their waiting rooms (Multimedia Appendix 6). Interested individuals were directed to the project website containing information about the study, participation, and registration. During the subject acquisition phase, individuals were able to download the app via the project webpage or directly via App Store or Google Play Store.

Inclusion Criteria, Informed Consent, and Intervention Process

Because of the exploratory nature of this study and to reach a sufficient sample size, we did not focus on a particular type of pain but rather included anyone suffering from ongoing pain. Inclusion criteria were checked from within the mobile SELMA app after the download and included: age (>18 years), language (German-speaking), owning a smartphone with internet access, pain duration (a minimum of 2 months of ongoing or cyclic pain), and not suffering from an acute mental crisis. Electronic informed consent was obtained from all participants who fulfilled the inclusion criteria (T0 screening), and these confirmed individuals were then guided to the baseline screening (T1). All screenings were carried out in the app's inbuilt assessment questionnaire via LimeSurvey (V3.4, LimeSurvey Project, Hamburg, Germany). After completion of the screenings, participants were automatically randomized by the MobileCoach system to either the intervention or control group. After the 8-week intervention, the follow-up screening (T2) was conducted with the app's inbuilt questionnaire. Individuals from the control group were informed of their group allocation and were offered participation in the program after the waiting time. These users received motivational messages with a quotation every week, which only involved content unrelated to chronic pain (Multimedia Appendix 7). The postintervention screening was conducted after the 8-week waiting time within the app.

Measures

Primary Outcome: Pain-Related Impairment

An overview of the measures and time points of measurement are outlined in Multimedia Appendix 8.

We measured the primary outcome at baseline and postintervention using a 7-item subscale of the German version of the validated Brief Pain Inventory (BPI) [77]. On a numeric scale from 0 (no impairment at all) to 10 (greatest impairment), impairments in a person's everyday life (eg, activity, work, relations) are quantified. We calculated the grand mean to determine the level of pain-related impairment.

Secondary Outcomes

Chronic Pain

To measure the dimension chronic pain intensity at baseline and postintervention, we used the self-reported mixed scale from the German Pain Questionnaire [78] with a visual rating scale ranging from 0 (no pain) to 10 (strongest imaginable pain), measuring current, average over the last 4 weeks, and highest pain intensity. The grand mean was used for calculating the average intensity of chronic pain. In addition, the duration of chronic pain was measured with a 5-scale item from "2 to 6 months" up to "more than 5 years" [78]. We assessed the type of pain by a choice from predefined multiple-answer options: back pain, headache, facial pain, toothache, extremity pain, pelvic pain, chest pain, joint pain, tumor pain, neuralgia, whole-body pain, don't know, other. Finally, we measured the cause of pain by the following predefined multiple-answer options: accident, surgery, inflammation, physical overload, disease, tumor, physical decline, stress, psychological strain, don't know, other.

General Well-Being

We measured general well-being by the Marburger Screening for Habitual Well-being (MFHW) [79], which is a 7-item self-report questionnaire that can be scored on a 0 (not at all applicable) to 5 (totally applicable) scale. The maximum score yields 35 and means particularly good well-being. In this study, we used the grand mean of the 7 items to determine general well-being.

Intention of Behavior Change

To measure participants' intention to change their behavior, we analyzed the intention to change behavior according to the Health Action Process (HAPA) [80], which differentiates three stages of change: non-intenders, intenders, and actors. Participants had to state whether they had recently been using psychological techniques for pain treatment by choosing from 1 out of 5 possible answers (No – and I do not intend to do so, No – but I'm considering it: nonintender; No – but I have the intention to do so: intender; Yes – but it's not easy, Yes – and it's easy: actors).

Working Alliance

We measured the working alliance between the participant and the TBHC SELMA with a context-adapted German version of the Working Alliance Inventory-Short Revised (WAI-SR) [81]. The WAI-SR consists of 3 dimensions targeting attachment bond (eg, "SELMA and I respect each other"), goal agreement (eg, "SELMA and I are working toward mutually agreed-upon goals"), and task agreement (eg, "I believe that the approach to working with my problem is correct"). Each subscale has 4 items with answer options ranging from never (1) to always (7). The WAI-SR was employed at the end of the intervention to assess the perceptions of the participating individuals with

respect to SELMA over the course of the 8-week intervention. In addition, and owing to the importance of first impressions when interacting with a social actor (in this case SELMA), we also used the attachment bond dimension during the baseline questionnaire (ie, after SELMA has introduced herself on the first day of conversational turns).

Acceptance

We assessed the acceptance of the SELMA app by applying single-item measures from technology acceptance research [82,83]. In particular, single-item measures for perceived usefulness ("I think that this app is useful"), perceived ease of use ("The app is easy to use"), and perceived enjoyment ("I enjoyed using the app"), ranging from strongly disagree (1) to strongly agree (7), were employed to reduce the burden of participants. We also employed the net promoter score [84] to assess patient experience and satisfaction with the SELMA intervention ("How likely is it that you would recommend the SELMA app to people with persistent or recurring pain?"), ranging from not at all likely (0) to extremely likely (10). We assessed the duration of the intervention by the following question: "The duration of the intervention was:" with the following answer options: "too short," "just right," "too long." The number of messages was assessed with the following question and answer option: "The number of chat messages was": "too few," just right," "too many." Finally, we assessed the quality of the intervention content of a chat sequence with the questions: "Was the content of the chat messages sufficient?" and the answer options "not sufficient," "just right," "too detailed." The questions "What did you like most?" and "What would you like to see improved?" could be answered with free text. All of these measures were assessed at the end of the intervention.

Adherence

Adherence was measured by the ratio of conversations replied by participants and all conversations initiated by SELMA for the intervention and control group separately. We also assessed the relationship between the adherence ratios and study outcomes by linear regression analysis for the intervention group.

Finally, age, sex, and the highest level of education attained were assessed at baseline to describe the population of participants.

Data Analysis

Descriptive statistics were used to summarize the characteristics of the intervention and control groups at baseline. To analyze baseline differences for demographics, we applied *t* tests and Chi-square tests. We used an independent *t* test to analyze the bond scale of the WAI-SR for differences between groups. We applied qualitative content analysis [85,86] for answers to open questions (eg, suggestions for future improvements). In particular, we explored answer text thematically with an inductive approach by defining main themes and subthemes. The frequencies of these themes were compiled in a thematic map [87]. To analyze differences between groups for pain-related impairment, pain intensity, and general well-being, we used an LMM for all participants. The model was estimated by time (T1/T2), intervention (intervene/wait), and the interaction of time and intervention as fixed factors, and participants as a random factor. We also tested the impact of the participants' intention to change their behavior as well as the duration of pain using an LMM. The model was conducted with a restricted maximum-likelihood approach and unstructured covariance type. Missing values were replaced as missing at random with the intention-to-treat analyses. We used an LMM approach because of its ability to handle missing data and to control for type 1 error in the case of incomplete or missing data [88,89]. We calculated effect sizes for both groups (intervene/wait) on the basis of a paired sample *t* test, which compared the baseline and postintervention measures for pain-related impairment, pain intensity, and general well-being. We conducted an independent *t* test to compare adherence rates between groups, and a linear regression analysis was used for comparison of adherence rates and study outcomes. SPSS 25 (IBM Corp. Released 2017. IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY, USA) was used for all quantitative analyses.

Deviations from the Study Protocol

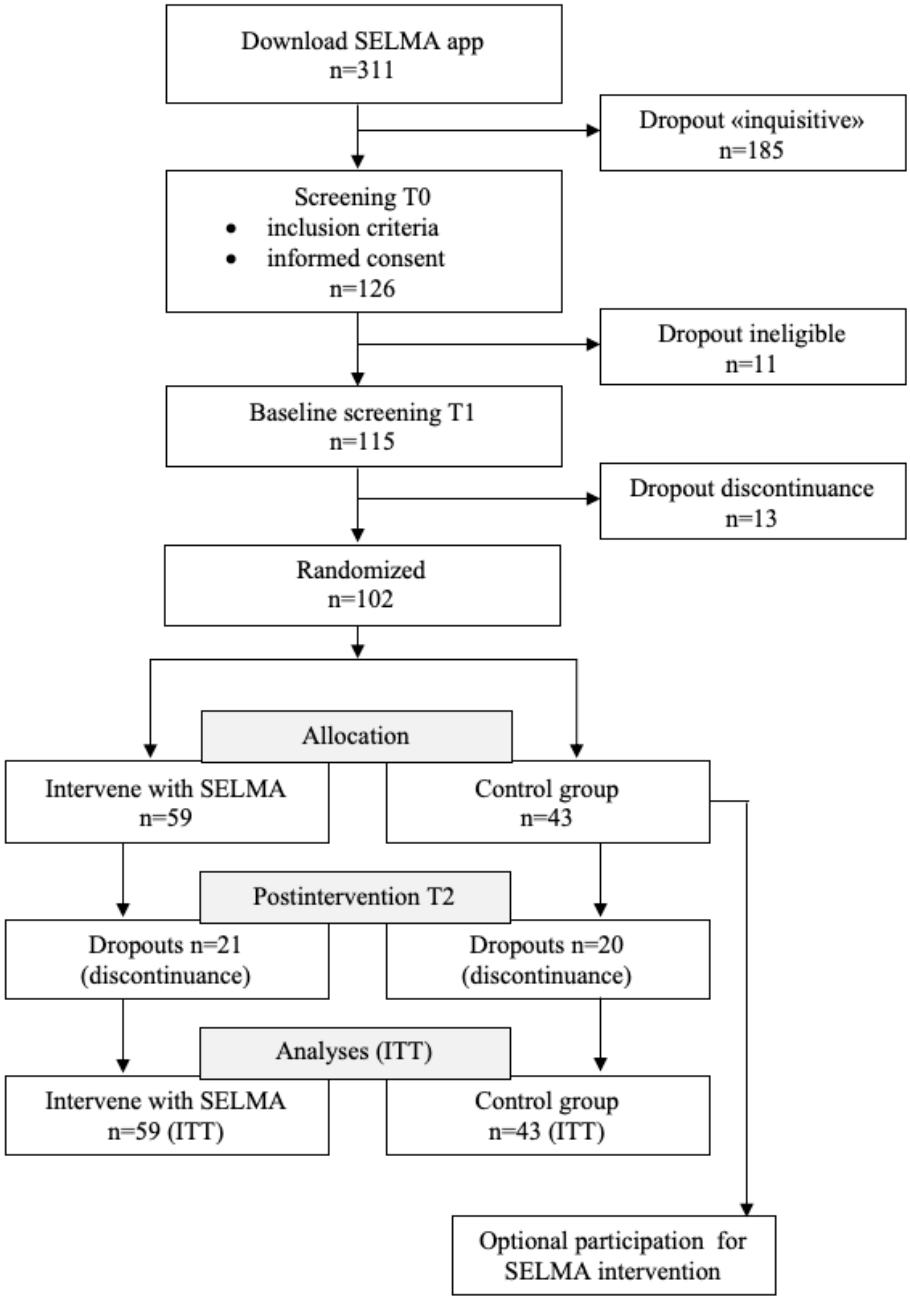
Instead of 3 variables for primary outcome (pain-related impairment, general well-being, pain intensity), we defined pain-related impairment as the primary outcome.

Results

Overview of Participation

A total of 311 downloads were made between May 5 and August 12, 2018, from which 126 individuals completed the T0 screening (inclusion criteria and informed consent), 11 of whom did not meet the criteria. The resultant sample of 115 started the T1 screening, and 13 individuals discontinued at this stage. After the onboarding process was completed, a total of 102 participants were randomized via MobileCoach automatically into either the intervention (59/102, 57.8%) or control (43/102, 42.2%) group. The participant flow chart is shown in Figure 3.

Figure 3. Participant flow chart. ITT: intention-to-treat.



As expected, the dropout rate during the onboarding process was high. This can be explained by the fact that the app was available for anyone in App Store (Apple) and Google Play Store (Android). Some people might have downloaded the app just out of curiosity and then deleted it soon after checking in. The overall dropout rate between the baseline and follow-up screening was 21/59 (36%) for the intervention group and 20/43 (47%) for the control group. We considered only participants who did not complete the T2 screening as dropouts.

Demographics and Baseline Scores

The demographic information and baseline scores on clinical variables for those who completed the baseline screening (N=102) are shown in Table 2. Participants were on average 43.77 (SD 12.72) years old, 82/102 (80.4%) were female, and 44/102 (43.1%) had a university degree. Overall, 55/102 (53.9%) of the participants had been suffering from pain for more than 5 years with a mean pain level of 5.61 (SD 1.66). Back pain was the most prevalent pain type in both groups (55/102, 53.9%), followed by headache (42/102, 41.2%). Means for pain-related impairment and for general well-being were 4.12 (SD 1.95) and 2.62 (SD 1.11), respectively. In terms of the intention to change behavior, 63/102 (61.0%) of the participants were classified as active, meaning they had been using psychological techniques (relaxation, mental distraction, mindfulness-based strategies, etc) recently, which were either easy or hard for them to use.

Table 2. Demographic and clinical variables of participants at baseline screening.

Variable	Control (n=43)	Intervention (n=59)	*P* value
Age (years), mean (SD)	44.88 (13.50)	42.97 (12.17)	.46
Gender, n (%)			**.07**
Male	12 (28)	8 (14)	
Female	31 (72)	51 (86)	
Education, n (%)			**.18**
Obligatory/High school	5 (12)	3 (5)	
Matriculation/A-Level	14 (32)	19 (32)	
Higher vocational training	9 (21)	8 (14)	
University	15 (35)	29 (49)	
Duration of pain, n (%)			**.69**
2 to 6 months	4 (9)	3 (5)	
6 months to 1 year	2 (5)	4 (7)	
1 to 2 years	7 (16)	9 (15)	
2 to 5 years	7 (16)	11 (19)	
Over 5 years	23 (54)	32 (54)	
Pain type, n (%)[a]			
Back pain	19 (44)	36 (61)	.09
Headache	17 (40)	25 (42)	.77
Extremities pain	17 (40)	20 (34)	.56
Neuralgia	14 (33)	19 (32)	.97
Joint pain	6 (14)	20 (34)	.02
Pelvic pain	3 (7)	9 (15)	.20
Whole-body pain	4 (9)	5 (9)	.89
Chest pain	3 (7)	3 (5)	.69
CRPS[b]	5 (12)	1 (2)	.04
Facial pain	1 (2)	4 (7)	.30
Unknown/other	1 (2)	2 (3)	.75
Cause of pain, n (%)[a]			
Overstraining	12 (28)	28 (48)	.05
Stress	9 (21)	23 (39)	.05
Accident	13 (30)	14 (24)	.46
Illness/migraine	9 (21)	15 (25)	.60
Degeneration	5 (12)	19 (32)	.02
Surgery	8 (19)	13 (22)	.68
Inflammation	9 (21)	6 (10)	.13
Psychogenic	6 (14)	9 (15)	.86
Unknown/other	8 (19)	6 (10)	.22
Gynecological	3 (7)	3 (5)	.69
Genetic	0 (0)	3 (5)	.13
Scale, mean (SD)			
Pain-related impairment (BPI[c])	4.21 (2.00)	4.06 (1.91)	.71

Variable	Control (n=43)	Intervention (n=59)	*P* value
Pain intensity (DSF[d])	5.72 (1.71)	5.52 (1.64)	.56
General well-being (MFHW[e])	2.65 (1.12)	2.61 (1.11)	.83
Intention to change behavior, n (%)[f]			**.99**
No, and I do not plan to	6 (14)	7 (12)	
No, but I think about it	8 (19)	10 (17)	
No, but I have the intention to	2 (5)	6 (10)	
Yes, but it is hard	18 (42)	26 (44)	
Yes, and it is easy	9 (21)	10 (17)	

[a]Note that multiple selection was possible.

[b]CRPS: complex regional pain syndrome.

[c]BPI: Brief Pain Inventory.

[d]DSF: Deutscher Schmerzfragebogen (German Pain Survey).

[e]MFHW: Marbuger Fragebogen zum habituellen Wohlbefinden (Marburger Screening for Habitual Well-being).

[f]Measured by the question: Have you recently used psychological techniques to treat your pain? This includes relaxation, mindfulness, distraction, scan thoughts, etc.

Effectiveness

Table 3 presents a comparison of the intervention and control groups from T1 to T2 for the primary outcome (pain-related impairment) as well as the secondary outcomes of pain intensity and general well-being. Means for the two groups were compared using a paired-sample *t* test. At T2, the intervention group showed significantly lower pain intensity compared to that at T1, (t_{37}=–2.8, *P*=.009) with a moderate to large effect size (*r*=0.42) and significantly higher well-being (t_{37}=2.41, *P*=.02) with a moderate effect size (*r*=0.37) according to Cohen [90].

Table 3. Results of a per-protocol paired-sample *t* test analysis comparing pre (T1) and post (T2) measures.

Outcome	Intervention (n=38)				Control (n=23)			
	T1, mean (SD)	T2, mean (SD)	*P* value	*r* [a]	T1, mean (SD)	T2, mean (SD)	*P* value	*r*
BPI[b]	4.18 (1.81)	3.98 (2.47)	.44	0.13	4.60 (2.00)	4.19 (2.05)	.28	0.23
DSF[c]	5.85 (1.57)	5.33 (1.70)	.009	0.42	5.88 (1.75)	5.42 (1.68)	.15	0.30
MFHW[d]	2.55 (1.10)	2.93 (1.11)	.02	0.37	2.50 (1.15)	2.54 (1.10)	.76	0.01

[a]*r*: effect size.

[b]BPI: Brief Pain Inventory.

[c]DSF: Deutscher Schmerzfragebogen (German Pain Survey).

[d]MFHW: Marbuger Fragebogen zum habituellen Wohlbefinden (Marburger Screening for Habitual Well-being).

The results of LMM analyses for the entire sample including covariates are given in Table 4. At T2, participants in the intervention group did not show significantly lower levels of pain-related impairment compared to those of the control group (t_{60}=0.42, *P*=.68).

Table 4. Results of the outcome intention-to-treat analysis using a linear mixed model.

Outcome	Estimate	SE	*P* value	95% CI
Pain-related impairment (BPI[a])				
Intercept	4.69	N/A[b]	N/A	N/A
Time[c]	–0.37	0.35	.29	–1.07 to 0.33
Group[d]	–0.13	0.38	.73	–0.89 to 0.63
Treatment[e]	0.18	0.44	.68	–0.70 to 1.07
HAPA[f]	0.38	0.15	.01	0.08 to 0.68
Duration of pain	–0.12	0.16	.45	–0.44 to 0.19
Pain intensity (DSF[g])				
Intercept	5.50	N/A	N/A	N/A
Time	–0.42	0.26	.10	–0.94 to 0.09
Group	0.20	0.33	.55	–0.86 to 0.46
Treatment	0.01	0.33	.97	–0.64 to 0.66
HAPA	0.33	0.12	.01	0.08 to 0.57
Duration of pain	0.06	0.13	.67	–0.21 to 0.32
General well-being (MFHW[h])				
Intercept	2.56	N/A	N/A	N/A
Time	0.01	0.17	.97	–0.34 to 0.35
Group	–0.05	0.22	.82	–0.49 to 0.49
Treatment	0.36	0.22	.11	–0.09 to 0.80
HAPA	–0.14	0.08	.09	–0.31 to 0.02
Duration of pain	0.02	0.09	.80	–0.15 to 0.20

[a]BPI: Brief Pain Inventory.

[b]Not applicable.

[c]Rate of improvement for both the intervention and control groups.

[d]Intervention or control group.

[e]Represented by the group-by-time interaction.

[f]HAPA: Health Action Process (intention to change behavior).

[g]DSF: Deutscher Schmerzfragebogen (German Pain Survey).

[h]MFHW: Marbuger Fragebogen zum habituellen Wohlbefinden (Marburger Screening for Habitual Well-being).

Intention to Change Behavior

As shown in Table 4, we found a significantly positive relation between the intention to change behavior and pain-related impairment (t_{98}=2.54, *P*=.01) as well as pain intensity (t_{98}=2.62, *P*=.01). The Pearson correlation coefficients between intention to change behavior and pain-related impairment (*r*=0.24; *P*=.02) as well as pain intensity (*r*=0.20; *P*=.05) revealed a significantly positive relation measured at T1. Descriptive analyses revealed that participants remained in the program by completing the T2 measure regardless of their classification as nonintenders and intenders at T1 (22/61).

Working Alliance

Table 5 shows the mean values of the intervention group for each of the 3 dimensions of the WAI-SR (attachment bond, goal agreement, task agreement) postintervention as well as the results of the mean comparison of attachment bond between the intervention and control group pre- and postintervention. At T2, groups differed significantly in their level of attachment bond ($t_{29.6}$=2.95, *P*=.005).

Table 5. Results of an independent t test for the subscale bond, means for subscales task and goal of the working alliance.

WAI-SR[a]	Preintervention (N=61)			Postintervention (N=61)		
	Intervention (n=38), mean (SD)	Control (n=23), mean (SD)	*P* value	Intervention (n=38), mean (SD)	Control (n=23), mean (SD)	*P* value
Total	N/A[b]	N/A	N/A	5.38	N/A	N/A
Bond	5.43 (1.27)	5.58 (1.44)	.69	5.89 (1.1)	4.51 (2.08)	.005
Task	N/A	N/A	N/A	4.95	N/A	N/A
Goal	N/A	N/A	N/A	5.3	N/A	N/A

[a]WAI-SR: Working Alliance Inventory-Short Revised (1-7 Likert scale).

[b]Not applicable.

Acceptance

We analyzed acceptance both descriptively (Table 6) and qualitatively (Figure 4 and Figure 5), demonstrating positive feedback from the majority of users. Overall, 24/38 (63%) of the participants fully agreed or agreed that they had fun using the app, 18/38 (47%) fully agreed or agreed that the app was useful, and 31/38 (84%) agreed that it was easy to use. The duration of the program was exactly right for 19/38 (50%) of the participants. The number of messages was too short for 28/38 (74%) and their content was not sufficiently profound for 29/38 (76%) of the participants. The net promoter score was high at 16 [91].

Table 6. Characteristics of app acceptability for the intervention group postintervention.

Characteristic	Mean (SD) (n=38)
The SELMA app was	
Enjoyable (1 to 7)	5.5 (1.45)
Easy to use (1 to 7)	6.34 (1.15)
Useful (1 to 7)	5.47 (1.41)
Intervention	
Duration[a]	1.87 (0.70)
Messages	
Number[b]	1.47 (0.76)
Content[c]	1.37 (0.71)

[a]1=too short, 2=just right, 3=too long.

[b]1=too seldom, 2=just right, 3=too often.

[c]1=not detailed enough, 2=just right, 3=too elaborate.

Figure 4. Thematic map and quotes from participants (righthand boxes) about positive aspects of the SELMA intervention. Note: numbers in brackets indicate the number of times the topic was mentioned by the participants.

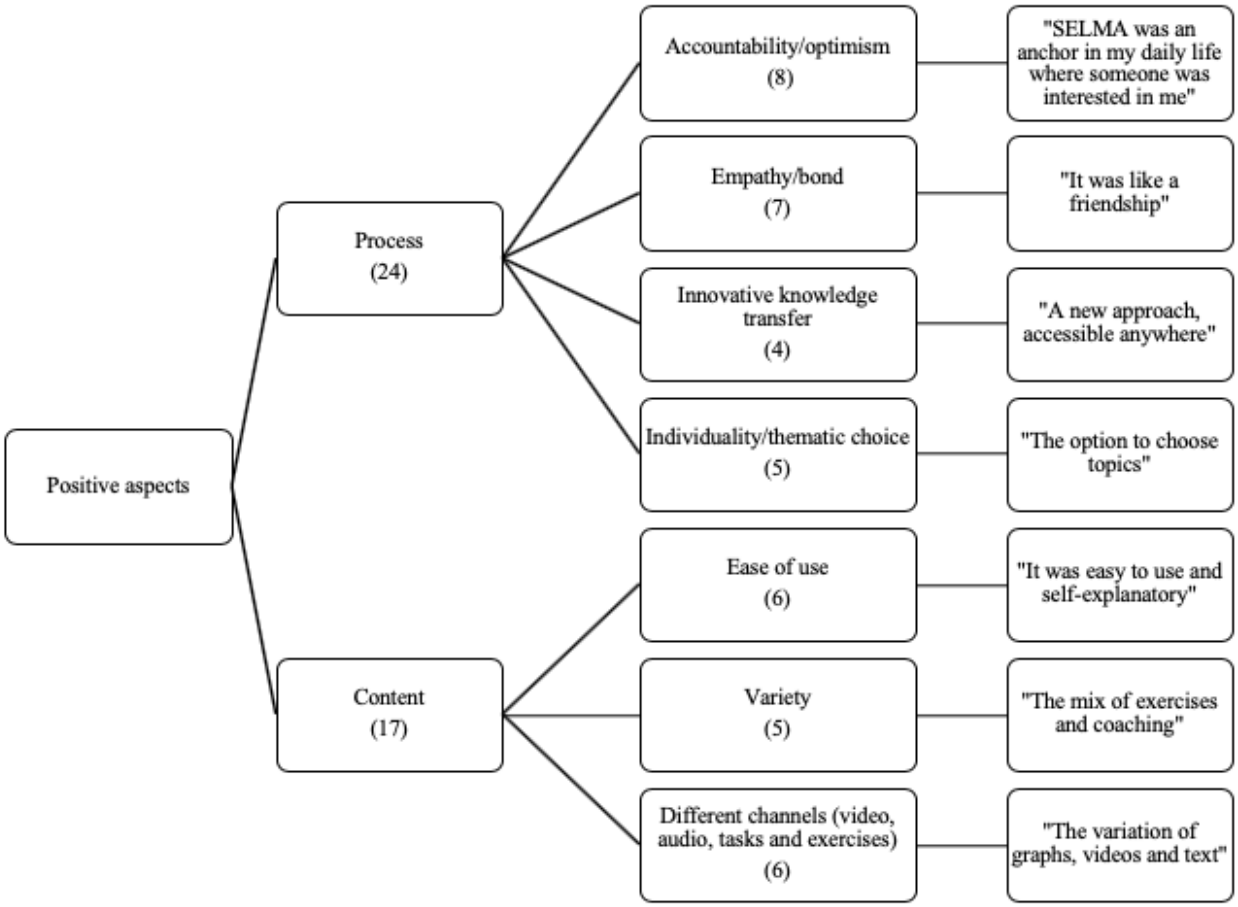


Figure 5. Thematic map and quotes from participants (righthand boxes) about negative aspects of the SELMA intervention. Note: Numbers in brackets indicate the number of times the topic was mentioned by the participants.

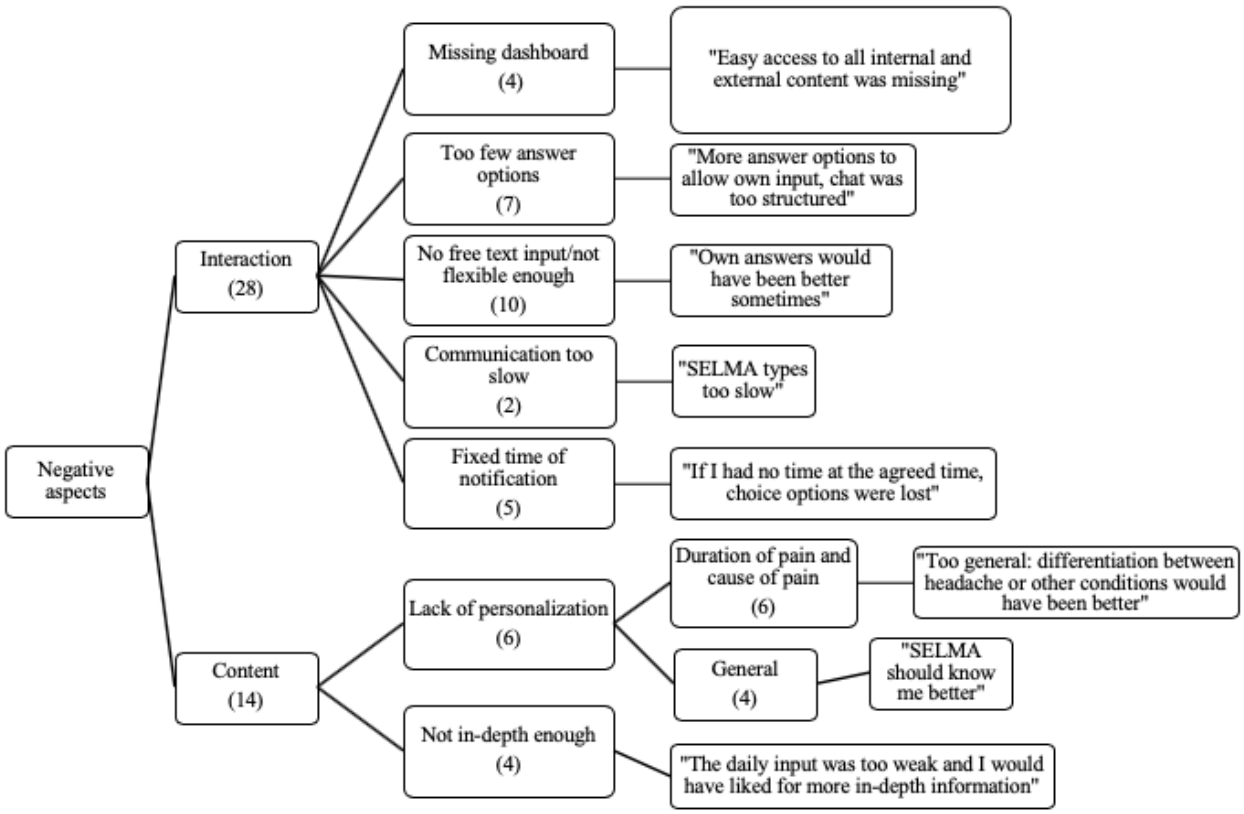


Figure 4 shows a thematic map of the participants' responses to the question "What did you like most about the app?" In the data-driven analyses, two major themes, process and content, emerged with respect to the positive aspects of the app. Regarding the process, accountability and optimism were stated most frequently (n=8), followed by empathy/bond (n=7) and ease of use and different channels (n=6). Individuality and variety were mentioned 5 times, followed by innovative knowledge transfer (n=4).

Figure 5 shows a thematic map of the participants' responses to the question "What should be improved in the app?" *Interaction* and *content* were revealed as the main themes. Within interaction, 10 users criticized the lack of free text answers and an overly static flow of interaction. Other aspects criticized included insufficient answer options (n=7), the fixed time point of notification (n=5), a missing dashboard (n=4), and the slow communication style (n=2). The lack of personalization was further split into the subthemes *duration and type of pain* (n=6) as well as *general* (n=4). Information did not go deeply enough for some users (n=4).

Adherence

The average adherence ratio was 71% (SD 20%) with 200 (SD 58.45) conversations initiated by SELMA in the intervention group. Participants of the control group responded with an adherence ratio of 60% (SD 27%) with 177 (SD 87.27) conversations initiated by SELMA. Multimedia Appendix 9 provides an overview of the number of conversations initiated by SELMA and responded by participants. There was no significant difference in adherence ratios between the intervention and control groups (t_{37}=1.81, *P*=.08). Multimedia Appendix 10 shows scatter plots of the relation of adherence ratios and outcomes. Linear regression analyses showed no significant relation between adherence ratios of the intervention group and the primary outcome pain-related impairment (beta=.18, SE 1.3, *P*=.29), and the secondary outcomes pain intensity (beta=–.06, SE 0.94, *P*=.71) and general well-being (beta=.22, SE 0.77, *P*=.19).

Discussion

Principal Findings

To the best of our knowledge, this is the first randomized controlled trial of a fully automated, unguided, text-based conversational agent designed for patients with chronic pain. The first goal of our study was to describe the design and implementation of SELMA. We clearly demonstrated that a TBHC-based intervention can be designed and implemented for chronic pain sufferers. The second goal was to explore whether a TBHC based on CBT could help users to better self-manage their pain over a period of 8 weeks. A comparison of baseline and postintervention measures of an 8-week trial for the intervention group revealed a small effect for pain-related impairment, strong effect for pain intensity, and medium effect for general well-being. However, the control group showed similar effects. LMM analyses showed that the intervention group did not differ significantly from the control group with regard to the primary outcome, pain-related impairment.

Moreover, we found a significant relationship between participants' intention to change behavior and both pain-related impairment as well as pain intensity. It seems plausible that people with a higher level of pain and pain-related impairment are more likely to show an increased intention to change their behavior and adopt new coping strategies in their pain self-management. Pain sufferers are more likely to change their behavior when their psychological strain increases. Additionally, participants who had no intention to change their behavior (measured at baseline) were participating and completing the intervention.

We also investigated the acceptance of the SELMA app as a secondary outcome. In general, participants enjoyed using the app, found it useful, and would recommend it to other people suffering from pain. Concerning the setting, half of the participants found the duration of the intervention to be adequate, and about the same number of users would have preferred a longer or shorter program duration. Over 70% of the users would have preferred a greater number of chat messages. Moreover, the content of messages was not extensive enough for more than 70% of the participants. Qualitative analyses revealed an insufficient degree of personalization, and many participants would have preferred to have more detailed or in-depth information about the relation of pain and behavior as well as coping strategies.

We measured adherence by assessing the ratio of conversations replied by participants and all conversations initiated by SELMA, revealing an average adherence ratio of 71% with 200 conversations initiated by SELMA; the intervention and control groups did not differ significantly. We found no significant relationship between adherence and study outcomes, although positive trends were observed. The drop-out rate was 21/59 (36%) for the intervention group and 20/43 (47%) for the control group.

Participants' qualitative feedback indicated that many users valued the reliable and empathic aspects of interaction. This shows that they perceived a sense of interpersonal closeness with the chatbot. Results from the bond scale of the WAI-SR [25] confirmed these statements; participants in the intervention group reported significantly higher values at the follow-up measure compared to those reported at the baseline measure. By contrast, many participants wanted the option to enter more free text, feeling that the interaction was too static and not flexible. This indicates a desire of participants to interact with a chatbot in the same way they do with humans, and supports the theory of media equation, which claims that people tend to treat computers or other media as if they were real people [92] and perceive them as social actors [28].

Comparison With Prior Work

Interventions based on a conversational agent have been recently deployed in the health sector. A recent study [93] showed that conversational agent-based interventions target neurological disorders [94-96], addictions [97,98], mental-physical wellness [99,100], nutritional metabolic disorders [101-103], and sexually transmitted disease [104]. To better compare or replicate studies, results should be reported consistently, for example according to the Consolidated Standards of Reporting Trials of electronic and mobile health applications [105]. In this study, conversational agents were used to monitor health condition, form attitudes toward health behavior, and finally intervene on cognitive or affective processes and health behavior. Scalability, personalization, consumability, asynchronicity, and anonymity were identified to be technical features of the greatest interest. Against this background, SELMA targets cognition, affect, attitude, and health behavior by using personalization. An intervention similar to SELMA [35] showed that a conversational agent in combination with medical group visits was able to reduce stress. Other related studies aim at mental-physical wellness. In one trial [99], mood was improved with a moderate effect size among users with high engagement. Another study [106] reported small to large effect sizes for improved well-being and stress among users with high engagement.

Moreover, conversational agent interventions seem to be well accepted by participants [35,101,106-108]. A meta-analysis showed that most interventions address mental health and demonstrate effectiveness [34]. Another review found that conversational agents are mainly implemented for healthy individuals rather than for those with chronic conditions [109]. Our study may contribute to closing this gap by supporting individuals suffering from chronic pain.

A more technical study categorized conversational agents regarding interaction paradigms, system architectures, and forms of dialog design [110]. This study also outlined that personalized interventions are a future challenge as are elderly populations because enhanced data are required to sustain improved support. Another study focused on conversational agents with unconstrained natural language input and found it to be an emerging field of research, but included studies that rarely evaluated efficacy or safety and were mainly quasiexperimental [111].

Furthermore, a review reported that conversational agent-based personalized interventions improved satisfaction, engagement, and dialog quality, and that outcome evaluations were often neglected [112]. These findings are promising, but because of limited evidence [113], further research is needed to identify appropriate design features for conversational agents that support pain self-management. SELMA offers personalization in various ways such as by offering the self-selection of intervention components. Qualitative analyses of SELMA revealed that participants appreciated this self-selection autonomy but wished for more personalization such as differentiation among types of pain.

Finally, there is no consensus on adherence measures for digital behavior change interventions. The ability to measure usage behavior of individuals, operationalization of intended use, and justification of intended use are deemed to be vital to measure adherence. An important aspect is to keep the goal of the app and the desired outcome in mind [114,115]; for example, assessing the time that participants spent on offline engagement with exercises, their motivation to exercise and engage with the intervention, or assessing specific elements of the intervention that were helpful to participants. Further, enjoyment of interaction with the digital coach and establishment of a working

alliance might contribute to adherence, as studies have shown establishment of a working alliance between humans and health care chatbots [22,106].

Limitations and Suggestions for Future Work

The current study has several limitations with respect to the generalizability of the results. First, as this was an exploratory pilot study, a heterogeneous sample was recruited and no follow-up data were collected to test long-term efficacy. The small sample size did not allow for further analyses such as comparing different types of pain. The heterogeneous sample also impeded establishing a definition of adequate psychoeducation. Future work should consist of larger and more homogeneous samples, along with a follow-up measure to investigate whether outcomes are persistent. This program was personalized by participants' interest and prior knowledge. However, future programs should expand on personalization of the content by processing more pain-related personal data.

Second, the majority of users were middle-aged women. This limiting factor was also found in several other studies [14,17,116]. Even though chronic pain is more common in women than in men of Western countries [4,117], this does not fit with the distribution in this trial, and future work should motivate more men to participate in technology-driven pain management approaches. The majority of participants had a university degree, which is also a characteristic reported by other studies [7,21]. Only 6% of the users had a migration background, which does not reflect the proportion in Swiss society (37%) [118]. Perhaps immigrants have reduced access to digital interventions due to language or cultural barriers, which should be considered in future research.

Third, some participants might have inadvertently missed the timeframe set to self-select modules, resulting in SELMA selecting predefined modules. It is possible that these default modules did not correspond to the participants' preferences and may have reduced their engagement. The default choice might have had a negative impact on the efficacy of the present study. Further research based on automated dialog systems should have flexible timepoints of communication and incorporate engagement analyses to find out more about when and for what reasons participants do not engage.

Fourth, due to technical problems, interaction with the chatbot was interrupted for some participants. The tracking of affected users was technically impossible, and some may have quit the program during this time. This factor limits an analysis of efficacy.

Fifth, in this trial, the participants had limited options to enter free text, which was criticized by many participants. Future programs should have a combination of predefined answers and free-text inputs to give users autonomy where needed. In addition, participants did not receive feedback on their progress. Individualized progress feedback is one of the most commonly used change techniques in smartphone-based health interventions [119], and helps users focus on their discrepancy awareness. Such feedback should be implemented in future programs so that participants can check to see if they are already approaching their set goals.

Finally, the control group underwent the onboarding process as well and received weekly text messages with quotations from the chatbot for ethical reasons. Because of this interaction with the chatbot, participants from this group might have started to actively challenge their pain self-management and therefore showed positive results on the primary outcomes. This might be due to the self-selecting bias that arises when individuals select themselves into a group or intervention. These individuals are particularly interested in the subject and cannot be considered to be a representative sample, as they might have different pre-existing characteristics [120]. The choice of this form of control group may have contributed to the improved levels of perceived pain-related impairment, pain intensity, and well-being.

Conclusions

The results of the present work must be interpreted with care and the findings need to be replicated. Nonetheless, our study clearly illustrates that a TBHC-based intervention can be designed and thus offers valuable information for future program adaptions. Our findings can help to design future studies with a larger and more homogeneous sample focusing on intentional behavior change and working alliance. It is also important to further examine the usage of and adherence to digital coaching programs. Chronic pain has a high impact on personal well-being and on health costs, and innovative treatment approaches are needed. A fully automated TBHC designed to guide self-management of chronic pain could have the potential to deliver a low-threshold CBT-based coaching program for people suffering from chronic pain.

Acknowledgments

The authors would like to thank Pathmate Technologies AG for providing technical development and support of the mobile SELMA app.

Conflicts of Interest

TK is affiliated with the Center for Digital Health Interventions (www.c4dhi.org), a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St Gallen, which is funded in part by the Swiss health insurer CSS. TK is also cofounder of Pathmate Technologies, a university spinoff company that creates and delivers digital clinical pathways and has used the open source MobileCoach platform for that purpose as well.

Multimedia Appendix 1
Overview of weekly core themes and tasks to complete.
[PDF File (Adobe PDF File), 133 KB - mhealth_v8i4e15806_app1.pdf]

Multimedia Appendix 2
Sympathy elements in the conversation between SELMA and a participant.
[PDF File (Adobe PDF File), 448 KB - mhealth_v8i4e15806_app2.pdf]

Multimedia Appendix 3
Examples of interaction with SELMA for the intervention group.
[PDF File (Adobe PDF File), 5969 KB - mhealth_v8i4e15806_app3.pdf]

Multimedia Appendix 4
Video with various examples of interactions with SELMA.
[MP4 File (MP4 Video), 22643 KB - mhealth_v8i4e15806_app4.mp4]

Multimedia Appendix 5
Project website.
[PDF File (Adobe PDF File), 5231 KB - mhealth_v8i4e15806_app5.pdf]

Multimedia Appendix 6
Flyer recruitment.
[PDF File (Adobe PDF File), 373 KB - mhealth_v8i4e15806_app6.pdf]

Multimedia Appendix 7
Example of an interaction with SELMA for the control group.
[PDF File (Adobe PDF File), 5377 KB - mhealth_v8i4e15806_app7.pdf]

Multimedia Appendix 8
List of self-reported screening measures.
[PDF File (Adobe PDF File), 57 KB - mhealth_v8i4e15806_app8.pdf]

Multimedia Appendix 9
Adherence illustrated by number of conversations.
[PDF File (Adobe PDF File), 77 KB - mhealth_v8i4e15806_app9.pdf]

Multimedia Appendix 10
Scatter plots of adherence and outcomes.
[PDF File (Adobe PDF File), 136 KB - mhealth_v8i4e15806_app10.pdf]

Multimedia Appendix 11
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 683 KB - mhealth_v8i4e15806_app11.pdf]

References

1. Schmidt CO, Raspe H, Pfingsten M, Hasenbring M, Basler HD, Eich W, et al. Back pain in the German adult population: prevalence, severity, and sociodemographic correlates in a multiregional survey. Spine 2007 Aug 15;32(18):2005-2011. [doi: 10.1097/BRS.0b013e318133fad8] [Medline: 17700449]
2. Reid KJ, Harker J, Bala MM, Truyers C, Kellen E, Bekkering GE, et al. Epidemiology of chronic non-cancer pain in Europe: narrative review of prevalence, pain treatments and pain impact. Curr Med Res Opin 2011 Feb;27(2):449-462. [doi: 10.1185/03007995.2010.545813] [Medline: 21194394]
3. Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: results of an Internet-based survey. J Pain 2010 Nov;11(11):1230-1239. [doi: 10.1016/j.jpain.2010.07.002] [Medline: 20797916]
4. Schmidt CO, Chenot JF, Kohlmann T. Epidemiologie und gesundheitsökonomische Aspekte des chronischen Schmerzes. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2011.

5. Roditi D, Robinson ME. The role of psychological interventions in the management of patients with chronic pain. Psychol Res Behav Manag 2011;4:41-49. [doi: 10.2147/PRBM.S15375] [Medline: 22114534]
6. Gatchel RJ, Okifuji A. Evidence-based scientific data documenting the treatment and cost-effectiveness of comprehensive pain programs for chronic nonmalignant pain. J Pain 2006 Nov;7(11):779-793. [doi: 10.1016/j.jpain.2006.08.005] [Medline: 17074616]
7. Vugts MAP, Joosen MCW, van der Geer JE, Zedlitz AMEE, Vrijhoef HJM. The effectiveness of various computer-based interventions for patients with chronic pain or functional somatic syndromes: A systematic review and meta-analysis. PLoS One 2018;13(5):e0196467 [FREE Full text] [doi: 10.1371/journal.pone.0196467] [Medline: 29768436]
8. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999 Mar;80(1-2):1-13. [doi: 10.1016/s0304-3959(98)00255-3] [Medline: 10204712]
9. Williams ACDC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012;11:CD007407. [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
10. Frettlöh J, Hermann C. Kognitiv-behaviorale Therapie. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:349-371.
11. AGS Panel on Persistent Pain in Older Persons. The management of persistent pain in older persons. J Am Geriatr Soc 2002 Jun;50(6 Suppl):S205-S224. [doi: 10.1046/j.1532-5415.50.6s.1.x] [Medline: 12067390]
12. Schmidt CO, Chenot J, Pfingsten M, Fahland RA, Lindena G, Marnitz U, et al. Assessing a risk tailored intervention to prevent disabling low back pain--protocol of a cluster randomized controlled trial. BMC Musculoskelet Disord 2010 Jan 05;11:5 [FREE Full text] [doi: 10.1186/1471-2474-11-5] [Medline: 20051119]
13. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA 1998 Jul 08;280(2):147-151. [doi: 10.1001/jama.280.2.147] [Medline: 9669787]
14. Garg S, Garg D, Turin TC, Chowdhury MFU. Web-Based Interventions for Chronic Back Pain: A Systematic Review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
15. Turk DC, Rudy TE. Neglected factors in chronic pain treatment outcome studies--referral patterns, failure to enter treatment, and attrition. Pain 1990 Oct;43(1):7-25. [doi: 10.1016/0304-3959(90)90046-g] [Medline: 2277718]
16. Turk DC, Rudy TE. Neglected topics in the treatment of chronic pain patients--relapse, noncompliance, and adherence enhancement. Pain 1991 Jan;44(1):5-28. [doi: 10.1016/0304-3959(91)90142-k] [Medline: 2038489]
17. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital Support Interventions for the Self-Management of Low Back Pain: A Systematic Review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
18. Lin J, Baumeister H. Internet-und Mobilebasierte Interventionen in der Psychotherapie. Public Health Forum 2015;23(3):176-179. [doi: 10.1515/pubhef-2015-0063]
19. Andersson G. Internet-Delivered Psychological Treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
20. Enock P, McNally R. How mobile apps and other web-based interventions can transform psychological treatment and the treatment development cycle. Behav Ther 2013;36(3):56-66.
21. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
22. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
23. Castonguay LG, Constantino MJ, Holtforth MG. The working alliance: Where are we and where should we go? Psychotherapy 2006;43(3):271-279. [doi: 10.1037/0033-3204.43.3.271] [Medline: 22122096]
24. Flückiger C, Del Re AC, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. J Couns Psychol 2012 Jan;59(1):10-17. [doi: 10.1037/a0025749] [Medline: 21988681]
25. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Counsel Psychol 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
26. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
27. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000 Jun;68(3):438-450. [Medline: 10883561]
28. Nass C, Steuer J, Tauber E. Computers Are Social Actors. : ACM; 1994 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 24, 1994; Boston, MA, USA p. 72-78. [doi: 10.1145/191666.191703]
29. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
30. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun 01;12(2):293-327. [doi: 10.1145/1067860.1067867]

31. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al. DesignEvaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. In: Maedche A, vom Brocke J, Hevner A, editors. Designing the Digital Transformation. Cham: Springer International Publishing; 2017:485-489.
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Herring SC. Slouching Toward the Ordinary: Current Trends in Computer-Mediated Communication. New Media Soc 2016 Jun 30;6(1):26-36. [doi: 10.1177/1461444804039906] [Medline: 22973420]
34. Ma T, Sharifi H, Chattopadhyay D. Virtual Humans in Health-Related Interventions: A Meta-Analysis. : ACM Press; 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-6. [doi: 10.1145/3290607.3312853]
35. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting Group Medical Visits with Conversational Agents for Stress Management Behavior Change. In: de Vries PW, editor. Persuasive Technology: Development And Implementation Of Personalized Technologies To Change Attitudes And Behaviors. PERSUASIVE 2017. Lecture Notes In Computer Science, vol 10171. Cham: Springer; 2020:55-67.
36. Kowatsch T, Nissen M, Rüegger D, Flückiger C, Allemand M, von Wangenheim WF. The Impact of Interpersonal Closeness Cues in Text-based Healthcare Chatbots on Attachment Bond and the Desire to Continue Interacting. 2018 Presented at: Twenty-Sixth European Conference on Information Systems (ECIS2018); June 23-28, 2018; Portsmouth p. 1-13.
37. Kowatsch T, Nissen M, Chen-Husan I, Rüegger D, Volland D, Filler A, et al. Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. In: Persuasive Embodied Agents for Behavior Change. 2017 Presented at: 17th International Conference on Intelligent Virtual Agents; August 27-30, 2017; Stockholm. [doi: 10.1007/978-3-319-59144-5_36]
38. von Wachter M. Chronische Schmerzen: Selbsthilfe und Therapiebegleitung, Orientierung für Angehörige und konkrete Tipps und Fallbeispiele. Dordrecht: Springer; 2014.
39. Sendera M, Sendera A. Chronischer Schmerz Schulmedizinische, komplementärmedizinische und psychotherapeutische Aspekte. Vienna: Springer; 2015.
40. Kröner-Herwig B, Frettlöh J. Behandlung chronischer Schmerzsyndrome: Plädoyer für einen interdisziplinären Therapieansatz. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:277-301.
41. Kröner-Herwig B, Lautenbacher S. Schmerzmessung und klinische Diagnostik. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:215-238.
42. Kröner-Herwig B. Schmerz als biopsychosoziales Phänomen – eine Einführung. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:3-16.
43. Nobis HG, Rolke R. Herausforderung Schmerz. In: Nobis HG, Rolke R, Graf-Baumann T, editors. Schmerz – Eine Herausforderung. Berlin, Heidelberg: Springer; 2016:1-15.
44. von Wachter M, Hendrischke A. Psychoedukation bei chronischen Schmerzen. Berlin, Heidelberg: Springer; 2016.
45. Sandkühler J. Schmerzgedächtnisntstehung, Vermeidung und Löschung. Dtsch Artzebl 2001;98(41):2725-2730 [FREE Full text]
46. Flor H. Neurobiologisch e und psychobiologische Faktoren der Chronifizierung und Plastizität. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:87-101.
47. Pfingsten M, Hildebrandt J. Rückenschmerzen. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:531-553.
48. Hasenbring M, Korb J, Pfingsten M. Psychologische Mechanismen der Chronifizierung – Konsequenzen für die Prävention. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:115-131.
49. Richter J. Schmerzen verlernen. Berlin, Heidelberg: Springer; 2013.
50. Lüking M, Martin A. Entspannung, Im agination, Biofeedback und Meditation. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:303-324.
51. Dobe M, Zernikow B. Therapie von Schmerzstörungen im Kindes- und Jugendalter in Manual für Psychotherapeuten, Ärzte und Pflegepersonal. Berlin, Heidelberg: Springer; 2013.
52. Diezemann A, Korb J. Akzeptanz- und Commitment- Therapie. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P, editors. Schmerzpsychotherapie. Berlin, Heidelberg: Springer; 2017:337-348.
53. Glier B. Chronischen Schmerz bewältigen: verhaltenstherapeutische Schmerzbehandlung. 4th ed. Stuttgart: Klett-Cotta; 2016.
54. Wengenroth M. Therapie-Tools Akzeptanz- und Commitmenttherapie (ACT): mit E-Book inside und Arbeitsmaterial. 2nd ed. Weinheim Basel: Beltz; 2017.
55. 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. : IEEE; 2015 Apr Presented at: 2015 Wireless Telecommunications Symposium (WTS); 2015; New York City, NY, USA p. 1-6. [doi: 10.1109/wts.2015.7117255]
56. MobileCoach. URL: https://www.mobile-coach.eu/ [accessed 2020-02-14]
57. Beun RJ, Fitrianie S, Griffioen-Both F, Spruit S, Horsch C, Lancee J, et al. Talk and Tools: the best of both worlds in mobile user interfaces for E-coaching. Pers Ubiquit Comput 2017 May 19;21(4):661-674. [doi: 10.1007/s00779-017-1021-5]
58. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: The working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
59. Aron A, Aron EN, Smollan D. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. J Person Soc Psychol 1992;63(4):596-612. [doi: 10.1037/0022-3514.63.4.596]
60. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
61. Bickmore T, Picard R. Towards caring machines. In: Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04.: ACM Press; 2004 Presented at: CHI '04 Conference on Human Factors and Computing Systems; April 24-29, 2004; Vienna, Austria p. 1489-1492. [doi: 10.1145/985921.986097]
62. Brave S, Nass C, Hutchinson K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Hum Comput Stud 2005 Feb;62(2):161-178. [doi: 10.1016/j.ijhcs.2004.11.002]
63. Partala T, Surakka V. The effects of affective interventions in human–computer interaction. Interact Comput 2004 Apr;16(2):295-309. [doi: 10.1016/j.intcom.2003.12.001]
64. Liu B, Sundar SS. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
65. Van der Zwaan JM, Dignum V. Robin, an empathic virtual buddy for social support. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems.: International Foundation for Autonomous Agents and Multiagent Systems; 2014 Presented at: 2013 International Conference on Autonomous Agents and Multi-agent Systems; 2014; Richland, SC p. 1413-1414.
66. Hone K. Empathic agents to reduce user frustration: The effects of varying agent characteristics. Interact Comput 2006 Mar;18(2):227-245. [doi: 10.1016/j.intcom.2005.05.003]
67. Haug S, Paz CR, 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 Dec;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
68. 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 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
69. Cassell J, Bickmore T. Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. User Model User Adapt Interact 2003;13(1):89. [doi: 10.1023/A:1024026532471]
70. Lin J, Paganini S, Sander L, Lüking M, Ebert D, Buhrman M, et al. An Internet-Based Intervention for Chronic Pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688. [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]
71. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
72. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
73. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
74. SchmerzLOS e.V. URL: https://www.schmerzlos-ev.de/ [accessed 2020-02-14]
75. MyHandicap. URL: http://www.myhandicap.ch [accessed 2020-02-14]
76. Paincompanion. URL: https://www.paincompanion.com [accessed 2020-02-14]
77. Radbruch L, Loick G, Kiencke P, Lindena G, Sabatowski R, Grond S, et al. Validation of the German version of the Brief Pain Inventory. J Pain Symptom Manage 1999 Sep;18(3):180-187 [FREE Full text] [doi: 10.1016/s0885-3924(99)00064-0] [Medline: 10517039]
78. Pfingsten M, Nagel B, Erich O, Seemann H, Lindena G, Korb J. Handbuch Deutscher Schmerz-Fragebogen. Deutschland: Deutsche Schmerzgesellschaft e.V; 2015.
79. Basler H, Herda C, Scharfenstein A. Marburger Fragebogen zum habituellen Wohlbefinden. Diagnostische Verfahren zu Lebensqulität und Wohlbefinden 2003;13:385-391. [doi: 10.1007/s004820050216]
80. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]

81. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
82. van der Heijden H. User Acceptance of Hedonic Information Systems. MIS Quart 2004;28(4):695-704. [doi: 10.2307/25148660]
83. Kamis A, Koufaris M, Stern T. Using an Attribute-Based Decision Support System for User-Customized Products Online: An Experimental Investigation. MIS Quart 2008;32(1):159-177. [doi: 10.2307/25148832]
84. Krol MW, de Boer D, Delnoij DM, Rademakers JJDJM. The Net Promoter Score – an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109. [doi: 10.1111/hex.12297] [Medline: 25345554]
85. Mayring P. Qualitative Content Analysis. Forum Qual Soc Res 2000;1(2):20. [doi: 10.17169/fqs-1.2.1089]
86. Mayring P. Qualitative Inhaltsanalyse. In: Flick U, v.Kardoff E, Keupp H, Rosenstiel L, Wolff S, editors. Handbook Qualitative Forsch Grundlagen Konzepte Methoden Anwendungen München. Weinheim, Germany: Beltz; 1991:209-213.
87. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
88. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharm Stat 2001;11(1-2):9-21. [doi: 10.1081/BIP-100104194] [Medline: 11459446]
89. Siddiqui O, Hung HMJ, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat 2009 Feb 12;19(2):227-246. [doi: 10.1080/10543400802609797] [Medline: 19212876]
90. Cohen J. Statistical Power Analysis. Curr Dir Psychol Sci 2016 Jun 24;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
91. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AHRW, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014 May;96-B(5):622-628. [doi: 10.1302/0301-620X.96B5.32434] [Medline: 24788496]
92. Reeves B, Nass C. The media equation: how people treat computers, television, and new media like real people and places. 1. paperback ed. Stanford, CA: CSLI Publ; 1998.
93. Pereira J, Díaz O. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
94. Atay C, Ireland D, Liddle J, Wiles J, Vogel A, Angus D, et al. Can a Smartphone-Based Chatbot Engage Older Community Group Members? The Impact of Specialised Content. Alzheimers Dement 2016 Jul 01;12:P1005-P1006. [doi: 10.1016/j.jalz.2016.06.2070]
95. Griol D, Molina J. In: Bajo J, Hallenborg K, Pawlewski P, Botti V, Sánchez-Pi N, Duque Méndez ND, et al, editors. An Ambient Assisted Living Mobile Application for Helping People with Alzheimer. Cham: Springer International Publishing; 2015:3-14.
96. Beun R, Brinkman W, Fitrianie S, Griffioen-Both F, Horsch C, Lancee J, et al. Persuasive Technology. In: Meschtscherjakov A, De Ruyter B, Fuchsberger V, Murer M, Tscheligi M, editors. Improving Adherence in Automated e-Coaching. Cham: Springer International Publishing; 2016:276-287.
97. Dubosson F, Schaer R, Savioz R, Schumacher M. Going beyond the relapse peak on social network smoking cessation programmes: ChatBot opportunities. Swiss Med Informatics 2017 Sep 20;33:00397. [doi: 10.4414/smi.33.00397]
98. Lisetti C, Amini R, Yasavur U. Now All Together: Overview of Virtual Health Assistants Emulating Face-to-Face Health Interview Experience. Künstl Intell 2015 Mar 31;29(2):161-172. [doi: 10.1007/s13218-015-0357-0]
99. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
100. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth '17.: ACM Press; 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare – PervasiveHealth '17; 2017; Barcelona p. 261-265. [doi: 10.1145/3154862.3154914]
101. Kowatsch T, Nissen M, Chen-Husan S, Rüegger D, Volland D, Filler A, et al. Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. In: Persuasive Embodied Agents for Behavior Change. 2017 Presented at: 17th International Conference on Intelligent Virtual Agents; August 27-30, 2017; Stockholm. [doi: 10.3929/ethz-b-000218776]
102. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al. A design and evaluation framework for digital health interventions. Inform Technol 2017;61(5-6):253-263. [doi: 10.1515/itit-2019-0019]
103. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. : IEEE; 2018 Presented at: 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); January 12-15, 2018; Las Vegas, NV p. 1-5. [doi: 10.1109/ccnc.2018.8319283]
104. Kobori Y, Osaka A, Soh S, Okada H. MP15-03 Novel application for sexual transmitted infection screening with an AI chatbot. J Urol 2018 Apr;199(4S):e189-e190. [doi: 10.1016/j.juro.2018.02.516]

105. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
106. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
107. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729. [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
108. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a Web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
109. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the Great Absentee of Virtual Coaching in Medical Care: Scoping Review. J Med Internet Res 2019 Oct 01;21(10):e12805 [FREE Full text] [doi: 10.2196/12805] [Medline: 31573902]
110. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
111. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
112. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
113. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
114. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
115. Sieverink F, Kelders SM, van Gemert-Pijnen JEWC. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
116. Lin J, Baumeister H, Lücking M, Ebert D, Buhrman M, Andresson G. Wirksamkeit und Kosteneffektivität einer begleiteten und unbegleiteten online-basierten Akzeptanz und Commitment Therapie für chronische Schmerzenine drei-armige randomisierte, kontrollierte Studie. Freiburg: Albert-Ludwigs-Universität Freiburg; 2017.
117. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open 2016 Jun 20;6(6):e010364 [FREE Full text] [doi: 10.1136/bmjopen-2015-010364] [Medline: 27324708]
118. Bundesamt FSB. Schweizerische Eidgenossenschaft. 2017. Die Bevölkerung der Schweiz 2017 URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/stand-entwicklung/bevoelkerung.assetdetail.6606496.html [accessed 2019-07-17]
119. 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]
120. Lavrakas P. Encyclopedia Of Survey Research Methods. Thousand Oaks, CA: Sage Publications, Inc; 2008.

Abbreviations

ANOVA: analysis of variance
BPI: Brief Pain Inventory
CBT: Cognitive Behavior Therapy
CRPS: complex regional pain syndrome
DRKS: German Clinical Trial Register
DSF: Deutscher Schmerzfragebogen (German Pain Survey)
HAPA: Health Action Process
KEK-ZH: Kantonale Ethikkommission Zürich (Cantonal Ethics Committee of Zurich)
LMM: Linear Mixed Model
MFHW: Marburger Fragebogen zum habituellen Wohlbefinden (Marburger Survey of Habitual Well-being)
SELMA: painSELfMAnagement chatbot
TBHC: Text-based healthcare chatbot
WAI-SR: Working Alliance Inventory-Short Revised

WHO: World Health Organization

Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by T Janmohamed, A Kocaballi; comments to author 30.09.19; revised version received 01.12.19; accepted 26.01.20; published 03.04.20.

Please cite as:
Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T
A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e15806
URL: http://mhealth.jmir.org/2020/4/e15806/
doi:10.2196/15806
PMID:32242820

Original Paper

Treatment Adherence and Secondary Prevention of Ischemic Stroke Among Discharged Patients Using Mobile Phone- and WeChat-Based Improvement Services: Cohort Study

Yuanjin Zhang[1], MD; Dongsheng Fan[1], MD, PhD; Hong Ji[2], PhD; Shudong Qiao[3], MD; Xia Li[3], MD

[1]Neurology Department, Peking University Third Hospital, Beijing, China

[2]Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

[3]Neurology Department, Peking University Shougang Hospital, Beijing, China

Corresponding Author:
Dongsheng Fan, MD, PhD
Neurology Department
Peking University Third Hospital
49 North Garden Road
Haidian District
Beijing
China
Phone: 86 01082265694
Email: dsfan2010@aliyun.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/4/e19454/

Abstract

Background: Real-world studies have indicated that adherence is important for guaranteeing medication effectiveness. Few studies have tested the feasibility and efficacy of WeChat-based improvement services, via mobile phone, in secondary prevention-specific follow-up among discharged stroke patients.

Objective: We evaluated a quadruple-domain, WeChat-based service for ischemic stroke secondary prevention designed to improve treatment adherence of discharged patients. This service focuses on sending reminders for drug use, blood pressure recording, and glucose recording; it also records medication use. We compared the endpoint event rate between WeChat self-monitoring and traditional monitoring.

Methods: A cohort study was used to determine the feasibility of a physician-assisted, WeChat-based improvement service and follow-up self-monitoring platform for the secondary prevention of ischemic stroke. The platform was developed by the Peking University Third Hospital based on the information-motivation-behavioral skills model. The overall adherence rate was calculated as the proportion of medication doses verified via uploading. The ischemic endpoint event rate and medication noncompliance rate were compared between traditional prevention monitoring and WeChat self-monitoring. Factors influencing adherence were summarized.

Results: The 1-year follow-up event rate of the WeChat self-monitoring group was 11.9% (12/101), which was less than that of the traditional group (21/157, 13.4%). Compared with the traditional group, the risk ratio of the WeChat group was 0.983 (95% CI 0.895-1.080); this difference was not noted to be significant. The 1-year medication noncompliance ratio tended to be lower in the WeChat monitoring group (3/101, 3.0%) than in the traditional group (11/157, 7.0%; χ^2=1.9, *df*=1, *P*=.16). Of the platform registry participants, 89.7% (210/234: 167 hospital-based and 43 community-based participants) adhered to inputting information into WeChat for 8-96 weeks. The average adherence time was 16.54 (SD 0.80, range 2-24) months. The average decrease in adherence was 4 participants (1.1%) per month. Being a member of a community-based population was an influencing factor for good adherence at the 2-year follow-up (OR 2.373, 95% CI 1.019-5.527, *P*=.045), whereas transient ischemic attack was an influencing factor for poor adherence at the 2-year follow-up (OR 0.122, 95% CI 0.016-0.940, *P*=.04).

Conclusions: Use of WeChat self-monitoring showed a trend of increasing medication compliance and decreasing ischemic endpoint event rate compared with traditional monitoring. However, there were ceiling effects in the outcomes, and a relatively small sample size was used. Male participants displayed better adherence to WeChat self-monitoring. The community-based population displayed good adherence when using WeChat self-monitoring.

Trial Registration: ClinicalTrials.gov NCT02618265; https://clinicaltrials.gov/ct2/show/NCT02618265

(JMIR Mhealth Uhealth 2020;8(4):e16496) doi:10.2196/16496

KEYWORDS

stroke; secondary prevention; WeChat; self-monitoring

Introduction

Stroke is one of the most common causes of death and long-term disability worldwide and in China [1-3]. Recurrent ischemic stroke accounts for nearly 17.7% of strokes in China according to the China National Stroke Registry [4]. Secondary prevention of ischemic stroke, including medical therapy and healthy lifestyle control, has been demonstrated from quantitative development and qualitative change for effectively preventing recurrent stroke or ischemic events. The combination of antiplatelet agents, statins, and antihypertensive agents is an effective therapeutic strategy and is recommended by various guidelines [5-8]. Good compliance of continuous monitoring to ensure that blood pressure, blood glucose, and lipids—particularly low-density lipoprotein cholesterol—reach therapeutic standards is essential for effective medication treatment to obtain prevention benefits. However, a real-world study has indicated that adherence is important for guaranteeing medication effectiveness [9]. Several studies have demonstrated that nearly 30% of patients failed to follow doctors' instructions continuously [10,11]. Because of distinctions in the concept of health, incomplete health education, and unbalanced distribution of medical service resources, secondary prevention is not necessarily conducted by each institution continuously and may even be terminated by the patient. Thus, self-management directly supervised by hospital researchers after discharge is a potentially feasible and effective method for ensuring medication compliance and follow-up of endpoint events. It is widely recognized that using mobile phones can improve outreach and interactions with populations in need for health promotion and disease prevention [12,13]. WeChat is a simple, widespread communication app. Moreover, WeChat is cost-effective, less resource intensive, and more accessible than other communication apps; thus, enhancing adherence of prevention strategies seems feasible. The authors reviewed emerging technologies and methods that show promise in supporting and tracking medication compliance, then brought these together into a WeChat-based, follow-up self-monitoring platform that provided improvement services and was specific to ischemic stroke secondary prevention. The platform is part of the Peking University Third Hospital's (PUTH) Health Service Account—the official WeChat account of PUTH—Department of Neurology follow-up domain (ie, the name of the platform's gateway). Few studies have tested the feasibility of this system as a prevention-specific follow-up approach among discharged stroke patients. We hypothesized that WeChat-based adherence monitoring, endpoint submission, and stroke-specific self-management will improve medication compliance rates and decrease recurrent stroke events. The objectives of this study were to (1) quantify the degree to which discharged participants adhere to prescribed modules of uploading medication adherence information and outcomes, (2) compare the event rate between traditional and WeChat adherence monitoring, and (3) explain longitudinal adherence rates.

Methods

Participants

Participants were recruited at three offline sites. The PUTH and the Peking University Shougang Hospital (PUSH) enrolled 300 acute ischemic stroke patients. The inclusion criteria of the hospital-based population were as follows:

1. Older than 18 years of age and can read or write.
2. Diagnosis of cerebral infarction with evidence from computed tomography or magnetic resonance imaging without coma—score of 0-25 on the National Institutes of Health Stroke Scale—and anterior (ie, internal carotid, anterior cerebral artery, and middle cerebral artery) circulation ischemia.
3. Diagnosis of transient ischemic attack (TIA) (ie, anterior circulation).
4. Had large vessel atherosclerotic and small vessel disease subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.

The exclusion criteria of this population were as follows:

1. Dementia.
2. Evidence of hemorrhage on computed tomography or magnetic resonance imaging.
3. Hematological disorders.
4. Any clinically relevant arrhythmia on admission, including atrial fibrillation.
5. Any major concurrent illness, including severe cardiovascular disease, liver or renal failure, and malignancies.
6. Fever, hypoxia, alterations in consciousness, or any relevant hemodynamic compromise on admission.
7. Any sensitivity to aspirin or clopidogrel.
8. Any other doctor-defined criterion as not suitable for enrollment.

All patients who were admitted to the neurology ward received therapeutic treatments according to the American Heart Association and American Stroke Association 2013 guidelines for acute ischemic stroke administration and the 2014 guidelines for ischemic stroke secondary prevention. All participants who

met the inclusion criteria gave informed consent before enrollment and received routine health education about secondary prevention while being discharged. The inclusion criteria for adherence verification for the community-based population at high risk of stroke in the Huayuanlu Community Healthcare Service Center were history of stroke or TIA with no restriction of duration; the exclusion criteria were the same as those used for the hospital-based population.

WeChat-Based Improvement Services and Self-Monitoring Platform for Discharged Patients

We developed WeChat-based, medication-related modules regarding the need for compliance. This was in accordance with the clinician's recommendations for the patient and based on the individual medications and needs of the patient; thus, the app could function as a *personal trainer*. These modules provide people with timely reminders, according to their medication input dosage and frequency information; this occurs after collecting the uploaded prescription for secondary stroke prevention, which can be reinput and changed by the participant in accordance with the clinician's prescription modifications. The individual's characteristics (eg, disease-related records and monitoring-parameter levels) were recorded daily, automatically and manually. After certifying identification and receiving permission, the participating user was able to access the main menu of the WeChat-based platform and insert her or his information. The platform services were free for participants.

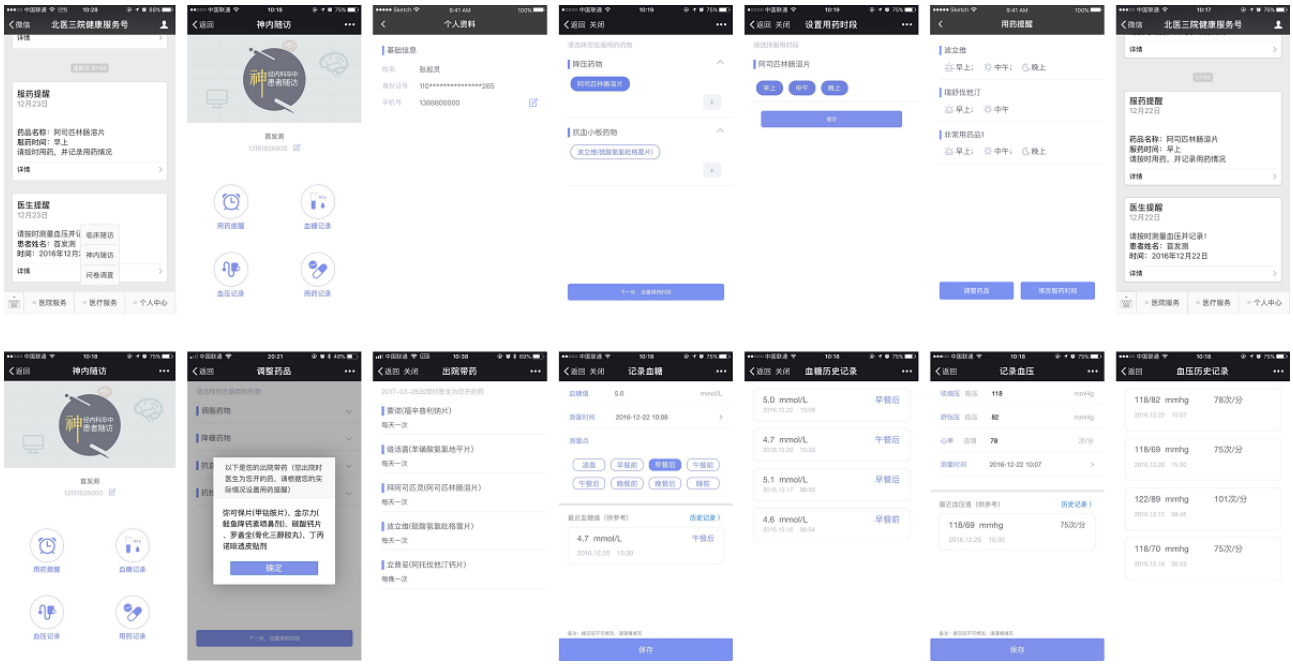
The PUTH Health Service Account, Department of Neurology follow-up domain, based on WeChat, created four modules and four items on the interface of the WeChat patient terminal. The modules emphasized the self-administration of prevention-therapy adherence, as explained by the designers and researchers. The construction and content of the four modules are described as follows:

1. Drug-use reminders: these provided therapeutic drug types, drug doses, original instructions for administration, and real-time modifications. The service platform automatically sent reminder messages according to a pre-established, inputted time point; feedback messages of "have taken medicine," "have not taken medicine," or no feedback were recorded by the central server.
2. Blood glucose recording: the module provided a recording option for glucose level (mmol/L) as well as a testing time and time point recording option that the patients could select; time point options included early morning, before breakfast, after breakfast, before lunch, after lunch, before evening dinner, after evening dinner, and before sleep. The interface of this module also provided a link for viewing the history record and the most recent blood glucose record.
3. Blood pressure recording: systolic blood pressure, diastolic blood pressure, heart rate, and measurement time option could be recorded. The interface of this module also provided a link for viewing the history record and the most recent blood pressure record.
4. Medication administration recording: the medication type and name, the period in which it was taken, and the condition in which it was taken could be reviewed and modified in this module.

The four WeChat-based services included the following:

1. Operating guide: instruction video that included steps on how to set up the medication reminder, how to complete the medication record, and how to view the medication record, as well as textual descriptions.
2. Event reporting: provided a recent event option, including no special event to report, where situation options were categorized as all is normal and not going to hospital or normal clinician visit and no special event; disease attack, including cerebral infarction, intracranial hemorrhage, transient ischemic attack, and cardiovascular disease; and other situation, which the patient could enter via text.
3. Discharge medication recording: another special route for discharged patients to enter their medication dosages and frequencies.
4. Message consultation: provided a real-time communication window between the researchers and participants; the text characters were limited by the WeChat app to 20 words for each question (see Figure 1). The communication dialog box between the researchers and participants was monitored as a pilot test, since we received content irrelevant to clinical practice and medication.

Figure 1. Screenshots of the WeChat-based self-monitoring program for prevention of secondary ischemic stroke, prevention-specific improvement services, and follow-up (Peking University Third Hospital, Health Service Account, Department of Neurology follow-up domain). The Web-based doctor terminal was included.



In the doctor management website terminal (ie, the patient information management system from the Department of Neurology), every participant has a record identification number. The administrative medical record information was transferred to the website. Meanwhile, the website included the following four modules:

1. Patient information: the demographic and clinical information was synchronized with the hospital administration clinical records. The medication records, as well as the blood pressure and glucose measurements obtained via the WeChat terminal, could be viewed here.
2. Compliance indication: the number of 1-week and 2-week periods of noncompliance were highlighted, and the patients' information could be obtained through this module.
3. Endpoint report information: patient-reported data in the WeChat terminal or manually inputted researcher follow-up data were shown, whereas only events defined by the researchers' telephone interviews were analyzed.
4. Medication management: the medication compliance of every dose that ought to have been taken could be observed in this module. All the potential medications were listed for the patients to be able to see their options. Figures 2 to 4 show the theoretical framework of the platform.

Figure 2. Information-interaction processing diagram of the WeChat platform.

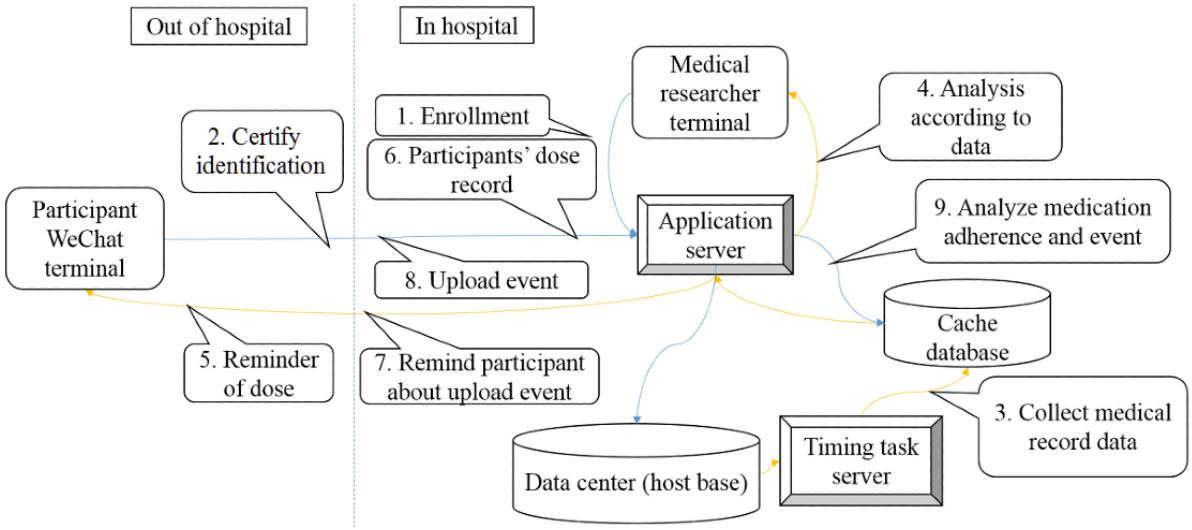


Figure 3. Synchronous web terminal of the WeChat information platform part 1.

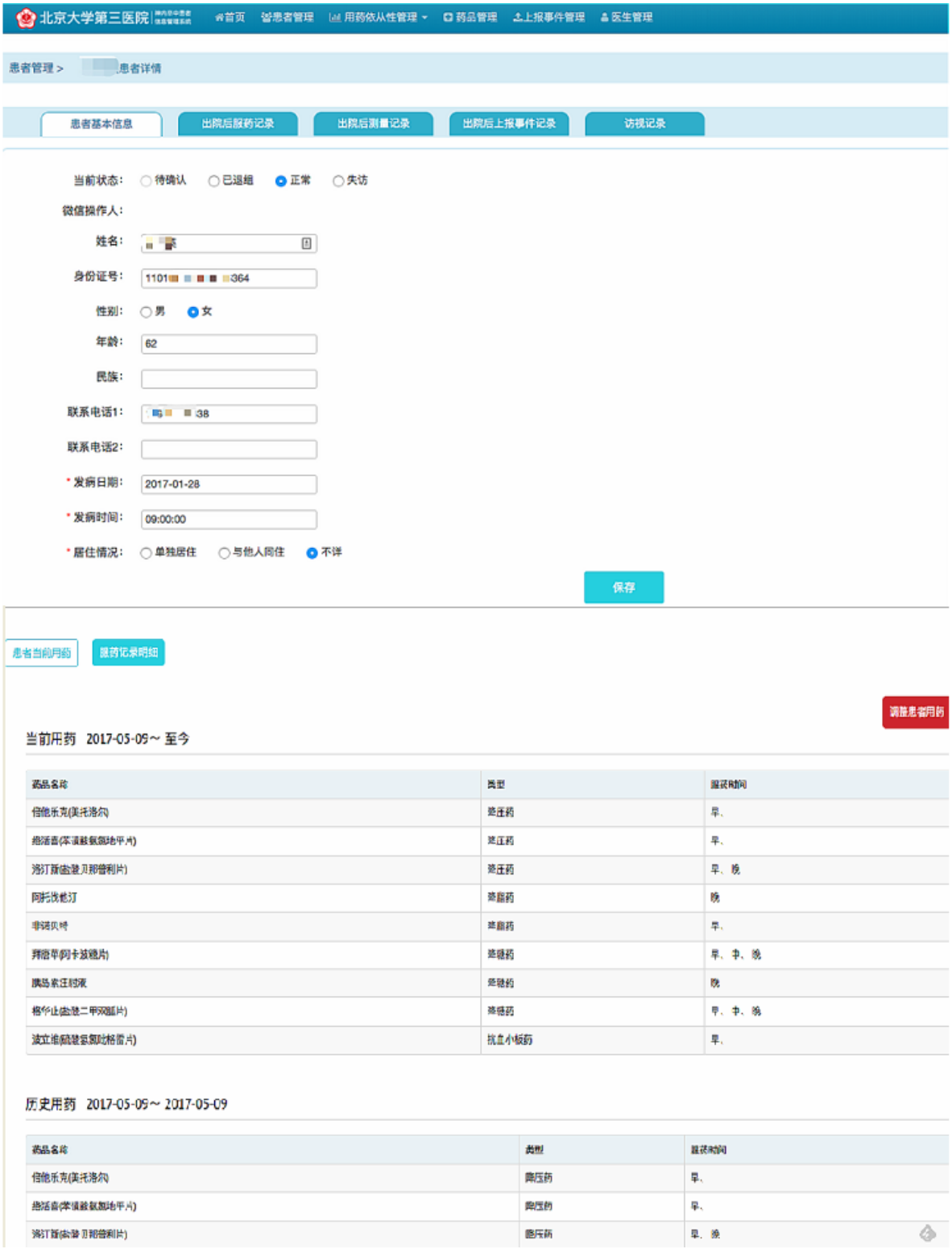
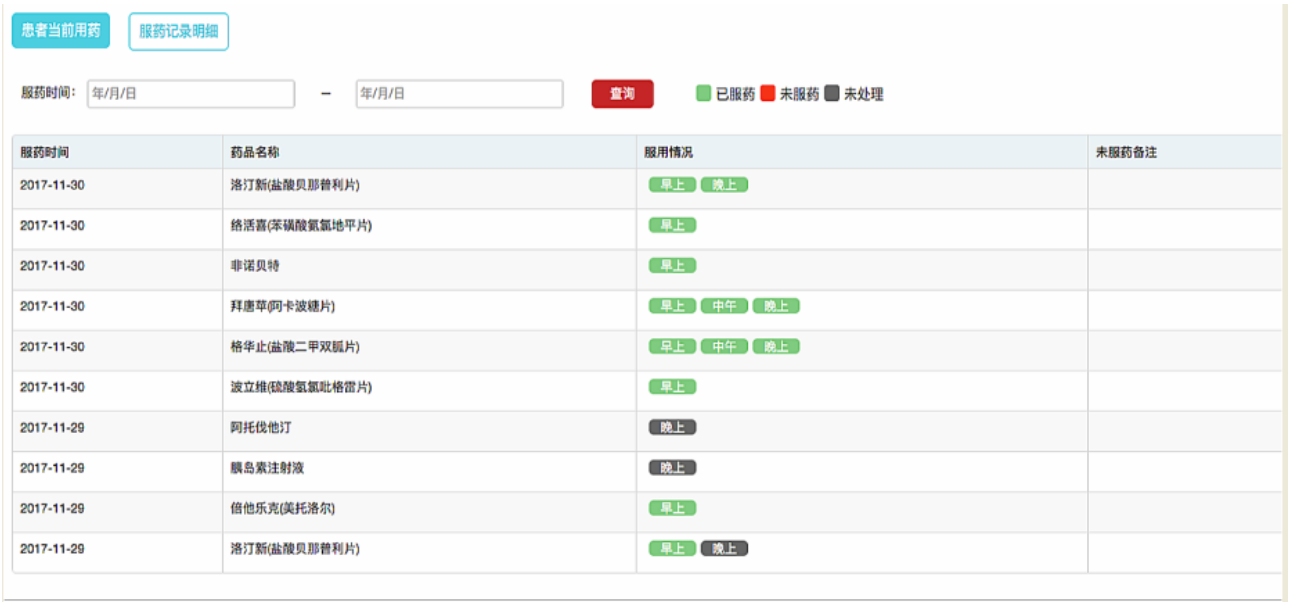


Figure 4. Synchronous web terminal of the WeChat information platform part 2.



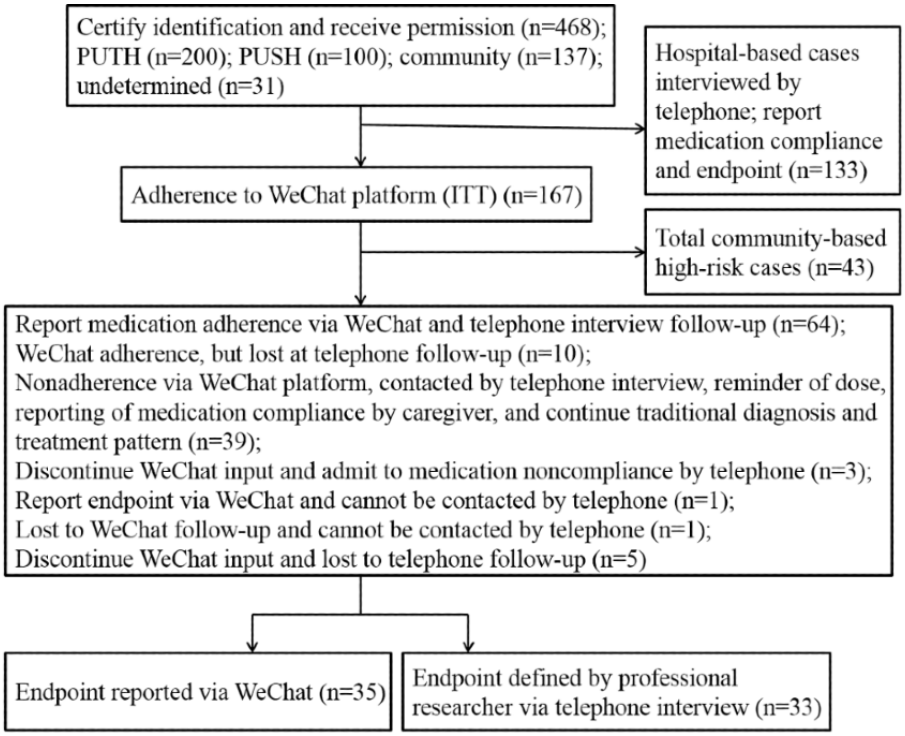
The Process for Implementing Monitoring

To ensure compliance of the demonstrated prevention elements, health education for secondary prevention began when the patients received antiplatelet agents, antihypertension agents, and statins in the stable disease period. Before discharge, on-site health education lectures and an introduction to the WeChat online interaction-monitoring platform were provided for each participant. The WeChat platform introduction, health education, and doctor recommendations were provided by telephone interview once every 3 months during the follow-up period if the participants had not received and adhered to the monitoring. The researchers logged in to the doctor terminal periodically to identify noncompliance; if noncompliance was identified, they contacted the participant to remind them to participate on the WeChat platform, routinely assessed follow-up endpoint event data, and investigated the reasons for nonadherence. The interactive communication via WeChat lasted 24 months and was facilitated and improved via modification from the academic members, based on participants' feedback, until the final version was deemed sufficient. During this process, some irrelevant information was identified and feedback was provided regarding symptom descriptions. Additionally, during special political periods, services were restricted and were temporarily replaced by manual secondary prevention outpatient services.

Research Design

This study was designed as a cohort study to examine the efficacy, feasibility, and acceptability of a WeChat-based service that was based on an information-motivation-behavioral skills model. The study protocol was approved by the Ethics Committee of the PUTH (2013-144) and PUSH (IRBK-2017-033-01), and participants gave written informed consent for participation. This study enrolled participants from September 2016 to September 2017. We compared the endpoint event rate between traditional prevention monitoring (ie, outpatient clinic visit, including health education and medication prescription) and WeChat monitoring (ie, medication adherence reminders, according to participants' prescriptions, plus traditional monitoring). The flowchart of the study design is shown in Figure 5. A total of 2 years of follow-up data were collected to analyze WeChat platform adherence and related factors.

Figure 5. Flowchart of the study design. PUSH: Peking University Shougang Hospital; PUTH: Peking University Third Hospital; ITT: intention-to-treat.



Pilot Study Adherence Parameters and Outcomes

Compliance, or adherence, to a medication is defined as the extent to which participants continuously take all secondary preventive medications as prescribed at hospital discharge, except if their health care provider instructs them to stop taking a medication [14]. Compliance to treatment was evaluated by the number of days that the drug was taken. For participants whose medical records could be attached, we calculated the prescription dosage and cycle to determine the rate of compliance. For residents who were outside the hospital region after discharge, we investigated the adherence rate by telephone communication. Not adhering to the medication as instructed for 2 weeks was defined as noncompliance. Recurrent events included ischemic stroke, intracranial hemorrhage, transient ischemic attack demonstrated by imaging examination, and cardiovascular disease. The recurrent event outcomes were applied in the analysis. The loss to follow-up was recorded. WeChat endpoint events reported by the participants themselves were defined and classified through telephone interviews with the researchers and through medical records.

Statistics

The data were analyzed using SPSS Statistics for Windows, version 22.0 (IBM Corp), and the figures were created using SPSS and Prism 5.0 (GraphPad). The feasibility analysis was conducted as intention-to-treat, based on the participants' recruitment. The risk ratio was calculated with SPSS and the recurrent event rate was compared between the WeChat self-monitoring group and the traditional group. The Kolmogorov-Smirnov test was used to determine the normality of the data. Normally distributed descriptive data were reported as mean and SD, and skewed distributions were reported as median with 25th and 75th percentiles (ie, quartiles). Demographic characteristics were compared using an independent *t* test for normally distributed continuous variables, based on the baseline data of the group with good adherence and the group with poor adherence. The nonparametric Mann-Whitney U test was used for continuous variables with skewed distributions, and the χ^2 test was used for categorical variables. Trends and bias analysis was described using a Prism scatter graph and bar graphs. A logistic regression and a survival analysis were calculated using SPSS. A two-sided *P* value of .05 was considered statistically significant.

Results

Baseline Characteristics and Selection Bias Analysis

A total of 468 participants (PUTH, 200/468, 42.7%; PUSH, 100/468, 21.4%; Huayuanlu Community Healthcare Service Center, 137/468, 29.3%; and undetermined, 31/468, 6.6%) were registered on the platform. Of those participants, 234 (50.0%) had used the platform at least once and had input data since enrollment. There were no statistically significant differences between the participants who did or did not input data (see Figure 6 and Table 1).

Figure 6. Selection bias analysis of the WeChat platform registry population. DM: diabetes mellitus; HLP: hyperlipidemia; HP: hypertension; TIA: transient ischemic attack.

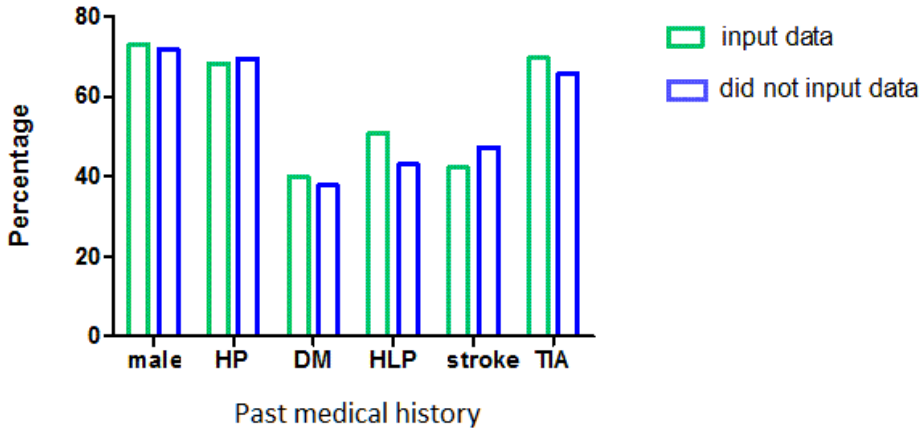


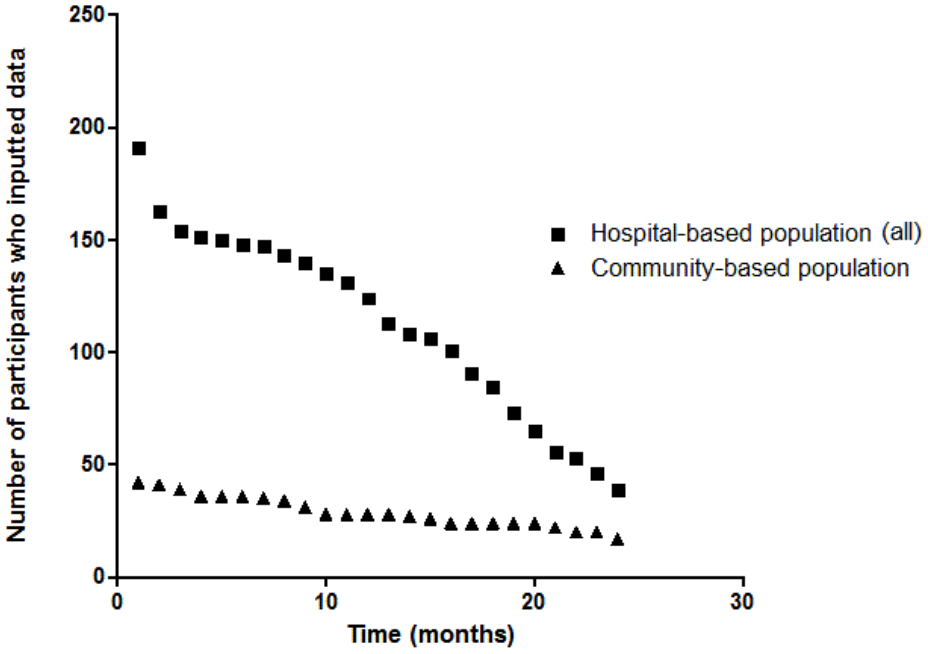
Table 1. Selection bias analysis of participants who did or did not input data into the WeChat platform registry.

Participant characteristic	Registry population		Statistic	*P* value
	Input data	Did not input data		
Demographic characteristic				
Age (years), mean (SD)	59.95 (14.42)	62.34 (11.67)	t_{220}=-1.925	.06
Gender (male), n/N (%)	171/234 (73.1)	168/234 (71.8)	χ^2_1=0.1	.84
Medical history, n/N (%)				
Hypertension	160/234 (68.4)	163/234 (69.7)	χ^2_1=0.1	.84
Diabetes	84/210 (40.0)	85/225 (37.8)	χ^2_1=0.2	.69
Hyperlipidemia	107/210 (51.0)	97/225 (43.1)	χ^2_1=2.7	.10
Stroke history	97/229 (42.4)	110/232 (47.4)	χ^2_2=2.1	.35
Transient ischemic attack	147/211 (69.7)	148/225 (65.8)	χ^2_1=0.8	.41

Within the research period, 167 hospital-based participants inputted data into WeChat after a 2-month learning and adaption period (see Figure 7). The *event report* domain was used by 13.9% (14/101) of participants until the 1-year follow-up, with endpoint data from the hospital-based population. The *drug use* module was used by 124 participants at 1 year, while 23 participants did not complete the telephone follow-up at 1 year.

The *blood pressure* module was used by 90% (61/68) of participants with hypertension, and the application time (ie, the time that the participants actively used WeChat) was 14.18 (SD 9.04) months. The *glucose* module was used by 66% (44/67) of participants with diabetes, and the application time was 13.47 (SD 8.37) months.

Figure 7. Participants' usage of the WeChat-based modules.



Subjects were enrolled at the PUTH via verbal notification during discharge, followed by a text message, telephone introduction, and self-study of the platform instructions. Enrollment at the PUSH included bedside education and instruction within the administration period and self-study of the platform instructions. Enrollment of the community-based population included collective mission education, face-to-face instructions, and self-study of the platform instructions.

Analysis of Medication Noncompliance: Follow-Up by Telephone and Reasons

The mean compliance time based on the inputted data, update of the platform vision based on participant feedback, and participant feedback regarding the WeChat platform's drug use module showed that patients complied with antihypertension drugs, antiplatelet drugs, statins, and hypoglycemic drugs for an average of 15.52 (SD 7.24), 14.78 (SD 7.45), 15.42 (SD 7.11), and 19.05 (SD 6.72) months, respectively. At the 1-year telephone interview and medical record follow-up, 3.0% (3/101) of patients receiving WeChat monitoring and 7.0% (11/157) of patients receiving traditional monitoring admitted to self-determined medication noncompliance because they had no symptoms, did not understand the effectiveness of the medication, or did not realize the importance of medication compliance. The chi-square was 1.9 (*df*=1, *P*=.16). Overall, there was less nonadherence in the WeChat monitoring group than in the traditional monitoring group.

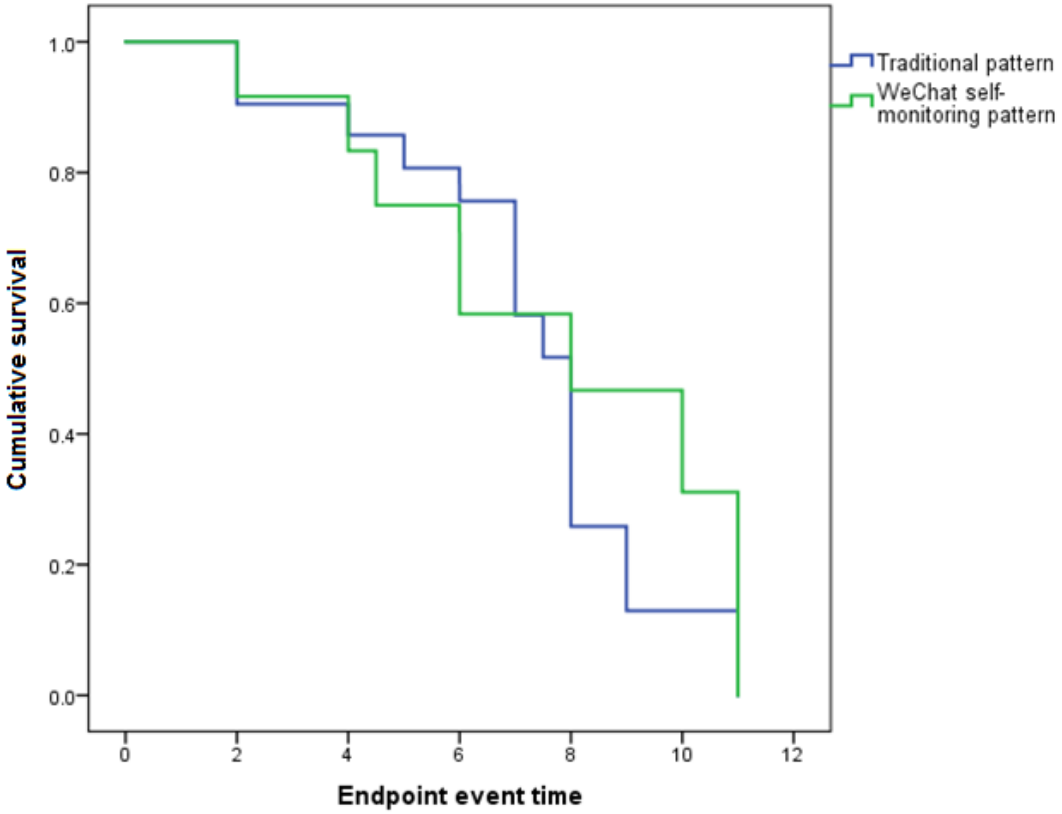
Endpoint Recurrent Event Analysis

There were no statistically significant differences in age and gender between the WeChat self-monitoring group and the traditional monitoring group: WeChat group, 58.1 (SD 14.8) years of age versus traditional group, 60.1 (SD 15.4) years of age, $t_{28.5}$=-0.999, *P*=.32; WeChat group, 22/101 (21.8%) males versus traditional group, 29/157 (18.5%) males, χ^2_2=0.4, *P*=.55.

According to the telephone interviews and endpoints identified by the researchers' investigation of the medical records, the rate of loss to follow-up of the total hospital-based population was 14.0% (42/300). A total of 101 participants in the WeChat self-monitoring group and 157 participants in the traditional monitoring group could be contacted.

The 1-year follow-up event rate, as defined by the researcher, of the WeChat self-monitoring group was 11.9% (12/101), which was less than that of the traditional monitoring group (21/157, 13.4%). Compared with the traditional group, the risk ratio of the WeChat group was 0.983 (95% CI 0.895-1.080), which was not statistically significant. The mean endpoint event time of the WeChat self-monitoring group was 7.04 (SD 2.85) months, whereas that of the traditional group was 6.74 (SD 2.35) months. Figure 8 shows the survival curve of the event (log-rank test, *P*=.49).

Figure 8. Survival functions: Kaplan-Meier analysis of the rate of endpoint events during the 1-year follow-up.



There was no statistically significant difference in adherence time to the WeChat platform between the endpoint group (defined by WeChat self-report) (mean 17.8 [SD 7.5] months) and the nonendpoint group (mean 15.5 [SD 7.1] months) (P=.16), where endpoint was defined by the WeChat platform as reported by participants. Similarly, there was no statistically significant difference in adherence time between the endpoint group (mean 16.3 [SD 11.9] months) and the nonendpoint group (mean 16.3 [SD 6.7] months) (P=.99), where endpoint was assessed via telephone interview and medical record.

Feasibility and Acceptability Parameters

The differences between patients with good adherence and poor adherence were examined. Using each study site's mean adherence time as the cutoff value (PUTH, 11.8 months; PUSH, 13.88 months; and community center, 16.63 months) with denoising, we summarized the valid data from 115 hospital-based and 43 community-based users with good adherence as well as 185 hospital-based and 94 community-based users with poor adherence. No significant difference in age was found between the patients with good adherence and those with poor adherence; however, more male participants adhered to inputting information into the WeChat platform (see Figure 9 and Table 2).

Figure 9. Demographics of the participants with good and poor adherence.

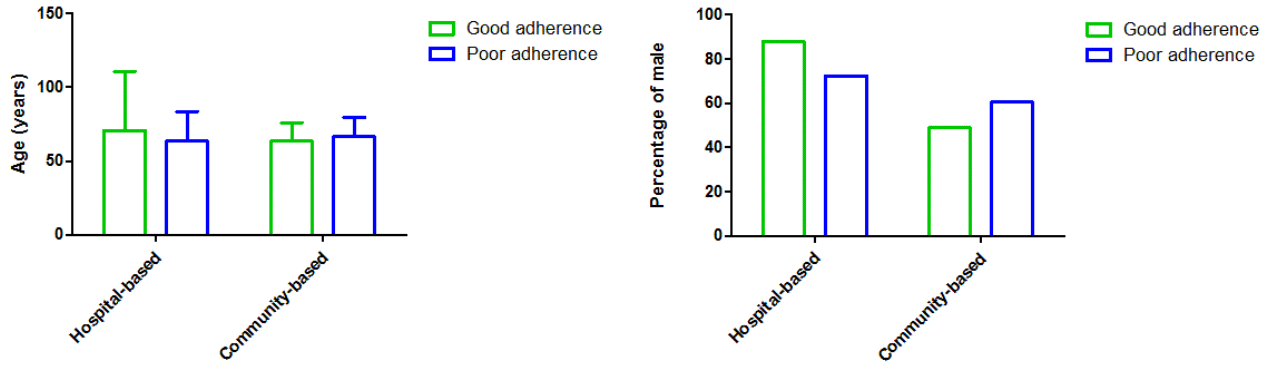


Table 2. Demographics of participants with good versus poor adherence in each population.

Demographic and population	Group with good adherence	Group with poor adherence	*P* value
Gender (male), n/N (%)			
Hospital-based population	101/115 (87.8)	134/185 (72.4)	.002
Community-based population	21/43 (49)	57/94 (61)	.17
Age (years), mean (SD)			
Hospital-based population (good adherence n=115; poor adherence n=185)	70.8 (40.0)	63.6 (20.0)	.07
Community-based population (good adherence n=43; poor adherence n=94)	63.5 (12.4)	66.5 (13.1)	.20

The number of participants who adhered decreased gradually each month (see Figure 7). The average adherence period was 16.5 (SD 0.8) months (range 2-24 months). The mean decrease in adherence was 4 participants (1.1%) per month. The slope for the decrease in adherence during the first year of follow-up (mean 4.0 [SD 0.2] cases/month) was significantly less than the slope during the second year of follow-up (mean 7.2 [SD 0.3] cases/month) (P<.001). The slope for the decrease in adherence of the hospital-based population (5.8 [SD 0.2] cases/month) was significantly greater than that for the community-based population (1.0 [SD 0.0] cases/month) (P<.001). A logistic regression showed that there were no risk factors for compliance during the first year (site: OR 1.529, 95% CI 0.906-2.579, P=.11; resident location: OR 2.014, 95% CI 0.648-6.262, P=.23; age: OR 0.979, 95% CI 0.948-1.011, P=.20; gender: OR 0.980, 95% CI 0.453-2.120, P=.96; stroke history: OR 0.979, 95% CI 0.948-1.011, P=.20; and TIA: OR 1.956, 95% CI 0.468-2.387, P=.90). Although community-based population status (OR 2.373, 95% CI 1.019-5.527, P=.045) and TIA (OR 0.122, 95% CI 0.016-0.940, P=.04) were factors that influenced 2-year compliance, other potential factors showed no statistical significance (resident location: OR 2.295, 95% CI 0.616-8.454, P=.22; age: OR 1.004, 95% CI 0.964-1.046, P=.85; gender: OR 1.529, 95% CI 0.567-4.121, P=.40; and stroke history: OR 1.277, 95% CI 0.575-2.840, P=.55).

Reasons for WeChat-Based Nonadherence According to the Telephone Interview Follow-Up

The telephone interview had a loss to follow-up rate of 14.0% (42/300) among the hospital-based population, mostly because the telephone numbers left during the hospitalization period changed or belonged to people who were not the caregivers. The analysis of the reasons for WeChat nonadherence in the hospital-based population at the 1-year follow-up showed that 11.3% (34/300) of participants or caregivers had difficulty learning to use WeChat or did not use the app. Additionally, 1.7% (5/300) of the participants lacked an internet connection while traveling, and 31.7% (95/300) had a caregiver who provided reminders for medication, regularly sought medical services from a local hospital, and did not intend to participate in WeChat monitoring.

Discussion

Previous studies have shown that an important explanation for medication noncompliance is lack of knowledge about reasons for adherence [15], dosages, and time of administration for the medication. In general, reasons for noncompliance reported by hospital-based neurological disease patients included modifications to the medication (41.7%) and forgetting to take the medication (33.2%) [16]. For patient-related factors, behavior approaches include either the use of aids, such as pill organizers, medication calendars, and blister packs, or active family involvement. We found an elevated level of medication compliance related to stroke secondary prevention among the WeChat monitoring group, which indicates the advantage of a WeChat reminder and confirms that mobile technology is potentially a promising means for improving health care.

The decreasing endpoint event rate reported in the WeChat monitoring group also indicated that recurrent events were delayed compared with traditional monitoring: the follow-up survival curve at 7 months after enrollment showed a better event rate among the WeChat group than among the traditional monitoring group. We attributed this to the accumulation effect of secondary prevention.

In this manuscript, we report that the innovative, stroke-specific, WeChat self-management program was well-received by participants at a local site as a useful supplement to medication adherence. The rate of decreasing adherence in inputting WeChat data was lower in the community-based population than in the hospital-based population; therefore, being a community-based participant was an influencing factor for good compliance in this study. This finding can be interpreted in that patient flow varies in different types of health institutions. Traditionally, discharged ischemic stroke patients undergo subsequent visits to the outpatient department of the hospital, community health care center, or local hospital of their choice near their residence, based on the available local medical resources. A proportion of participants with sudden-onset ischemic stroke would return to their local residential region after discharge. Moreover, participant-doctor relationships at community health service centers are often more interactive and friendly. The WeChat app has the advantages of being generalizable and able to reach more hospitals with the monitoring abilities and team needed for the high-risk stroke population. Promoting the use of local patient sites is an impactful and appropriate method to increase adherence.

We analyzed the factors that influenced self-monitoring adherence of the WeChat-based ischemic stroke secondary prevention, according to the slope parameters and the authors' explanations. We found that WeChat self-management was better received by males, and we interpreted this finding to the stronger intrinsic motivation of male participants [17]. TIA is an influencing factor for poor adherence; except for selection

bias, we think the lack of attention can be attributed to insufficient health education. The decrease in adherence rate was similar between the two hospital sites (PUTH slope: mean 3.0 [SD 0.2] cases/month; PUSH slope: mean 2.9 [SD 1.1] cases/month). The adherence rate was similar to other studies, in that it tended to be higher in the earlier months before dropping off over time [18]. The authors explained that there is a *learning effect*, in which participants would be encouraged to adhere only after having enough time to experience benefits from the WeChat-generated reminders and would then intend to benefit from adhering to the intervention method.

This study has several limitations. First, the lower WeChat adherence rate in our research was partially attributed to the fact that there was no statistically significant difference in the endpoint event rate between WeChat and traditional monitoring for secondary prevention, although the total recurrent ischemic stroke endpoint rate was less than that in previous reports. Second, the data from the blood glucose and blood pressure modules were not sufficiently transformed, since there was not enough input. The absence of sufficient risk-factor-control information, as well as relative communication via the WeChat terminal between doctor and participants, may have reduced the effectiveness of the secondary prevention. The research team members should provide ongoing support for the construction of the patient platform and doctor terminal; this would ensure delivery of the WeChat-based learning activities though WeChat videos and telephone calls during the implementation phase in order to avoid user fatigue over time. To our knowledge, the following are influencing factors of the feasibility of WeChat monitoring: performance expectancy (ie, improvements in ischemic stroke management and mind relaxation), effort expectancy (ie, ease of, or interest in, technical use), facilitating conditions (ie, availability of technical support and interactive doctor-participant communication), social influence (ie, support from caregivers), and habit (ie, the degree to which participating became a daily routine) [19]. Although we defined the endpoint of this study, we noticed that participants' definitions of an event were not identical, so we struggled to upload an event as a potential influencing factor for adherence. Third, during the app progression and modification process, the module construction and improvements in function focused on ease of use, based on feedback from the participants. In future research, more effective types of therapy that could be conducted and popularized by the WeChat platform, or other attractive mobile health (mHealth) methods, should be introduced to increase the adherence rate.

Indeed, discharged patients are occasionally not compliant with their medication; however, we discovered that more discharged patients prefer face-to-face, traditional follow-up in clinical practice with a clinician. From the telephone follow-up, we found that daily, regular, medical practices of principal prevention strategies were well-adopted among our research population.

Since the diagnosis and treatment levels and medical resources are not identical across different hospitals and health service centers, the use of mHealth programs after discharge, such as WeChat platform monitoring, is convenient for patients and caregivers in both hospital-based and community-based populations. We should provide users with the best program, which is sustainable, dependable, and usable, in order to ensure that the self-monitoring service is well-received by the participants. The disadvantages of mHealth apps include incomplete network coverage and the participants' or caregivers' lack of internet technology skills regarding mobile phones and the WeChat terminal. Additionally, we expect that enabling innovative changes in secondary prevention practice to decrease the rate of recurrent stroke events will motivate learning. A two-arm, clustered, adaptability randomized controlled trial with lower loss to follow-up should be developed to demonstrate the effectiveness of mHealth programs.

Acknowledgments

We would like to thank all the participants for participating in this study. This study was funded by the Beijing Municipal Natural Science Foundation (2014-1-4092) and the National Natural Sciences Foundation of China (81471184).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2820 KB - mhealth_v8i4e16496_app1.pdf]

References

1. GBD 2016 Lifetime Risk of Stroke Collaborators, Feigin VL, Nguyen G, Cercy K, Johnson CO, Alam T, et al. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. N Engl J Med 2018 Dec 20;379(25):2429-2437 [FREE Full text] [doi: 10.1056/NEJMoa1804492] [Medline: 30575491]
2. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):439-458 [FREE Full text] [doi: 10.1016/S1474-4422(19)30034-1] [Medline: 30871944]

3. Yang G, Wang Y, Zeng Y, Gao GF, Liang X, Zhou M, et al. Rapid health transition in China, 1990-2010: Findings from the Global Burden of Disease Study 2010. Lancet 2013 Jun 08;381(9882):1987-2015. [doi: 10.1016/S0140-6736(13)61097-1] [Medline: 23746901]
4. Wang Y, Xu J, Zhao X, Wang D, Wang C, Liu L, et al. Association of hypertension with stroke recurrence depends on ischemic stroke subtype. Stroke 2013 May;44(5):1232-1237. [doi: 10.1161/STROKEAHA.111.000302] [Medline: 23444308]
5. Pandian JD, Gall SL, Kate MP, Silva GS, Akinyemi RO, Ovbiagele BI, et al. Prevention of stroke: A global perspective. Lancet 2018 Oct 06;392(10154):1269-1278. [doi: 10.1016/S0140-6736(18)31269-8] [Medline: 30319114]
6. Wang Y, Zhang S, Zhang L, Wang C, Dong Q, Gao S, et al. Chinese guidelines for the secondary prevention of ischemic stroke and transient ischemic attack 2010. CNS Neurosci Ther 2012 Feb;18(2):93-101 [FREE Full text] [doi: 10.1111/j.1755-5949.2011.00290.x] [Medline: 22313945]
7. Davis SM, Donnan GA. Clinical practice: Secondary prevention after ischemic stroke or transient ischemic attack. N Engl J Med 2012 May 17;366(20):1914-1922. [doi: 10.1056/NEJMcp1107281] [Medline: 22591297]
8. Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, American Heart Association Stroke Council, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Peripheral Vascular Disease. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014 Jul;45(7):2160-2236. [doi: 10.1161/STR.0000000000000024] [Medline: 24788967]
9. Sappok T, Faulstich A, Stuckert E, Kruck H, Marx P, Koennecke HC. Compliance with secondary prevention of ischemic stroke: A prospective evaluation. Stroke 2001 Aug;32(8):1884-1889. [doi: 10.1161/01.str.32.8.1884] [Medline: 11486121]
10. Bushnell CD, Olson DM, Zhao X, Pan W, Zimmer LO, Goldstein LB, AVAIL Investigators. Secondary preventive medication persistence and adherence 1 year after stroke. Neurology 2011 Sep 20;77(12):1182-1190 [FREE Full text] [doi: 10.1212/WNL.0b013e31822f0423] [Medline: 21900638]
11. Wei JW, Wang J, Huang Y, Liu M, Wu Y, Wong LK, ChinaQUEST Investigators. Secondary prevention of ischemic stroke in urban China. Stroke 2010 May;41(5):967-974. [doi: 10.1161/STROKEAHA.109.571463] [Medline: 20224061]
12. Dubey D, Amritphale A, Sawhney A, Amritphale N, Dubey P, Pandey A. Smart phone applications as a source of information on stroke. J Stroke 2014 May;16(2):86-90 [FREE Full text] [doi: 10.5853/jos.2014.16.2.86] [Medline: 24949314]
13. Wu N, Gong E, Wang B, Gu W, Ding N, Zhang Z, et al. A smart and multifaceted mobile health system for delivering evidence-based secondary prevention of stroke in rural China: Design, development, and feasibility study. JMIR Mhealth Uhealth 2019 Jul 19;7(7):e13503 [FREE Full text] [doi: 10.2196/13503] [Medline: 31325288]
14. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: Terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]
15. Sabaté E. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1 [accessed 2020-02-08]
16. Prell T, Grosskreutz J, Mendorf S, Franke GH, Witte OW, Kunze A. Clusters of non-adherence to medication in neurological patients. Res Social Adm Pharm 2019 Dec;15(12):1419-1424. [doi: 10.1016/j.sapharm.2019.01.001] [Medline: 30772239]
17. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: A review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
18. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient compliance in home-based self-care telehealth projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]
19. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: A longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]

Abbreviations

mHealth: mobile health
PUSH: Peking University Shougang Hospital
PUTH: Peking University Third Hospital
TIA: transient ischemic attack
TOAST: Trial of Org 10172 in Acute Stroke Treatment

Edited by G Eysenbach; submitted 05.10.19; peer-reviewed by G Xie, D Gandhi; comments to author 18.11.19; revised version received 08.01.20; accepted 24.01.20; published 15.04.20.

Please cite as:
Zhang Y, Fan D, Ji H, Qiao S, Li X
Treatment Adherence and Secondary Prevention of Ischemic Stroke Among Discharged Patients Using Mobile Phone- and WeChat-Based Improvement Services: Cohort Study
JMIR Mhealth Uhealth 2020;8(4):e16496
URL: https://mhealth.jmir.org/2020/4/e16496
doi:10.2196/16496
PMID:32293574

Original Paper

British South Asian Patients' Perspectives on the Relevance and Acceptability of Mobile Health Text Messaging to Support Medication Adherence for Type 2 Diabetes: Qualitative Study

Suman Prinjha[1], BSc, MSc, PhD; Ignacio Ricci-Cabello[2,3,4], PhD; Nikki Newhouse[1], BA, MSc, PhD; Andrew Farmer[1], DM, FRCGP

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Balearic Islands Health Services, Primary Care Research Unit of Mallorca, Palma de Mallorca, Illes Balears, Spain

[3]Health Research Institute of the Balearic Islands (IdISBa), Palma de Mallorca, Illes Balears, Spain

[4]Ciber de Epidemiologia y Salud Pública (CIBERESP), Madrid, Spain

Corresponding Author:
Suman Prinjha, BSc, MSc, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford,
United Kingdom
Phone: 44 (0)7774 629231
Email: suman.prinjha@phc.ox.ac.uk

Abstract

Background: The prevalence of type 2 diabetes (T2D) is greater in South Asian populations and health outcomes are poorer compared with other ethnic groups. British South Asians are up to six times more likely to have T2D than the general population, to develop the condition at a younger age, and to experience diabetes-related complications. Interventions to support people in managing their condition can potentially reduce debilitating complications. Evidence to support the use of digital devices in T2D management, including mobile phones, has shown positive impacts on glycemic control. There is increasing recognition that health interventions that are culturally adapted to the needs of specific groups are more likely to be relevant and acceptable, but evidence to support the effectiveness of adapted interventions is limited and inconclusive.

Objective: This formative study aimed to explore the perceptions and views of British South Asian patients with T2D on mobile health SMS text messaging to support medication adherence, aimed at the general UK population.

Methods: Eight exploratory focus groups were conducted in Leicester, the United Kingdom, between September 2017 and March 2018. A diverse sample of 67 adults took part.

Results: British South Asian people with T2D who use digital devices, including mobile phones, felt that short messages to support medication adherence would be acceptable and relevant, but they also wanted messages that would support other aspects of self-management too. Participants were particularly interested in content that met their information needs, including information about South Asian foods, commonly used herbs and spices, natural and herbal approaches used in the United Kingdom and in South Asia, and religious fasting. Short messages delivered in English were perceived to be acceptable, often because family members could translate for those unable to read or understand the messages. Suggestions to support patients unable to understand short messages in English included having them available in different formats, and disseminated in face-to-face groups for those who did not use digital devices.

Conclusions: Exploring the views of British South Asian patients about SMS text messaging aimed at the general UK population is important in maximizing the potential of such an intervention. For such a digital system to meet the needs of UK South Asian populations, it may also have to include culturally relevant messages sent to those who opt to receive them. It is equally important to consider how to disseminate message content to patients who do not use digital devices to help reduce health inequalities.

(JMIR Mhealth Uhealth 2020;8(4):e15789) doi:10.2196/15789

KEYWORDS

type 2 diabetes; South Asians; text messages; self-management; medication adherence; mobile health; mHealth; eHealth

Introduction

Digital Health Interventions to Support Patients With Type 2 Diabetes

The prevalence of type 2 diabetes (T2D) is increasing globally, representing a serious clinical and financial challenge [1]. Diabetes prevalence is greater in South Asian populations and health outcomes are poorer [2]. South Asians account for approximately one-fifth of the global population. T2D occurs at 50% higher rates in South Asian patients compared with the general population [3], developing 5 to 10 years earlier, and is one of the main causes of premature death in this population [4]. South Asians constitute the largest minority ethnic group in the United Kingdom, around 7% of the total British population [5]. UK South Asians are up to six times more likely to have T2D than the general population, to develop the condition at a younger age, and to experience diabetes-related complications [6].

Effective diabetes self-management (DSM)—healthy diet, physical activity, and medication adherence—is associated with improved glycemic control [7], leading to a reduction in complications and mortality [8]. Interventions to support people in effectively self-managing their condition can potentially reduce costly, debilitating complications [9,10]. Interventions to support medication adherence can lead to improved health outcomes [11]. Digital health technologies have the potential to deliver low-cost interventions aimed at supporting healthier lifestyles [12] and disease self-management [13,14]. The number of mobile phone users surpassed 2 billion in 2016 and is expected to increase to 2.86 billion by 2020 [15]. In the United Kingdom, 73% of adults accessed the internet through a smartphone or mobile device in 2017 [16]. The growing evidence supporting the use of mobile phone–based technologies in T2D management, including SMS text messages, has shown positive impacts on glycemic control [17] and health care costs [18].

Cultural Adaptation of Health Interventions

Research into the cultural adaptation of health promotion interventions, including SMS text messaging interventions, is growing steadily, with an increasing recognition that lifestyle and behavior change interventions that are culturally adapted to the needs of specific groups are more likely to be effective [19-21]. Cultural adaptation involves grounding interventions in the lived experience of end users, taking account of language, cultural patterns, and values [22,23]. Mobile health (mHealth) interventions have been culturally adapted and piloted or trialed in several countries, including in New Zealand to support healthy lifestyles in Māori and Pasifika communities [24,25]; maternal health in Māori, Pacific, Asian, and South Asian families [26]; and smoking cessation [27]. Interventions aimed at the general population tend to be less effective for Māori and Pasifika communities [28], possibly contributing to increased health inequalities [29]. These interventions have often been designed with little input from these communities, have lacked tailoring to cultural needs, and have had poor uptake [24].

The few culturally adapted complex interventions for UK South Asian populations with T2D have been lifestyle modification interventions (diet-, weight-, and physical activity–related) rather than digital health interventions [30,31]. However, evidence to support the effectiveness of these interventions is inconclusive in demonstrating sufficient health-related gain to make such interventions cost effective [32,33]. Head-to-head comparisons of adapted and nonadapted interventions are rare [34], though the use of adaptations has been documented to increase process outcomes such as acceptability, uptake, satisfaction, and retention [35].

Previous research into DSM in British South Asian populations has explored perceptions and experiences of taking prescribed diabetes medications and traditional medicines [36,37], and barriers and facilitators to diet management [38,39] and physical activity [40]. Patients' perceptions of oral diabetes medications are complex and ambivalent, with good patient-provider communication and an understanding of the cultural factors that inform beliefs and practices reported as key to improving medication adherence [41]. Diet-management has been highlighted to be the most difficult aspect of living with diabetes, with recommendations including designing interventions that involve family members as well as patients [42].

At present, there is no published research into digital health interventions for UK South Asian patients with T2D, or the cultural adaptation of digital health interventions aimed at the general population. The EuroDHYAN study currently underway in the United Kingdom includes a pilot study of SMS text messages for T2D prevention in women of Pakistani origin in Scotland [43]. We know little about how best to adapt interventions to meet the needs of South Asian migrant populations with T2D [44]. This formative research aimed to explore the perceptions and views of British South Asian patients with T2D on mHealth SMS text messaging to support medication adherence.

Study Design

Recent research into digital health intervention development highlights the importance of an iterative, multidisciplinary approach in which the end user is placed at the heart of the system and the technology is grounded in user wants and needs [45-47]. This is increasingly seen as crucial to person-centered or person-based intervention development [48-50]. Research examining SMS text messaging interventions, in particular, has highlighted the use of multiple methodologies in intervention development, including qualitative research in the planning stage to gather information from relevant stakeholders and to inform decisions about message frequency, timing, and level of tailoring [51].

Exploratory focus groups were conducted with a range of South Asian communities in Leicester, one of the most ethnically diverse cities in England [52]. The focus group study was part

of a larger project, the SuMMiT-D (Support Through Mobile Messaging and Digital Health Technology for Diabetes) study, which aimed to explore supporting people with T2D in effective use of their medicine through a system comprising digital health technology integrated with clinical care, including use of SMS text messages [53]. Messages were developed using a taxonomy of behavior change techniques to identify ways of structuring a wide range of messages derived from patients, health care professionals, and health promotion literature [54]. An extensive library of messages was developed by a meeting of psychologists and health care professionals, checked for fidelity to behavior change techniques and checked for acceptability to patients [55]. The focus group study was a collaboration between the University of Oxford and the Centre for Black and Minority Ethnic Health (CBMEH), University of Leicester.

Methods

Ethical Approval

Ethical approval was obtained from the University of Oxford Central University Research Ethics Committee (Ref R50751/RE001).

Recruitment and Participants

British South Asian populations include first, second, and third generation people of Indian, Pakistani, Bangladeshi, and Sri Lankan descent. Participants were purposively sampled to include a broad range of views [56], reflecting the heterogeneity within and across UK South Asian communities, including people of different age groups, educational and occupational backgrounds, fluency in English, place of birth, and time since diagnosis. We also aimed to include users and nonusers of digital devices. Adults with T2D were recruited from community centers, places of worship, and a South Asian women's center in Leicester, with the support of the CBMEH who had strong links with local community organizations. Potential participants were informed about the focus groups verbally and with written information in English by a CBMEH project support worker and by community center managers.

Focus Group Discussions

We conducted eight exploratory focus groups between September 2017 and March 2018. Discussions were held in venues that participants were familiar with and could easily access, such as community centers. Each group met once, and the discussions lasted between 1.5 and 2 hours. The focus groups were facilitated by SP, a qualitative researcher, with cofacilitators from the CBMEH. Before the start of the focus group, the facilitators briefly introduced the study and discussed informed consent and confidentiality in the preferred language of participants. All participants gave signed consent. A topic guide was used to facilitate discussion, informed by the multidisciplinary literature on British South Asian experiences of T2D (Textbox 1). A detailed description of our methods is discussed elsewhere [57].

The focus groups aimed to explore South Asian perspectives of DSM, engagement with digital devices, support needs, and whether an SMS could help them to manage their condition. Examples of SMS text messages from the SuMMiT-D program were also discussed (Figure 1). Data collection continued until no new themes relating to the system were identified.

Textbox 1. Focus group topic guide.

Part 1: Challenges of living with type 2 diabetes and self-management

What do you find most difficult about living with diabetes?

- *Medication*
  - When is it difficult or challenging to take the tablets as recommended by the doctor?
  - When do you find it easiest to stick to a routine with your medications?
  - Do you use any herbal or non-Western approaches to help you with your diabetes?
- *Physical activity*
  - When is it difficult or challenging to exercise, go for a walk, or keep fit and healthy physically?
  - When do you find it easiest to stick to a routine with exercise and keeping fit and well?
- *Diet*
  - When is it difficult or challenging to eat healthily?
  - When do you find it easiest to stick to a routine with healthy eating?
  - Explore everyday diet and diet during festivities/family gatherings/during fasting.
  - What sorts of things do you do to help you to look after your diabetes and to stay as healthy as you can?
  - What might help you to manage your diabetes?

Part 2: Views on brief digital messaging system

- If we were going to design a new brief messaging system (show images), what kind of system and messages would help you most?
- Explore thoughts and feelings about the system and whether participants use mobile phones/digital devices.
- What do you think a system like this could include?
- What would make a system like this acceptable to you? What would make a system like this unacceptable to you?
- *System messages*
  - Do you have any problems with understanding any of the messages?
  - Explore suggestions to improve the messages so they can be more easily understood.
  - Were the messages clear and easy to understand?
  - How would you change the messages (use of language, tone, personalization)?
  - Do you think it is helpful to receive messages in English or to translate these messages? What are your thoughts about translated messages?
  - Do you think the proposed messages would help you personally to take your diabetes medication?
  - Who would you want the messages to come from? (National Health Service? General Practitioner? Pharmacist? Researchers?)
  - How often would it be helpful to receive messages like these?
  - Do you think this system would help you manage your health and diabetes better?
  - Would you sign up to a system like this? Explore reasons for and against.

Figure 1. Examples of messages from the Support Through Mobile Messaging and Digital Health Technology for Diabetes system.

Data Analysis

Discussions were conducted in Punjabi, Bengali, Sylheti, Urdu, Hindi, and English. They were audio-recorded, and translated and transcribed verbatim by SP or a professional transcriber. SP conducted the initial thematic analysis [58], coding inductively for main themes using a qualitative software package (NVivo, QSR International Pty Ltd, Melbourne, Australia, Version 12, 2018) [59]. Transcripts were then independently coded by NN. All four research team members (SP, RC, NN, and AF) discussed the data and themes, and finalized themes using a consensus process. Discrepancies were resolved and agreed by consensus.

Results

A total of 67 participants (including four carers) were recruited from some of the largest South Asian communities in the United Kingdom: Indian Punjabi Sikh, Pakistani Muslim, Indian Gujarati Hindu, Bangladeshi Muslim, and Indian Gujarati Muslim. Participants ranged in age from 18 to 84 years and included first- and second-generation South Asians. Four groups were mixed and four were single sex; groups ranged in size from n=5 to n=12 (Table 1).

Users and nonusers of digital devices discussed a range of challenges in terms of DSM and suggested a number of potential ways that SMS messages could help. Views ranged from those who felt that such a system would be a very useful addition to supporting existing self-management endeavors to those who stated that they would not use SMSs as they rarely used digital devices at all, including mobile phones. Users of digital devices discussed not only their own needs and preferences relating to the system but also their views and perceptions of what would be helpful for nonusers in their families and communities. Likewise, nonusers discussed their preferences and how family members who were users of digital technologies could help them. Five main themes were identified from the data, relating to system content and usability: (1) message content and design features, (2) language preferences, (3) family involvement, (4) different digital formats for different groups, and (5) face-to-face groups for those who do not use digital devices (Multimedia Appendix 1).

Table 1. Focus group composition and participant demographics (N=67).

Language/cultural group	Participants[a], n	Male[b], n	Female[c], n	Age range (years)	Country of birth (number of participants)
Punjabi Sikh men and women	11	5	6	47-78	• India (11)
Bangladeshi Muslim men	11	11	0	41-81	• Bangladesh (10) United Kingdom (1)
Pakistani Muslim men and women	7	3	4	39-66	• Pakistan (3) • India (1) • Bangladesh (1) • Malawi (1) • Mozambique (1)
Gujarati Hindu men and women	8	4	4	56-84	• India (4) • Kenya (2) • Uganda (1) • Trinidad (1)
South Asian women	12	0	12	18-71	• Bangladesh (3) • Pakistan (3) • India (2) • Sri Lanka (1) • Uganda (1) • Malawi (1) • United Kingdom (1)
Bangladeshi Muslim women	7	0	7	34-45	• Bangladesh (7)
Gujarati Muslim men	5	5	0	50-75	• India (4) • Malawi (1)
Younger people aged 18-45 years	6	1	5	28-47	• Bangladesh (6)

[a]Total=67.

[b]Total=29.

[c]Total=38.

Message Content and Design Features

Participants who used mobile phones or other digital devices were positive about receiving short messages about diabetes medication and the importance of taking medicines. However, they felt strongly that a messaging service should help support all aspects of self-management and not only medication adherence (Textbox 2).

The content of messages was a common theme, with many people discussing the kinds of messages that they would like to receive. Medication-related messages that they wanted included: information about diabetes symptoms, side effects, the risks associated with not taking medicines, the long-term effects of medications on kidney health, whether tablets should be taken before or after food, routine blood tests, and how to improve glycemic control. Participants expressed strong interest in messages that addressed unmet information needs, particularly about diet and physical activity, including messages about South Asian foods, portion sizes, and fasting. Participants also expressed a need for information about stress and stress management, natural and complementary approaches used in the United Kingdom and South Asia, and "reversing" diabetes:

> *In my opinion some kind of a diet plan should be introduced showing, for example, if you eat this amount of food in the morning, and in the afternoon one small bowl of lentils and one chapatti for example... If there was a diet plan, we could look at it and follow it. That would help.... When you go to see your doctor, they could show you that this is a plate and this is the portion size that should fill this plate....But these diet plans and portion sizes don't exist for Punjabi diets. They don't exist for our diets and foods.* [53-year-old Punjabi Sikh female]

Textbox 2. Patients' preferences for message content.

General information

- *Medication-related information*
  - Diabetes symptoms
  - Side effects
  - Risks associated with not taking medicines
  - Long-term effects of medications on kidney health
  - Whether tablets should be taken before or after food
  - Routine blood tests
  - New medications for diabetes
  - How to improve glycemic control
  - Why some people have to take tablets as well as insulin to manage diabetes?
- *Information about diet*
  - Healthy eating
  - Portion sizes
  - Sugar content in foods including fruits (eg, mangoes and bananas)
  - Dietary guidelines
  - Effects of commonly consumed foods on blood sugar levels
- *Information about physical activity*
  - Recommendations for walking
  - Exercises for people unable to walk because of disability (eg, exercises they could do while sitting)
- *Other information to support self-management*
  - Stress and stress management
  - "Reversing" diabetes
  - Information about current research
  - New research findings
  - Details of local diabetes-related events such as talks and discussions

Information relating to South Asian culture and self-management

- Healthy South Asian diets
- Effects of South Asian foods on blood sugar levels (eg, rice, chapatti, and different types of chapatti flour)
- South Asian diet plans and portion sizes
- Herbs and spices that may have health benefits (eg, turmeric, cinnamon, fennel, and neem)
- Health benefits of honey and dates
- Natural and complementary approaches used in the United Kingdom and South Asia
- Fasting and safe medicine taking
- Medicine taking when traveling to South Asia
- Local women-only gyms and swimming classes
- Details of local diabetes-related events for South Asian communities

Participants also discussed system design features, with views varying vastly and little consensus. In terms of the frequency of messages, views ranged from participants who preferred to have messages daily to those who wanted them weekly, fortnightly, or monthly. Views on the personalization of messages varied between those who were in favor of having their first name being used to those who did not want this. Messages sent by

general practitioners (GPs) or researchers were seen to be credible and trustworthy:

> *If we're going to get too many of these messages, people are just going to ignore them...[...]* [58-year-old man, Gujarati Hindu focus group]

> *That was what I was suggesting, that if you get too many of them you don't even look at it. Yeah, [then] they're a sheer waste of effort and time.* [69-year-old man, Gujarati Hindu focus group]

Language Preferences

Participants felt that messages in English would be acceptable as "everyone understands English" and that those who spoke no or little English could receive help to translate messages from their children, often providing examples of a son or daughter who could help in these situations. Several who preferred messages in English noted that the written version of some South Asian languages might be too formal and difficult for most people to understand, that some dialects have no written form, and that many people who spoke a South Asian language could not always read or write it. Some younger participants who spoke little English (and had recently migrated to the United Kingdom) believed that receiving SMS text messages in English might help them to improve their English and that Google Translate could be used when needed:

> *If it is translated into Bengali we will not learn English as we have an alternative.* [34-year-old Bangladeshi Muslim woman]

> *This is a very good point actually i.e. if it is said in our respective language, then the motivation to learn English would decrease because we will find whatever it is we want, so we probably won't learn.* [47-year-old Bangladeshi Muslim man]

The translation of short messages into South Asian languages was seen as a possible option to consider only if resources were available. Some participants felt that while multi-language options would provide choice and should perhaps be ideally available, these options were unnecessary given how brief the messages were and quick and easy for family members to translate for patients unable to read them:

> *English is okay, it's just information [laughs]. [If you don't read English] you'd show someone else briefly to tell you that's what it says about diabetes. It's easy, and that's it, finish.* [50-year-old Gujarati Muslim man]

Family Involvement

Family involvement, key for participants—particularly those unable to communicate well or fluently in English—was a theme discussed in all focus groups. This included the role of young and adult children in helping patients to self-care, and the support of spouses and carers. Participants often discussed how family members (eg, spouses) reminded the patient to take medications, cooked healthier meals since their diagnosis, and made or attended GP appointments with them. One participant, a 66-year-old Pakistani Muslim man, noted that his wife reminded him to take his medications, helped him watch his diet, and that messages should be sent to her phone rather than his. Another stated that messages for his father's physiotherapy appointments were already being sent to his phone because his father was unable to read them. He, like others, felt that brief messages about T2D should be sent to family members when the patient did not use or have a mobile phone, or was unable to read English, as family members could translate or explain messages to the patient:

> *For instance, my dad, he has physiotherapy appointments. My dad had a fall, but sending a text message to my dad is not good. Okay. So I have given them my number. So I get the message...I take him for his physiotherapy appointment.* [47-year-old Bangladeshi Muslim man]

Different Digital Formats for Different Groups

Although younger people generally engaged more readily with digital devices than their older counterparts, participants felt that messages in different formats could enable all members of their communities to benefit from text messages, noting the diversity within families and communities in terms of education and fluency in English. Discussions around different formats for different groups included SMS text messages in English for those who could read and understand them, audio messages in English for those who could understand but not read English, possibly audio messages in South Asian languages for those who required this, and illustrations or images to help those unable to read English or a South Asian language or dialect.

M1: *What should we do for those who are not educated and cannot read or write?*

M7: *We can send messages via Whatsapp.*

M1: *Do you mean send images via Whatsapp?*

M7: *Images or recorded messages [...]*

M6: *Images, animations.*

[M1: 47-year-old; M7: 41-year-old; M6: 46-year-old; Bangladeshi Muslim men's focus group]

Face-to-Face Groups for Those Who Do Not Use Digital Devices

Information provision was seen as essential for good DSM, but participants felt that it was not reaching all sections of their communities. They emphasized that not everyone, particularly older patients, had a mobile phone: "Older people don't even have mobile phones," carry them on their person, or charge them regularly. Two suggestions were offered in these cases: (1) sending messages to other family members who were involved in their care and (2) face-to-face meetings as an opportunity for people to share information and learn from one another, or preferably with a health professional present to answer questions.

Participants felt that, although an mHealth short messaging system could be helpful for people who "understand how to use a mobile phone" or other digital devices, face-to-face groups would be more helpful for those who did not. Some participants also felt that face-to-face groups would be helpful for those who

could not rely on their children to help them translate SMS text messages, noting that their children were often busy with their own lives. "Human contact" with health professionals and other patients was also seen as important, particularly in the context of receiving information:

> *I have come here twice and have informed others about what I learned here. They are all interested in joining. We can learn many things about the disease from which we are suffering. I am giving my opinion, like Mr X has said. When many people are involved in the discussion we can learn from them, like this son who has talked about many issues beautifully...If places are available, then I think it will be good for everyone...It will be better if a GP or doctor is present.* [81-year-old Bangladeshi Muslim man]

Other suggestions for disseminating message content included having educational programs or short adverts on South Asian television channels. Participants in the Bangladeshi Muslim men's focus group stated that most Bangladeshi Muslim households in Leicester had audio receivers through which prayers and sermons from the local mosques were delivered. They felt that information about diabetes could be disseminated by faith leaders in local mosques and via the audio receivers to family members at home.

Discussion

Comparison With Prior Work

This is the first study exploring British South Asian views of a digital health intervention for patients with T2D. Our focus groups included a diverse sample, including the "seldom heard" views of people unable to speak in English, some of whom were also unable to write in English or a South Asian language.

Previous research into the perspectives of British South Asians with T2D has focused on experiences of, and barriers and facilitators to, DSM, but to date no research has explored the views of British South Asian people in relation to digital health interventions to support DSM. Our findings corroborate those of the HeLP-Diabetes study, guided by the Corbin and Strauss framework [60], where the features of digital health interventions desired by the general population included specific content relating to diabetes, reliable, accessible dietary advice, and guidance on emotional management [61]. Corbin and Strauss [60] discussed three types of work involved in living with and managing chronic illness: illness work, everyday life work, and biographical work. Illness-related work—that of managing symptoms, diagnosis, medications, and crises—differs from everyday life work, which includes managing everyday living, emotions, and relationships. Biographical work is the work that is done to find meaning from the condition and life experience in light of the disruptions to a person's biographical narrative caused by chronic illness. For our participants, having messages that helped with illness-related work and everyday life work was of most importance in terms of DSM. Future DSM interventions aimed at South Asian populations could consider including culturally relevant information to support this "work" that people living with diabetes undertake in their everyday lives, such as information about the health benefits of different South Asian foods and diets, healthy portion sizes, fasting, and different kinds of physical activity.

The cultural adaptation of SMS text message interventions has generally included language translation, so that health messages reflect health beliefs, norms, and social practices [62]. However, there is little evidence to support the effectiveness and desirability of cultural adaptation of such interventions and a lack of research into the cultural appropriateness of messages [9]. Although a culturally adapted SMS has never been trialed in the United Kingdom, culturally adapted health promotion interventions aimed at British South Asians have shown only moderate effect and were inconclusive in demonstrating cost effectiveness [33]. None of these interventions involved head-to-head comparisons with interventions aimed at the general population. Study materials included adaptation into South Asian languages, which researchers state is challenging, time-consuming, costly, and not just a process of linguistic translation [63]. Participants in our study felt that messages delivered in South Asian languages were not essential or necessary for a short messaging intervention in which messages were less than 160 characters.

Previously adapted health promotion interventions for British South Asians with T2D have emphasized the importance of family involvement because of the strong cultural emphasis on family life. In a lifestyle intervention on weight change in British South Asian people at high risk of T2D, families rather than individuals were randomized in the trial. However, researchers working on the trial were unable to recruit family volunteers for many families, concluding that the added value of family involvement remains to be explored [64]. Our participants were positive about the involvement of family members or carers possibly because very little would be required of them to translate short messages. Recent work in intervention development recommends putting the needs and wants of end users at the heart of any intervention [50], highlighting that interventions that can be well integrated into everyday life and health care routines, that are easy to use, compatible with patients' existing skills, and that do not significantly disrupt patients' lives are more likely to lead to successful implementation [65]. Interventions aimed at South Asian populations may also need to consider these design aspects in relation to family and carers too.

A range of factors need to be considered when culturally adapting a digital health intervention, including levels of mobile phone or technology use. Access to smartphones can be influenced by various factors, including age, gender, education, and affordability [27]. It is also important to consider the power dynamics that determine different groups' access to technologies [26] and how health inequalities may be created or perpetuated. A study of the use of electronic health among patients in Norway with type 1 and T2D found a strong association between a high level of education and the use of search engines but no educational differences for the use of apps, social media, and video services, indicating that adequate communication strategies for audiences with varying education levels should be a focus in efforts to reduce health inequalities in health outcomes [66]. As our participants suggested—users and nonusers alike—different digital formats for different groups

have to be considered when thinking about the needs of heterogeneous populations with varied levels of literacy and education, as well as addressing the needs of family and community members who do not use digital devices, such as face-to-face meetings.

Our formative research sheds light on the varied needs of British South Asian patients in relation to short messages to support DSM. These findings are consistent with those from focus groups and interviews that we conducted separately with the general population, where participants also expressed a preference for message content that addressed all aspects of DSM and not just medication adherence [67]. This suggests that an mHealth SMS designed for the general population, such as the SuMMiT-D system, can be acceptable and relevant to UK South Asian populations but may need to include additional content with culturally adapted messages about South Asian foods, natural and herbal approaches used in the United Kingdom and South Asia, safe medicine taking when fasting, and exercise in women-only groups—messages that can be available to those who opt to receive them from a system that can provide individual choice. Other design implications worth considering include using images and audio messages for patients unable to read message content in English, as suggested by our participants. There is limited evidence of the extent to which culturally adapted messages might lead to specific changes in behavior, but they are likely to enhance engagement with the wider intervention. Exploring the views of British South Asian patients on an SMS aimed at the general UK population is important in maximizing the potential of such an intervention. Although head-to-head comparisons of adapted and nonadapted interventions are rare, our qualitative data with British South Asian communities and the general population suggest that formative work comparisons can yield helpful insights into which interventions need to be culturally adapted, why, and what this might involve. Future research should explore how best to codesign and test culturally adapted messages that could be incorporated into a general digital messaging system aimed at all UK patients with T2D.

Limitations

The study was exploratory and, although it included some family/carer views, focus groups specifically with families would shed further light into their views on receiving short messages and translating these for patients. Although our sample included a diverse range of views, fewer second generation and no third generation participants took part in the study, people who may have had further ideas about system content and usability. As family/carers, third generation British South Asians could shed further light into the translating/explaining of message content to first- or second-generation family members (eg, to grandparents). As young patients, they may have differing views about the need for messages on, for example, South Asian foods and diets. As regular users of the internet, they may also feel that they have access to all the information that they need and have no need for short digital messages. The focus groups were held during the day. Evening focus groups might have encouraged more second and third generation people to attend, potentially a more convenient time for those in full-time education and employment.

Conclusions

An mHealth short messaging intervention that addresses all aspects of DSM is more relevant and acceptable to British South Asian people with T2D than one that focuses only on medication adherence. For such an intervention to meet the needs of UK South Asian populations, it may also have to include culturally relevant messages sent to those who opt to receive them. It is equally important to consider how to disseminate message content to patients who do not use digital devices to help reduce health inequalities, including face-to-face groups.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR; Programme Grants for Applied Research programme [RP-PG-1214-20003]). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. AF is an NIHR Senior Investigator and receives funding from the NIHR Oxford Biomedical Research Centre. The authors would like to thank the participants who took part in this study, Nasima Miah who helped with the recruitment and cofacilitation of focus groups, and Ebrahim Ali for cofacilitating two focus groups. They would also like to thank the CBMEH, University of Leicester, Diabetes Research Centre, for their support with this work. They also thank their reviewers for their helpful comments.

Authors' Contributions

AF, RC, and SP designed the focus group study, and SP facilitated the focus groups. Transcripts were coded by SP and NN. Drafting the manuscript was led by SP, with all other authors contributing comments and reflections, and approving the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participants' quotes about system content and usability.
[DOCX File , 19 KB - mhealth_v8i4e15789_app1.docx]

References

1. International Diabetes Federation. IDF Diabetes Atlas 9th Edition 2019. Brussels, Belgium: International Diabetes Federation; 2019. URL: https://www.diabetesatlas.org/en/ [accessed 2020-02-11]
2. Sohal T, Sohal P, King-Shier KM, Khan NA. Barriers and facilitators for type-2 diabetes management in South Asians: a systematic review. PLoS One 2015;10(9):e0136202 [FREE Full text] [doi: 10.1371/journal.pone.0136202] [Medline: 26383535]
3. Khan NA, Wang H, Anand S, Jin Y, Campbell NR, Pilote L, et al. Ethnicity and sex affect diabetes incidence and outcomes. Diabetes Care 2011 Jan;34(1):96-101 [FREE Full text] [doi: 10.2337/dc10-0865] [Medline: 20978094]
4. Forouhi NG, Sattar N, Tillin T, McKeigue PM, Chaturvedi N. Do known risk factors explain the higher coronary heart disease mortality in South Asian compared with European men? Prospective follow-up of the Southall and Brent studies, UK. Diabetologia 2006 Nov;49(11):2580-2588. [doi: 10.1007/s00125-006-0393-2] [Medline: 16972045]
5. Gov.UK. 2018. Population of England and Wales URL: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest#main-facts-and-figures [accessed 2020-01-27]
6. Hills AP, Arena R, Khunti K, Yajnik CS, Jayawardena R, Henry CJ, et al. Epidemiology and determinants of type 2 diabetes in south Asia. Lancet Diabetes Endocrinol 2018 Dec;6(12):966-978. [doi: 10.1016/S2213-8587(18)30204-3] [Medline: 30287102]
7. Bailey TS, Zisser HC, Garg SK. Reduction in hemoglobin A1C with real-time continuous glucose monitoring: results from a 12-week observational study. Diabetes Technol Ther 2007 Jun;9(3):203-210. [doi: 10.1089/dia.2007.0205] [Medline: 17561790]
8. Macisaac RJ, Jerums G. Intensive glucose control and cardiovascular outcomes in type 2 diabetes. Heart Lung Circ 2011 Oct;20(10):647-654. [doi: 10.1016/j.hlc.2010.07.013] [Medline: 20807681]
9. Dobson R, Whittaker R, Jiang Y, Shepherd M, Maddison R, Carter K, et al. Text message-based diabetes self-management support (SMS4BG): study protocol for a randomised controlled trial. Trials 2016 Apr 2;17:179 [FREE Full text] [doi: 10.1186/s13063-016-1305-5] [Medline: 27039300]
10. Gatwood J, Balkrishnan R, Erickson SR, An LC, Piette JD, Farris KB. The impact of tailored text messages on health beliefs and medication adherence in adults with diabetes: A randomized pilot study. Res Social Adm Pharm 2016;12(1):130-140. [doi: 10.1016/j.sapharm.2015.04.007] [Medline: 25980582]
11. Farmer AJ, McSharry J, Rowbotham S, McGowan L, Ricci-Cabello I, French DP. Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with Type 2 diabetes: a systematic review of randomized trials. Diabet Med 2015 Oct 15;33(5):565-579. [doi: 10.1111/dme.12987] [Medline: 26470750]
12. Ramachandran A, Kumar R, Nanditha A, Raghavan A, Snehalatha C, Krishnamoorthy S, et al. mDiabetes initiative using text messages to improve lifestyle and health-seeking behaviour in India. BMJ Innov 2018;4(4):155-162. [doi: 10.1136/bmjinnov-2017-000265]
13. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 9;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
14. Thakkar J, Barry T, Thiagalingam A, Redfern J, McEwan AL, Rodgers A, et al. Design considerations in development of a mobile health intervention program: the TEXT ME and TEXTMEDS experience. JMIR Mhealth Uhealth 2016 Nov 15;4(4):e127 [FREE Full text] [doi: 10.2196/mhealth.5996] [Medline: 27847350]
15. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: a global overview. Diabetes Metab Res Rev 2018 May;34(4):e2990. [doi: 10.1002/dmrr.2990] [Medline: 29431916]
16. Office of National Statistics. 2017. Internet access – households and individuals, Great Britain: 2017 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [accessed 2020-01-27]
17. Rami B, Popow C, Horn W, Waldhoer T, Schober E. Telemedical support to improve glycemic control in adolescents with type 1 diabetes mellitus. Eur J Pediatr 2006 Oct;165(10):701-705. [doi: 10.1007/s00431-006-0156-6] [Medline: 16670859]
18. Nundy S, Dick JJ, Chou C, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. Health Aff (Millwood) 2014 Feb;33(2):265-272 [FREE Full text] [doi: 10.1377/hlthaff.2013.0589] [Medline: 24493770]
19. Razum O, Spallek J. Addressing health-related interventions to immigrants: migrant-specific or diversity-sensitive? Int J Public Health 2014 Dec;59(6):893-895. [doi: 10.1007/s00038-014-0584-4] [Medline: 25012801]
20. Osei-Assibey G, Kyrou I, Adi Y, Kumar S, Matyka K. Dietary and lifestyle interventions for weight management in adults from minority ethnic/non-White groups: a systematic review. Obes Rev 2010 Nov;11(11):769-776. [doi: 10.1111/j.1467-789X.2009.00695.x] [Medline: 20059708]
21. Ockene IS, Tellez TL, Rosal MC, Reed GW, Mordes J, Merriam PA, et al. Outcomes of a Latino community-based intervention for the prevention of diabetes: the Lawrence Latino Diabetes Prevention Project. Am J Public Health 2012 Feb;102(2):336-342. [doi: 10.2105/AJPH.2011.300357] [Medline: 22390448]

22. Castro FG, Barrera M, Holleran SLK. Issues and challenges in the design of culturally adapted evidence-based interventions. Annu Rev Clin Psychol 2010;6:213-239 [FREE Full text] [doi: 10.1146/annurev-clinpsy-033109-132032] [Medline: 20192800]
23. Terragni L, Beune E, Stronks K, Davidson E, Qureshi S, Kumar B, et al. Developing culturally adapted lifestyle interventions for South Asian migrant populations: a qualitative study of the key success factors and main challenges. Public Health 2018 Aug;161:50-58. [doi: 10.1016/j.puhe.2018.04.008] [Medline: 29902781]
24. Verbiest M, Borrell S, Dalhousie S, Tupa'i-Firestone R, Funaki T, Goodwin D, et al. A co-designed, culturally-tailored mHealth tool to support healthy lifestyles in Māori and Pasifika communities in New Zealand: protocol for a cluster randomized controlled trial. JMIR Res Protoc 2018 Aug 22;7(8):e10789 [FREE Full text] [doi: 10.2196/10789] [Medline: 30135054]
25. Gorton D, Dixon R, Maddison R, Mhurchu CN, Jull A. Consumer views on the potential use of mobile phones for the delivery of weight-loss interventions. J Hum Nutr Diet 2011 Dec;24(6):616-619. [doi: 10.1111/j.1365-277X.2011.01163.x] [Medline: 21781188]
26. Dobson R, Whittaker R, Bartley H, Connor A, Chen R, Ross M, et al. Development of a culturally tailored text message maternal health program: TextMATCH. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e49 [FREE Full text] [doi: 10.2196/mhealth.7205] [Medline: 28428159]
27. McCool J, Tanielu H, Umali E, Whittaker R. Assessing the cross-cultural adaptation and translation of a text-based mobile smoking cessation program in Samoa (TXTTaofiTapaa): pilot study. JMIR Mhealth Uhealth 2018 Aug 31;6(8):e173 [FREE Full text] [doi: 10.2196/mhealth.9033] [Medline: 30170994]
28. Rush E, Reed P, McLennan S, Coppinger T, Simmons D, Graham D. A school-based obesity control programme: Project Energize. Two-year outcomes. Br J Nutr 2012 Feb;107(4):581-587. [doi: 10.1017/S0007114511003151] [Medline: 21733268]
29. Ellison-Loschmann L, Pearce N. Improving access to health care among New Zealand's Maori population. Am J Public Health 2006 Apr;96(4):612-617. [doi: 10.2105/AJPH.2005.070680] [Medline: 16507721]
30. Bellary S, O'Hare JP, Raymond NT, Gumber A, Mughal S, Szczepura A, UKADS Study Group. Enhanced diabetes care to patients of south Asian ethnic origin (the United Kingdom Asian Diabetes Study): a cluster randomised controlled trial. Lancet 2008 May 24;371(9626):1769-1776. [doi: 10.1016/S0140-6736(08)60764-3] [Medline: 18502301]
31. Muilwijk M, Stronks K, Qureshi SA, Beune E, Celis-Morales C, Gill J, et al. Dietary and physical activity strategies to prevent type 2 diabetes in South Asian adults: protocol for a systematic review. BMJ Open 2017 Jul 2;7(6):e012783 [FREE Full text] [doi: 10.1136/bmjopen-2016-012783] [Medline: 28674123]
32. Brown T, Smith S, Bhopal R, Kasim A, Summerbell C. Diet and physical activity interventions to prevent or treat obesity in South Asian children and adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2015 Jan 9;12(1):566-594 [FREE Full text] [doi: 10.3390/ijerph120100566] [Medline: 25584423]
33. Wallia S, Bhopal RS, Douglas A, Bhopal R, Sharma A, Hutchison A, et al. Culturally adapting the prevention of diabetes and obesity in South Asians (PODOSA) trial. Health Promot Int 2014 Dec;29(4):768-779. [doi: 10.1093/heapro/dat015] [Medline: 23574693]
34. Barrera M, Castro FG, Strycker LA, Toobert DJ. Cultural adaptations of behavioral health interventions: a progress report. J Consult Clin Psychol 2013 Apr;81(2):196-205 [FREE Full text] [doi: 10.1037/a0027085] [Medline: 22289132]
35. Liu J, Davidson E, Bhopal R, White M, Johnson M, Netto G, et al. Adapting health promotion interventions to meet the needs of ethnic minority groups: mixed-methods evidence synthesis. Health Technol Assess 2012;16(44):1-469 [FREE Full text] [doi: 10.3310/hta16440] [Medline: 23158845]
36. Lloyd CE, Mughal S, Roy T, Raymond NT, O'Hare JP, Barnett AH, et al. What factors influence concordance with medications? Findings from the UK Asian Diabetes study. Diabet Med 2014 Dec;31(12):1600-1609. [doi: 10.1111/dme.12554] [Medline: 25073479]
37. Porqueddu T. Herbal medicines for diabetes control among Indian and Pakistani migrants with diabetes. Anthropol Med 2017 Apr;24(1):17-31. [doi: 10.1080/13648470.2016.1249338] [Medline: 28292208]
38. Patel N, Ferrer HB, Tyrer F, Wray P, Farooqi A, Davies MJ, et al. Barriers and facilitators to healthy lifestyle changes in minority ethnic populations in the UK: a narrative review. J Racial Ethn Health Disparities 2017 Dec;4(6):1107-1119 [FREE Full text] [doi: 10.1007/s40615-016-0316-y] [Medline: 27928772]
39. Lawton J, Ahmad N, Hanna L, Douglas M, Bains H, Hallowell N. 'We should change ourselves, but we can't': accounts of food and eating practices amongst British Pakistanis and Indians with type 2 diabetes. Ethn Health 2008 Sep;13(4):305-319. [doi: 10.1080/13557850701882910] [Medline: 18701991]
40. Lawton J, Ahmad N, Hanna L, Douglas M, Hallowell N. 'I can't do any serious exercise': barriers to physical activity amongst people of Pakistani and Indian origin with Type 2 diabetes. Health Educ Res 2006 Feb;21(1):43-54. [doi: 10.1093/her/cyh042] [Medline: 15955792]
41. Lawton J, Ahmad N, Hallowell N, Hanna L, Douglas M. Perceptions and experiences of taking oral hypoglycaemic agents among people of Pakistani and Indian origin: qualitative study. Br Med J 2005 May 28;330(7502):1247 [FREE Full text] [doi: 10.1136/bmj.38460.642789.E0] [Medline: 15883139]

42. Singh H, Cinnirella M, Bradley C. Support systems for and barriers to diabetes management in South Asians and Whites in the UK: qualitative study of patients' perspectives. BMJ Open 2012;2(6):pii: e001459 [FREE Full text] [doi: 10.1136/bmjopen-2012-001459] [Medline: 23151392]
43. EuroDHYAN. URL: https://www.eurodhyan.eu/ [accessed 2020-08-02]
44. Tabak RG, Sinclair KA, Baumann AA, Racette SB, Sebert KA, Johnson-Jennings MD, et al. A review of diabetes prevention program translations: use of cultural adaptation and implementation research. Transl Behav Med 2015 Dec;5(4):401-414 [FREE Full text] [doi: 10.1007/s13142-015-0341-0] [Medline: 26622913]
45. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. Br Med J 2007 Mar 3;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
46. 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. Br Med J 2008;337:a1655 [FREE Full text] [Medline: 18824488]
47. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
48. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
49. O'Brien N, Heaven B, Teal G, Evans EH, Cleland C, Moffatt S, et al. Integrating evidence from systematic reviews, qualitative research, and expert knowledge using co-design techniques to develop a web-based intervention for people in the retirement transition. J Med Internet Res 2016 Aug 03;18(8):e210 [FREE Full text] [doi: 10.2196/jmir.5790] [Medline: 27489143]
50. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
51. Ricci-Cabello I, Bobrow K, Islam SM, Chow CK, Maddison R, Whittaker R, et al. Examining development processes for text messaging interventions to prevent cardiovascular disease: systematic literature review. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e12191 [FREE Full text] [doi: 10.2196/12191] [Medline: 30924790]
52. Office for National Statistics. 2011 Census URL: https://www.ons.gov.uk/census/2011census [accessed 2020-01-27]
53. Nuffield Department of Primary Care Health Sciences. Support through Mobile Messaging and dIgital health Technology for Diabetes URL: https://www.phc.ox.ac.uk/research/participate/summit-d [accessed 2020-01-27]
54. 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]
55. Long H, Bartlett YK, Farmer AJ, French DP. Identifying brief message content for interventions delivered via mobile devices to improve medication adherence in people with type 2 diabetes mellitus: a rapid systematic review. J Med Internet Res 2019 Jan 09;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]
56. Coyne IT. Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? J Adv Nurs 1997 Sep;26(3):623-630. [doi: 10.1046/j.1365-2648.1997.t01-25-00999.x] [Medline: 9378886]
57. Prinjha S, Miah N, Ali E, Farmer A. Including 'seldom heard' views in research: opportunities, challenges and recommendations from focus groups with British South Asian people with type 2 diabetes. BMC Medical Research Methodology (forthcoming) 2020 . [doi: 10.4324/9781315611358-4]
58. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
59. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. Br Med J 2000 Jan 08;320(7227):114-116 [FREE Full text] [doi: 10.1136/bmj.320.7227.114] [Medline: 10625273]
60. Corbin J, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. San Francisco and London: Jossey-Bass; 1988.
61. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
62. Whittaker R, Matoff-Stepp S, Meehan J, Kendrick J, Jordan E, Stange P, et al. Text4baby: development and implementation of a national text messaging health information service. Am J Public Health 2012 Dec;102(12):2207-2213. [doi: 10.2105/AJPH.2012.300736] [Medline: 23078509]
63. Hunt SM, Bhopal R. Self report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. J Epidemiol Community Health 2004 Jul;58(7):618-622 [FREE Full text] [doi: 10.1136/jech.2003.010074] [Medline: 15194728]
64. Bhopal RS, Douglas A, Wallia S, Forbes JF, Lean MEJ, Gill JMR, et al. Effect of a lifestyle intervention on weight change in south Asian individuals in the UK at high risk of type 2 diabetes: a family-cluster randomised controlled trial. Lancet Diabetes Endocrinol 2014 Mar;2(3):218-227 [FREE Full text] [doi: 10.1016/S2213-8587(13)70204-3] [Medline: 24622752]

65. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015;10:59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
66. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen O, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. J Med Internet Res 2019 May 29;21(5):e13615 [FREE Full text] [doi: 10.2196/13615] [Medline: 31144669]
67. Bartlett YK, Newhouse N, Long HA, Farmer AJ, French DP. What do people with type 2 diabetes want from a brief messaging system to support medication adherence? Patient Prefer Adherence 2019;13:1629-1640 [FREE Full text] [doi: 10.2147/PPA.S217843] [Medline: 31686789]

Abbreviations

CBMEH: Centre for Black and Minority Ethnic Health
DSM: diabetes self-management
GP: general practitioner
mHealth: mobile health
NIHR: National Institute for Health Research
SuMMiT-D: Support Through Mobile Messaging and Digital Health Technology for Diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by S Drabble, S Kanthawala; comments to author 26.09.19; revised version received 01.11.19; accepted 26.01.20; published 20.04.20.

Please cite as:
Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A
British South Asian Patients' Perspectives on the Relevance and Acceptability of Mobile Health Text Messaging to Support Medication Adherence for Type 2 Diabetes: Qualitative Study
JMIR Mhealth Uhealth 2020;8(4):e15789
URL: http://mhealth.jmir.org/2020/4/e15789/
doi:10.2196/15789
PMID:32310150

Original Paper

The Development and Evaluation of a Text Message Program to Prevent Perceived Insufficient Milk Among First-Time Mothers: Retrospective Analysis of a Randomized Controlled Trial

Jill R Demirci[1], PhD; Brian Suffoletto[2], MD, MSc; Jack Doman[3], MS; Melissa Glasser[1], PhD; Judy C Chang[4], MD, MPH; Susan M Sereika[5], PhD; Debra L Bogen[6], MD

[1]Department of Health Promotion & Development, University of Pittsburgh School of Nursing, Pittsburgh, PA, United States

[2]Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[3]Office of Academic Computing, Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, PA, United States

[4]Department of Obstetrics, Gynecology & Reproductive Sciences, and Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[5]Department of Health & Community Systems, Center for Research and Evaluation, University of Pittsburgh School of Nursing, Pittsburgh, PA, United States

[6]Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Jill R Demirci, PhD
Department of Health Promotion & Development
University of Pittsburgh School of Nursing
3500 Victoria St
Suite 440
Pittsburgh, PA, 15261
United States
Phone: 1 4126489236
Email: Jvr5@pitt.edu

Abstract

Background: Several recent trials have examined the feasibility and efficacy of automated SMS *text messaging* to provide remote breastfeeding support to mothers, but these texting systems vary in terms of design features and outcomes examined.

Objective: This study examined user engagement with and feedback on a theory-grounded SMS text messaging intervention intended to prevent perceived insufficient milk (PIM)—the single, leading modifiable cause of unintended breastfeeding reduction and cessation.

Methods: We recruited 250 nulliparous individuals intending to breastfeed between 13 and 25 weeks of pregnancy in southwestern Pennsylvania. Participants were randomly assigned with equal allocation to either an SMS intervention to prevent PIM and unintended breastfeeding reduction or cessation (MILK, a Mobile, semiautomated text message–based Intervention to prevent perceived Low or insufficient milK supply; n=126) or a control group receiving general perinatal SMS text messaging–based support via the national, free Text4Baby system (n=124). Participants in both groups received SMS text messages 3 to 7 times per week from 25 weeks of pregnancy to 8 weeks postpartum. The MILK intervention incorporated several automated interactivity and personalization features (eg, keyword texting for more detailed information on topics and branched response logic) as well as an option to receive one-on-one assistance from an on-call study lactation consultant. We examined participant interactions with the MILK system, including response rates to SMS text messaging queries. We also sought participant feedback on MILK content, delivery preferences, and overall satisfaction with the system via interviews and a remote survey at 8 weeks postpartum.

Results: Participants randomized to MILK (87/124, 70.2% white and 84/124, 67.7% college educated) reported that MILK texts increased their breastfeeding confidence and helped them persevere through breastfeeding problems. Of 124 participants, 9 (7.3%) elected to stop MILK messages, and 3 (2.4%) opted to reduce message frequency during the course of the study. There were 46 texts through the MILK system for individualized assistance from the study lactation consultant (25/46, 54% on weekends or after-hours). The most commonly texted keywords for more detailed information occurred during weeks 4 to 6 postpartum and addressed milk volume intake and breastfeeding and sleep patterns. MILK participants stated a preference for anticipatory guidance on potential breastfeeding issues and less content addressing the benefits of breastfeeding. Suggested improvements

included extending messaging past 8 weeks, providing access to messaging for partners, and tailoring content based on participants' pre-existing breastfeeding knowledge and unique breastfeeding trajectory.

Conclusions: Prenatal and postpartum evidence–based breastfeeding support delivered via semiautomated SMS text messaging is a feasible and an acceptable intervention for first-time mothers. To optimize engagement with digital breastfeeding interventions, enhanced customization features should be considered.

Trial Registration: ClinicalTrials.gov NCT02724969; https://clinicaltrials.gov/ct2/show/NCT02724969

(JMIR Mhealth Uhealth 2020;8(4):e17328) doi:10.2196/17328

KEYWORDS

breast feeding; perceived insufficient milk; text messaging; short message service; cell phone; mobile phone; telemedicine; mHealth; randomized controlled trial

Introduction

Over 80% of mothers in the United States begin breastfeeding after their baby is born [1], indicating a broad awareness of its health and economic advantages [2,3]. However, only few mothers in the United States meet the national recommendations for breastfeeding duration (1+ years) and exclusivity (6 months) [4]. By 6 months, only 25% of mothers are breastfeeding without formula, and by 12 months, 36% are breastfeeding at all [1]. Only one-third of mothers reach their intended breastfeeding goals [5,6]. These data indicate that, although most mothers wish to breastfeed, they face significant barriers to doing so.

Perceived insufficient milk (PIM) is the single most common reason for formula supplementation and its corollary, premature breastfeeding cessation [5,7,8]. Approximately 35% of cases of weaning before intended breastfeeding duration are attributable to PIM [9]. Excluding those with adverse metabolic profiles (eg, diabetes and polycystic ovarian syndrome), PIM is rarely rooted in primary anatomical or physiological abnormalities [10,11]; rather, there is a strong relationship between PIM, low maternal breastfeeding self-efficacy, and misinformation about normal infant breastfeeding behavior, milk volume trajectories, and principles of milk production [12-14]. When left unchecked, these issues have the potential to propagate a ripple effect of formula supplementation, less frequent milk removal from the breast, and responsive physiological reductions in breast milk volume [9,15].

Few interventions exist to prevent or correct PIM and its consequences. One exception to this is a recent single group, pretest/posttest pilot study that examined a home-based education intervention to reduce PIM and increase maternal breastfeeding self-efficacy. The authors found that among 14 included mothers, self-efficacy increased and the attribution of infant crying to PIM decreased over time [16]. However, with no comparison group, it is difficult to separate the intervention from time effects. Moreover, the logistics of disseminating such an intervention are dependent on qualified personnel and resources, which are unavailable at many sites.

Automated SMS *text messaging* may be an effective platform to deliver targeted, evidence-based breastfeeding education and support to pregnant and lactating parents addressing the root causes of PIM. Mobile health interventions, most predominantly SMS text messaging–based interventions, to address global maternal-child health issues have proliferated in recent years. This trend reflects the ubiquity of cell phones and SMS text messaging, particularly among Generation X and Y women [17-19]. Several recent trials conducted both in developed and developing countries demonstrate significant increases in exclusive breastfeeding from 9 weeks to 6 months postpartum among women who received automated SMS text message breastfeeding support vs control group women [20-22]. However, to date, no published studies have examined SMS text messaging to address PIM specifically, and few have incorporated theory-based content and advanced functionality, such as automated interactivity and personalization, to engage breastfeeding mothers. In this study, we have reported on the usability and acceptability as well as design and implementation considerations of a semiautomated SMS text message system (ie, employing both automated responses as well as opportunities for live interactions) to prevent PIM, examined within the MILK (a Mobile, semiautomated text message–based Intervention to prevent perceived Low or insufficient milK supply) trial.

Methods

Design

MILK was a randomized controlled trial examining the effect of a theory-driven SMS text message breastfeeding support system vs an attention control condition (general perinatal text-based support from the national Text4Baby system) on PIM and breastfeeding outcomes among first-time mothers in the United States. We hypothesized that MILK would be feasible and, at 8 weeks postpartum, MILK participants would have a perception of greater breast milk volume/supply, higher self-reported breastfeeding confidence and satisfaction, lower anxiety related to breastfeeding, and higher rates of any and exclusive breastfeeding compared with the control group. Data collection was completed in May 2019, with breastfeeding outcome data pending (ClinicalTrials.gov registration: NCT02724969). This study reported on feasibility data only.

Participants and Setting

Participants were recruited at prenatal visits during the second pregnancy trimester at University of Pittsburgh Medical Center Magee-Women's Hospital prenatal clinics. Recruitment also occurred through local advertising (posted flyers and bus advertisements), a university research registry, and social media (TrialSpark). Eligible individuals were nulliparous, aged 18 years or older, English-speaking, 13 to 25 gestational weeks, pregnant with 1 infant, owned a cell phone with internet access

(ie, a *smartphone*) and an unlimited SMS text message plan, and intended to exclusively or nearly exclusively breastfeed (<2 ounces of formula per day) for at least two months. Breastfeeding intention was assessed via a list of potential feeding options in the screener form; no information was provided for or against breastfeeding during the study introduction before screening or randomization. The exclusion criteria included any contraindications to breastfeeding [4] and maternal or fetal conditions likely to compromise breastfeeding or milk supply (eg, breast reduction surgery and major congenital anomalies). All participants provided written informed consent for study participation. The study was approved by the University of Pittsburgh Institutional Review Board.

Procedures

After screening for eligibility and enrollment at 13 to 25 weeks of pregnancy, participants were randomized with equal allocation to the intervention or control group. Enrollment timing was based on both capturing a participant pool with viable pregnancies and maximizing between-group comparability in intervention timing (ie, commencement of breastfeeding-specific texts at 25 weeks for both groups). Control group participants received sign-up information for the freely available national Text4Baby program, which delivered automated texts immediately following enrollment; these texts included content on various aspects of infant care, including breastfeeding (seven prenatal breastfeeding-specific messages and three postpartum breastfeeding-specific messages before the primary study end point at 8 weeks postpartum). The Text4Baby program was developed by the National Healthy Moms, Healthy Babies Coalition in collaboration with the Centers for Disease Control and Prevention through a rigorous process including expert review, research, and input from pregnant and postpartum women. Text4Baby messages are continually updated in partnership with various public health organizations, and evidence suggests that the messages impact both maternal attitudes and health behaviors [23-25]. Participants assigned to MILK received messages beginning at 25 weeks of gestation that focused specifically on establishing breastfeeding confidence and behaviors to prevent PIM. All participants received prenatal and postpartum text messages 3 to 7 times per week. Although there was no postpartum cutoff point for Text4Baby messages, MILK texts continued till 8 weeks postpartum.

Baseline data, including demographics and health history, were collected via electronic medical record abstraction and maternal self-reporting at the enrollment visit. A remotely administered 8-week postpartum survey assessed participants' overall enthusiasm for the MILK system and content preferences. MILK participants were additionally invited to participate in an audio-recorded phone or in-person interview about their study experiences at 8 weeks postpartum or at the time of breastfeeding cessation. Interviews were conducted by the study coordinator, followed a semistructured script, and were professionally transcribed.

MILK Development

Development of the MILK intervention was based on the breastfeeding self-efficacy (social cognitive theory) conceptual model that theorizes self-efficacy as a driving force of breastfeeding behavior and PIM as a consequence of impaired breastfeeding self-efficacy [26,27]. In this framework, text content targeted antecedents of self-efficacy, including performance accomplishments (eg, development of early technical, hands-on breastfeeding skills), vicarious experiences (eg, video- and vignette-based breastfeeding exposure), verbal persuasion (eg, information on breastfeeding benefits), and physiological and affective states (eg, strategies to reduce anxiety around breastfeeding and milk production). Text content and delivery strategies were also informed by (1) principles of health communication and behavioral economics, (2) mapping of the breastfeeding trajectory and problems encountered by primiparous mothers in an ecological momentary assessment (EMA) study [28], (3) consultation with experts in human lactation and social marketing (Best for Babes Foundation), and (4) a focus group with pregnant and postpartum individuals to obtain feedback on draft messages.

The final text bank contained a series of 63 messages to be delivered from 25 to 40 weeks of pregnancy, and a series of 47 messages to be delivered from the day of delivery to 8 weeks postpartum. Content differed by the perinatal stage—antenatal messaging focused on positive reinforcement regarding the decision to breastfeed, the impact of breastfeeding on maternal and child health, current breastfeeding recommendations and goal setting, lactation-related body changes during pregnancy, and anticipatory guidance for how breastfeeding looks and works. In the postpartum period, texts encompassed anticipatory guidance about breastfeeding milestones and infant breastfeeding behavior, milk volume expectations, technical aspects of breastfeeding and problem resolution (eg, positioning and latch), referrals to local and online breastfeeding resources, and encouragement to begin and continue breastfeeding.

MILK Platform and Design Features

The MILK text message system was built by investigators using an encrypted Structured Query Language server database at the University of Pittsburgh, with all outgoing texts and incoming participant responses managed through a Microsoft Access front-end server. The server stored all prespecified libraries of automated messages and contingent responses. Messages were typically scheduled to be delivered at 10 AM, with those longer than 160 characters split into several separate SMS text messages. The server additionally sent the study staff email notifications of participant replies and allowed the exchange of nonautomated texts between participants and study personnel. To ensure timely transition from prenatal to postpartum messages, participants were sent SMS reminders in the third trimester to text the keyword BIRTH after delivery; hospital delivery records were also checked daily to manually drive this transition if necessary.

MILK texts included free-standing content as well as embedded links to Web pages, infographics, photos, and videos. Texts also featured automated personalization and interactivity, such that mothers and their infants were addressed by name, the content

was tailored to the infant's gestational and chronological age, and participants could text keywords to receive more detailed topical information by email or text (37 texts included keyword prompts). A portion of texts also attempted to engage mothers by requesting a response and using branched logic algorithms to respond in kind. For example, in one text message series, participants were queried about the number of breastfeeding sessions in the last 24 hours, and automated responses provided feedback on whether the frequency was considered adequate, along with potential recourses if it was not. Additional features of the MILK system included the ability for participants to text a key phrase to stop messages or reduce message frequency to one series per week. Participants could also text HELP to communicate directly with a study-based international board-certified lactation consultant (IBCLC) via an SMS text message, an email, or a telephone call.

Analysis

MILK metadata addressing interactions with the system were abstracted from the Access database. Summary statistics were calculated for quantitative survey data using IBM SPSS Statistics version 24 (IBM Corporation, Armonk, New York, 2016). We used an editing approach for the qualitative coding within the qualitative coding program ATLAS.ti (ATLAS.ti GmbH, Berlin, 2019) [29], adding and refining codes in an iterative manner. A total of 2 authors trained in qualitative methods coded each interview transcript independently and met to compare the coding and organize codes into major and minor themes. Any discrepancies in interpretation were to be adjudicated by a third investigator; and none occurred.

Results

Recruitment and Sample Characteristics

A total of 250 maternal participants were recruited and randomized over a 15-month period (n=126 for MILK; n=124 for Text4Baby). The majority of the sample was recruited at prenatal visits (198/250, 79.2%). Among those assessed for eligibility, 3.2% (8/250) declined participation and 19.2% (48/250) were ineligible (Figure 1). The most common reason for ineligibility was not planning to exclusively breastfeed (n=31); just 3 individuals were ineligible because of not having a smartphone or an unlimited SMS plan. Participants were predominantly white, college educated, and married (Table 1).

Figure 1. Study flow diagram. MILK: a Mobile, semiautomated text message–based Intervention to prevent perceived Low or insufficient milK supply.

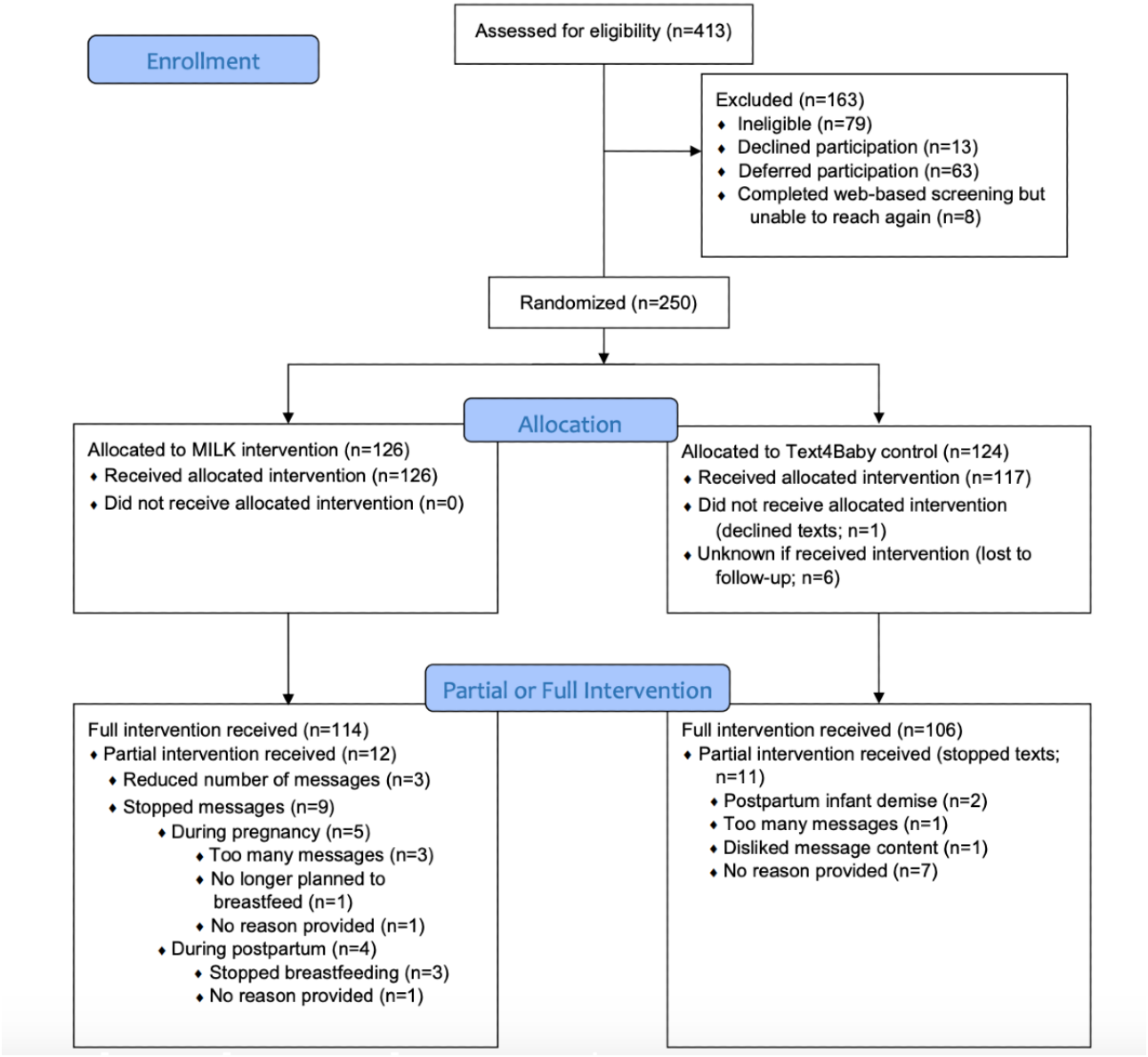


Table 1. Demographics of study participants at baseline (13-25 gestational weeks; N=247).

Characteristic	MILK[a] (n=124)	Text4Baby (n=123)	Total sample (n=247[b])	*P* value[c]
Age (years), mean (SD)	28.9 (5.5)	28.7 (5.2)	28.8 (5.3)	.76
Marital status, n (%)				.81
Married	81 (65.3)	77 (62.6)	158 (64.0)	
Living with a partner	19 (15.3)	18 (14.6)	37 (15.0)	
Single	24 (19.4)	28 (22.8)	52 (21.1)	
Education, n (%)				.81
High school or less	15 (12.1)	13 (10.6)	28 (11.3)	
Some college or vocational program	25 (20.2)	31 (25.2)	56 (22.7)	
Bachelor's degree	36 (29.0)	33 (26.8)	69 (27.9)	
Postgraduate degree	48 (38.7)	46 (37.4)	94 (38.1)	
Race, n (%)				.64
White	87 (70.2)	94 (76.4)	181 (73.3)	
Black/African American	25 (20.2)	22 (17.9)	47 (19.0)	
Asian/Indian	7 (5.6)	5 (4.1)	12 (4.9)	
Mixed/biracial	3 (2.4)	2 (1.6)	5 (2.0)	
Other	2 (1.6)	0 (0.0)	2 (1.0)	
Hispanic ethnicity, n (%)	3 (2.4)	7 (5.7)	10 (4.0)	.20
WIC[d] recipient[e], n (%)	32 (25.8)	28 (22.8)	60 (24.3)	.58
Employed, n (%)	104 (83.9)	108 (87.8)	212 (85.8)	.38
Smoke cigarettes (current and/or month before pregnancy), n (%)	14 (11.3)	16 (13.0)	30 (12.1)	.68
Prepregnancy BMI, mean (SD)	26.7 (7.1)	26.1 (5.5)	26.4 (6.3)	.43
Prepregnancy BMI ≥30 (obese), n (%)	24 (19.4)	24 (19.5)	48 (19.4)	.98
Intended duration of any breastfeeding, n (%)				.71
<6 months	7 (5.6)	5 (4.1)	12 (4.9)	
6-11 months	29 (23.4)	33 (26.8)	62 (25.1)	
12 months	49 (39.5)	48 (39.0)	97 (39.3)	
>12 months	18 (14.5)	12 (9.8)	30 (12.1)	
Unsure/as long as possible	21 (16.9)	25 (20.3)	46 (18.6)	
Intended duration of exclusive breastfeeding, n (%)				.07
<6 months	36 (29.0)	21 (17.1)	57 (23.1)	
6 months or longer	51 (41.1)	64 (52.0)	115 (46.6)	
Unsure/as long as possible	37 (29.8)	38 (30.9)	75 (30.4)	

[a]MILK is acronym for the intervention group treatment: a Mobile, semiautomated text message-–based Intervention to prevent perceived Low or insufficient milK supply.

[b]A total of three randomized participants did not complete the baseline demographic survey.

[c]*P* value for between-group differences were calculated with independent samples *t* tests for continuous type variables; for categorical variables, we used Pearson chi-square tests or, if sparse cells were encountered, Fisher exact tests.

[d]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

[e]Of 246 participants answering the item.

Uptake and Interactions with MILK (Usability)

In total, 38,284 MILK texts were sent to participants. Of those, 0.15% (58/38,284) texts sent to 38 participants were undelivered after multiple attempts. Nearly all MILK participants (122/126, 96.8%) engaged in at least one exchange with the system (eg, keyword text and response to query). Of 124 participants, 9 (7.3%) elected to stop MILK texts after receiving at least one message, and 3 (2.4%) participants elected to reduce messages to one series per week. In the control arm, 1 participant (1%)

declined to receive any Text4Baby messages after randomization, and 11 participants (9%) were confirmed to have stopped Text4Baby messages (Figure 1).

There were 46 HELP requests from 25.4% (32/126) MILK participants for individualized assistance from the study's IBCLC. Of these requests, 54% (25/46) occurred on a weekend, a holiday, or outside business hours (5 PM-7 AM). Most HELP requests specified SMS text messaging as the preferred medium to communicate with the study's IBCLC (27/46, 59%). The most commonly texted keywords for more detailed information occurred during the postpartum period and focused on two issues: (1) assessing adequate breast milk volume/infant intake and (2) optimizing sleep within the breastfeeding relationship (Table 2).

Table 2. Most commonly texted MILK[a] keywords.

Keyword	Timing of texts with a keyword prompt	Topical area	Number of participants texting, n
Sleep	Postpartum, weeks 6 and 7	Week 6: Maximizing sleep length/quality for the breastfed infant; Week 7: Tips for consolidating sleep for the breastfeeding mother	68; 54
Schedule	Postpartum, week 6	Norms for sleeping/feeding pattern emergence in infants; tips for encouraging the formation of desired sleeping/feeding habits	65
Norms	Postpartum, week 4	Average milk intake volume and weight gain for a 3- to 4-week-old infant	59
Reasons	Postpartum, week 5	Potential rationale for perceived decrease in milk volume between postpartum weeks 5 and 6	54
Sign	Postpartum, week 2	Indicators of satiety among breastfed infants	54

[a]MILK: Mobile, semiautomated text message-–based Intervention to prevent perceived Low or insufficient milK supply.

Acceptability

At 8 weeks postpartum, 84% (82/98) of MILK participants completing a survey reported that MILK messages were "helpful" or "very helpful" in achieving their breastfeeding goals or helping them to breastfeed; 21% (18/88) in the control group reported the same for Text4Baby messages (between-group difference: $\chi^2_{1,186}=74.5$; $P<.001$). According to the same survey, the most preferred MILK content was information about potential breastfeeding problems and solutions, while the least preferred was links to breastfeeding support resources (Table 3). The primary MILK dislike or criticism pertained to technical problems, though less than 10% (9/90) endorsed experiencing such issues (Table 4).

Table 3. Most preferred MILK[a] text message content (n=101 MILK participants completing an 8-week survey). Participants could select more than one preferred message content type. No participants selected "other" or "did not like any messages.

Content type	Participants endorsing, n (%)
Encouragement to begin or continue breastfeeding	57 (56.4)
Information on how to prevent and manage breastfeeding problems	74 (73.3)
Information on breastfeeding recommendations	63 (62.4)
Information on breastfeeding benefits	42 (41.6)
Links to breastfeeding articles and websites	58 (57.4)
Links to videos featuring real parents breastfeeding	22 (21.8)
Links to connect with breastfeeding support groups or persons	14 (13.9)

[a]MILK: Mobile, semiautomated-automated text message–based Intervention to prevent perceived Low or insufficient milK supply.

Table 4. MILK[a] text message dislikes and criticisms (n=101 MILK participants completing an 8-week survey). Participants could select more than one issue. No participants selected "content was hard to read/understand" [not represented in the table].

Dislike or criticism	Participants endorsing, n (%)
No dislikes	60 (59.4)
Too many texts	8 (7.9)
Too few texts	2 (2.0)
Messages too lengthy	1 (1.0)
Sent at an inconvenient time	7 (6.9)
Not helpful or applicable	4 (4.0)
Content offensive	3 (3.0)
Content poorly timed	5 (5.0)
Technical problems	9 (8.9)

[a]MILK: Mobile, semiautomated-automated text message–based Intervention to prevent perceived Low or insufficient milK supply.

Qualitative Findings

A total of 35 MILK participants (28%) provided qualitative feedback on the intervention; 34 interviews were conducted at weeks 2 to 10 postpartum, and one was conducted at 6 months postpartum. We identified three major themes pertaining to participants' experiences with MILK: (1) ascribed value rooted in the perceived impact on breastfeeding experience and trajectory; (2) preferred content prioritized practicalities, realities, and complexity in the breastfeeding relationship; and (3) appreciation for design features offering personalization and control juxtaposed with critiques about technical issues.

Theme 1: Perceived Impact on Breastfeeding Experience and Trajectory

Overwhelmingly, participants gave positive reviews of the MILK intervention, stating that it increased their breastfeeding knowledge and/or confidence, impacted their decision to begin or continue breastfeeding, and strengthened their breastfeeding support networks. In terms of knowledge/confidence, participants found MILK informative and reassuring in that the breastfeeding issues they were experiencing were surmountable and not uncommon:

> *I'm glad I signed up for the study, 'cause I got like so much knowledge and was able to persevere through the tough times.*

> *It gave me hope that other people had similar issues and made it through.*

Participants stated that texts were often comprehensive enough that they did not need to consult with their pediatrician or use other resources (eg, books and the internet) for breastfeeding questions, though in-person counseling was still considered necessary to resolve complex breastfeeding issues. Participants also preferred MILK to other breastfeeding support resources for convenience and trustworthiness:

> *When I would get websites from you guys, I would know that this was a valuable resource, that it wasn't just a random forum with somebody saying things, so I felt like it helped point you in the direction of the better information.*

> *If I had certain kinds of [breastfeeding] questions, I'll ask his doctor, but most of the time, the texts were just as helpful, so I didn't have to bother them.*

Participants noted that MILK texts reinforced their decision to breastfeed and provided encouragement to continue breastfeeding. In particular, texts were perceived as well-timed *(spot on; hit the nail right on the head*) to anticipate and address critical issues and misperceptions that might have otherwise led to formula supplementation or breastfeeding discontinuation:

> *I had almost given up breastfeeding even earlier, like after a week, and then one of the texts I got from the study was like, 'This is the hardest part. This is a common growth spurt time, and it'll get easier,' and I kept going because of that.*

A total of 1 participant noted that learning about the health impacts of breastfeeding through MILK provided a useful counternarrative to marketing tactics used by formula companies:

> *...they sent me Similac, some samples of it. And there are times when breastfeeding gets hard or it's completely exhausting, and I thought, "Oh my gosh, it would be so easy just to give her formula right now." But just knowing [from MILK] how beneficial it is for her to have purely breast milk was, I guess, was reinforcing of my decision* [to exclusively breastfeed]

Participants reported that particularly relevant or interesting texts were shared with partners or friends and "sparked" breastfeeding conversations. Texts were also thought to provide an advantage in breastfeeding discussions with pediatricians or IBCLCs:

> *They give me good terminology and language of how to talk about things...I was able to describe what was happening better because of some of the text messages.*

Theme 2: Preferences for Content Encapsulating Breastfeeding Practicalities, Realities, and Complexity

Participants favored content sent in the postpartum period that provided anticipatory guidance about breastfeeding expectations at each stage of infant development (eg, first hour and first week), as well as when breastfeeding issues were most likely to surface, their causes, and how to avoid or resolve them. Participants particularly liked messaging that addressed breastfeeding frequency, how growth spurts impact breastfeeding, and indicators that the infant was getting enough breast milk. They reported often accessing video and weblinks from texts for more detailed information and "bookmarking" certain sites to return to later. Participants also found motivational messaging encouraging, particularly texts geared toward the achievement of incremental breastfeeding milestones:

> *And so when you're looking at how long—"Oh my God, how long am I going to have to [breastfeed]," you can at least see those milestones, and feel like you're accomplishing something when you hit, like, "Ok, like every single week that I do this, I'm helping my baby."*

In the antenatal period, participants noted that content addressing practicalities in preparation for breastfeeding was most helpful (eg, breast pumps and nursing clothing and bras). With several exceptions, antenatal messaging around the benefits of breastfeeding was deemed interesting but "too simplistic." Participants felt that these texts could be replaced with information on how to practically integrate breastfeeding into the "real world." In particular, they desired more guidance on breastfeeding in public spaces, pumping breast milk (eg, flange fit and recommended accessories), weaning and introduction of solid foods, formula supplementation, breastfeeding after primary teeth eruption, and resolving nipple soreness and plugged ducts:

> *Just tell me...like how to make [breastfeeding] work in my life, when I'm going to the mall.*

Finally, some participants felt that several MILK messages "vilified" formula and that there was a general lack of messaging tailored for combination-feeding mothers. Conversely, some mothers thought texts were equally supportive of breastfeeding and formula feeding:

> *I felt like at the beginning, it was all about how breast milk was so much better than formula, and I definitely appreciate that breast milk has a lot of benefits, but at the same time, there's, like, all sorts of reasons why people have to go on formula, you know?*
>
> *You sent a website that said, about how moms like maybe only breastfeed at night and that still works, which I didn't know that was a possibility. So it did help me to learn that there's were like other options for me...and it kind of took away some of that mom guilt of "oh no, I'm giving my kid formula."*

Additional suggestions to improve MILK included tailoring content to address existing breastfeeding knowledge and more "complex" breastfeeding trajectories (eg, delayed lactogenesis and exclusively pumping), compiling all messages on a companion website for quick back-reference, adapting content for partners, and extending messaging beyond 8 weeks in recognition of continued breastfeeding barriers, such as the mother's return to employment.

Theme 3: Appreciation for Design Features Offering Personalization and Control Juxtaposed With Critiques About Technical Issues

Participants liked receiving SMS text messaging–based breastfeeding support for various reasons, including convenience, autonomy in deciding when to read/access information, ability to refer back to texts, and "chunking" into a "consumable set of information":

> *The information will be forever with me...I saved all my messages so I can go back and look at them...I would read them twice, maybe three or four times, just depending on how busy my day was.*

Participants also responded positively to MILK's automated personalization and interactivity features, including how they were addressed ("I loved how they used my baby's name...it felt personal, even though I know that they're automated text messages") and the option to access more detailed information via keywords and links within messages. This interactivity seemed to contribute to some degree of anthropomorphism of the system, such that some participants began responding conversationally to automated messages. Participants also liked the ability to receive on-demand individual help from a study's IBCLC. Those who solicited IBCLC assistance remarked that the consultant explained complex issues well, referred them to reputable resources, and provided "nonjudgmental," personalized advice. Among those who did not use the IBCLC's services, some had access to other lactation help, whereas others simply forgot and suggested weekly reminders.

Participants reported technical glitches in the system that somewhat diminished their enthusiasm for the program, including repeated audio alerts from multiple-part messages, broken weblinks and video links, and no response when texting a keyword (response windows "timed out" within 24 hours). To address these problems, multiple-part messages were consolidated where possible and weblinks were updated upon notification of an issue and during regular checks by the study staff. Similarly, we programmed an automated email alert for unsolicited texts. These alerts were reviewed daily to check for timed-out keyword texts, in which case the response was sent manually.

Opinions differed on delivery timing of automated messages. Although participants appreciated the predictability of a standard daily message time (10 AM) and felt that the message frequency was acceptable (no mention of preference for a particular day of the week), some reported that messages were disruptive during work or sleep. Some suggested early evening as a more convenient message delivery option, but there was no consensus.

Discussion

We found that MILK was both a feasible and an acceptable breastfeeding support intervention among first-time mothers

with a strong prenatal intention to breastfeed. During the trial period, we experienced rapid recruitment, a low rate of direct declines, and few withdrawals or elections to reduce message frequency. There was also a high rate of message receipt and participant interaction with the system. Participants reported that MILK increased their breastfeeding confidence, solidified their decision to initiate breastfeeding, and potentially helped them to breastfeed longer than they might have otherwise.

Other automated SMS breastfeeding support programs have been developed and trialed in China [22], Australia [21], Kenya [20], mainland United States [30], Hawaii, and Puerto Rico [31]. These systems all delivered approximately one automated text per week to breastfeeding mothers but varied in terms of initiation (pregnancy vs postpartum), duration (8 weeks to 12 months postpartum), personalization (eg, tailoring to key participant characteristics, such as breastfeeding status), access to an IBCLC, and rigor employed in content development. All but one program [31] reported positive results in terms of participant satisfaction and/or exclusive breastfeeding through 6 months postpartum. The MILK system differs from these existing programs in terms of its target population of first-time mothers, primary focus on a single breastfeeding issue (PIM), advanced automated functionality (eg, keyword texting and branched response logic), and grounding in both theory and EMA data [28]. The latter informed the delivery of near-daily texts tailored to breastfeeding issues experienced at the intersection of a particular gestational and chronological age. Although the impact of MILK on objectively measured breastfeeding outcomes are pending, the findings presented here illustrate a high degree of participant satisfaction with the overall program and its features. On the basis of suggested modifications to MILK, future SMS text messaging and technologically based breastfeeding support programs should consider messaging customized to one's existing breastfeeding knowledge, current breastfeeding concerns, and unique breastfeeding trajectory. Such tailoring may be achieved via the incorporation of more advanced branching algorithms or possibly through the application of machine learning principles, though the user interface of any system should remain simple and intuitive [32].

A somewhat unexpected study finding was the relatively low utilization of MILK's HELP feature providing consultation with an IBCLC. Although interview data indicate that participants may have simply forgotten about it, it is also possible that MILK prevented some common breastfeeding issues for which parents typically seek assistance. Another possibility, supported by the fact that more than half of the HELP requests came during off-hours, is that participants found community lactation resources more familiar or convenient. This is in line with the technology acceptance model, which states that for a technology to be adopted, the technology must fulfill both a need (perceived usefulness) and be more convenient than other options (perceived ease-of-use) [33]. Thus, future technologic breastfeeding innovations should consider what gap they fill relative to existing resources.

Our previous research indicated that among 146 surveyed postpartum mothers, the most commonly desired technology-based breastfeeding support involved encouragement or "cheerleading" [34]. However, mothers' preferences for motivational breastfeeding support messages in this study was somewhat unanticipated, given that the automation of such content could be construed as generic or impersonal. Potentially, some of the personalization features of MILK, including addressing mothers and infants by name and referencing the infant's age, contributed to a sense of connection to the system and openness to emotive appeals. This is supported by our observations that some participants began to actively converse with the system, even when a response was not indicated. Preferences for messages of encouragement also suggest that pregnant and postpartum individuals are not necessarily receiving this type of breastfeeding (or emotional) support within their existing networks [35]. Most participants also preferred practical how-to breastfeed messages rather than those addressing breastfeeding benefits and recommendations. It is unclear whether preferences would be similar among more diverse groups or among those with more ambivalent views toward breastfeeding.

The extrapolation of findings from MILK is limited by the sample demographics, consisting of predominantly white college-educated women without other biological children from a single geographic area and with a strong commitment to breastfeeding. In addition, MILK was designed for English speakers, though PIM is a pervasive lactation challenge across cultures [36-39]. Future research and programmatic development should consider SMS text messaging breastfeeding support for more diverse groups and for other common lactation challenges in addition to PIM.

Acknowledgments

This study was funded by the National Institutes of Health, National Institute of Nursing Research, grant R00 NR015106, "Preventing Perceived Insufficient Milk: Development of a Text Message-Based Intervention," PI: Demirci. For their input into text message content development and delivery, the authors would like to thank Danielle Rigg and Bettina Forbes of the Best for Babes Foundation, Diane Spatz (University of Pennsylvania School of Nursing, The Children's Hospital of Philadelphia), Elizabeth Felter (University of Pittsburgh Graduate School of Public Health), and numerous other colleagues at the University of Pittsburgh who provided feedback in the context of presentations and workshops. The authors would also like to thank Meg Beard (Santa Barbara County Public Health Department Nutrition Services) for sharing the *Bfed* SMS breastfeeding support program as an exemplar.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 532 KB - mhealth_v8i4e17328_app1.pdf]

References

1. Centers for Disease Control and Prevention. Breastfeeding Report Card: United States, 2018 URL: https://www.cdc.gov/breastfeeding/data/reportcard.htm [accessed 2019-07-18]
2. Bartick MC, Schwarz EB, Green BD, Jegier BJ, Reinhold AG, Colaizy TT, et al. Suboptimal breastfeeding in the United States: maternal and pediatric health outcomes and costs. Matern Child Nutr 2017 Jan;13(1). [doi: 10.1111/mcn.12366] [Medline: 27647492]
3. Victora CG, Bahl R, Barros AJ, França GV, Horton S, Krasevec J, Lancet Breastfeeding Series Group. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 2016 Jan 30;387(10017):475-490. [doi: 10.1016/S0140-6736(15)01024-7] [Medline: 26869575]
4. Section on Breastfeeding. Breastfeeding and the use of human milk. Pediatrics 2012 Mar;129(3):e827-e841. [doi: 10.1542/peds.2011-3552] [Medline: 22371471]
5. Odom EC, Li R, Scanlon KS, Perrine CG, Grummer-Strawn L. Reasons for earlier than desired cessation of breastfeeding. Pediatrics 2013 Mar;131(3):e726-e732 [FREE Full text] [doi: 10.1542/peds.2012-1295] [Medline: 23420922]
6. Perrine CG, Scanlon KS, Li R, Odom E, Grummer-Strawn LM. Baby-Friendly hospital practices and meeting exclusive breastfeeding intention. Pediatrics 2012 Jul;130(1):54-60 [FREE Full text] [doi: 10.1542/peds.2011-3633] [Medline: 22665406]
7. Brown CR, Dodds L, Legge A, Bryanton J, Semenic S. Factors influencing the reasons why mothers stop breastfeeding. Can J Public Health 2014 May 9;105(3):e179-e185 [FREE Full text] [doi: 10.17269/cjph.105.4244] [Medline: 25165836]
8. Chantry CJ, Dewey KG, Peerson JM, Wagner EA, Nommsen-Rivers LA. In-hospital formula use increases early breastfeeding cessation among first-time mothers intending to exclusively breastfeed. J Pediatr 2014 Jun;164(6):1339-45.e5 [FREE Full text] [doi: 10.1016/j.jpeds.2013.12.035] [Medline: 24529621]
9. Gatti L. Maternal perceptions of insufficient milk supply in breastfeeding. J Nurs Scholarsh 2008;40(4):355-363 [FREE Full text] [doi: 10.1111/j.1547-5069.2008.00234.x] [Medline: 19094151]
10. Stuebe AM. Does breastfeeding prevent the metabolic syndrome, or does the metabolic syndrome prevent breastfeeding? Semin Perinatol 2015 Jun;39(4):290-295 [FREE Full text] [doi: 10.1053/j.semperi.2015.05.008] [Medline: 26187772]
11. Riddle SW, Nommsen-Rivers LA. Low milk supply and the pediatrician. Curr Opin Pediatr 2017 Apr;29(2):249-256. [doi: 10.1097/MOP.0000000000000468] [Medline: 28134707]
12. Galipeau R, Dumas L, Lepage M. Perception of not having enough milk and actual milk production of first-time breastfeeding mothers: is there a difference? Breastfeed Med 2017 May;12:210-217. [doi: 10.1089/bfm.2016.0183] [Medline: 28326807]
13. Wambach K, Watson GC. Anatomy and physiology of lactation. In: Wambach K, Riordan J, editors. Breastfeeding and Human Lactation. Fifth Edition. Burlington: Jones and Bartlett Learning; 2016:79-120.
14. Dykes F, Williams C. Falling by the wayside: a phenomenological exploration of perceived breast-milk inadequacy in lactating women. Midwifery 1999 Dec;15(4):232-246. [doi: 10.1054/midw.1999.0185] [Medline: 11216257]
15. Chantry CJ. Supporting the 75%: overcoming barriers after breastfeeding initiation. Breastfeed Med 2011 Oct;6:337-339 [FREE Full text] [doi: 10.1089/bfm.2011.0089] [Medline: 22007824]
16. Wood NK, Sanders EA, Lewis FM, Woods NF, Blackburn ST. Pilot test of a home-based program to prevent perceived insufficient milk. Women Birth 2017 Dec;30(6):472-480. [doi: 10.1016/j.wombi.2017.04.006] [Medline: 28529088]
17. Abaza H, Marschollek M. mHealth application areas and technology combinations. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 8;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
18. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
19. Heinig MJ. Breastfeeding promotion for generations X and Y: why the old ways won't work. J Hum Lact 2009 Aug;25(3):263-265. [doi: 10.1177/0890334409341450] [Medline: 19666853]
20. Unger JA, Ronen K, Perrier T, DeRenzi B, Slyker J, Drake A, et al. Short message service communication improves exclusive breastfeeding and early postpartum contraception in a low- to middle-income country setting: a randomised trial. Br J Obstet Gynaecol 2018 Nov;125(12):1620-1629 [FREE Full text] [doi: 10.1111/1471-0528.15337] [Medline: 29924912]
21. Gallegos D, Russell-Bennett R, Previte J, Parkinson J. Can a text message a week improve breastfeeding? BMC Pregnancy Childbirth 2014 Nov 6;14:374 [FREE Full text] [doi: 10.1186/s12884-014-0374-2] [Medline: 25369808]

22. Jiang H, Li M, Wen LM, Hu Q, Yang D, He G, et al. Effect of short message service on infant feeding practice: findings from a community-based study in Shanghai, China. JAMA Pediatr 2014 May;168(5):471-478. [doi: 10.1001/jamapediatrics.2014.58] [Medline: 24639004]
23. Bushar JA, Kendrick JS, Ding H, Black CL, Greby SM. Text4baby influenza messaging and influenza vaccination among pregnant women. Am J Prev Med 2017 Dec;53(6):845-853 [FREE Full text] [doi: 10.1016/j.amepre.2017.06.021] [Medline: 28867143]
24. Evans WD, Wallace Bihm J, Szekely D, Nielsen P, Murray E, Abroms L, et al. Initial outcomes from a 4-week follow-up study of the Text4baby program in the military women's population: randomized controlled trial. J Med Internet Res 2014 May 20;16(5):e131 [FREE Full text] [doi: 10.2196/jmir.3297] [Medline: 24846909]
25. Evans WD, Wallace JL, Snider J. Pilot evaluation of the Text4baby mobile health program. BMC Public Health 2012 Nov 26;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
26. Dennis C. Breastfeeding initiation and duration: a 1990-2000 literature review. J Obstet Gynecol Neonatal Nurs 2002;31(1):12-32. [doi: 10.1111/j.1552-6909.2002.tb00019.x] [Medline: 11843016]
27. Dennis CL. Theoretical underpinnings of breastfeeding confidence: a self-efficacy framework. J Hum Lact 1999 Sep;15(3):195-201. [doi: 10.1177/089033449901500303] [Medline: 10578797]
28. Demirci JR, Bogen DL. An ecological momentary assessment of primiparous women's breastfeeding behavior and problems from birth to 8 weeks. J Hum Lact 2017 May;33(2):285-295 [FREE Full text] [doi: 10.1177/0890334417695206] [Medline: 28418803]
29. Crabtree BF, Miller WL. Clinical research: a multimethod typology and qualitative roadmap. In: Miller WL, Crabtree BF, editors. Doing Qualitative Research. Second Edition. Thousand Oaks: Sage Publications; 1999:3-30.
30. Beard M. 'Bfed' texting program and 'Breastfeeding: A Smart Choice' class: using cell phones to reach Gen Y mothers. Clin Lact 2014;5(4):123-127. [doi: 10.1891/2158-0782.5.4.123]
31. Palacios C, Campos M, Gibby C, Meléndez M, Lee JE, Banna J. Effect of a multi-site trial using short message service (SMS) on infant feeding practices and weight gain in low-income minorities. J Am Coll Nutr 2018;37(7):605-613 [FREE Full text] [doi: 10.1080/07315724.2018.1454353] [Medline: 29708471]
32. Tang K, Gerling K, Chen W, Geurts L. Information and communication systems to tackle barriers to breastfeeding: systematic search and review. J Med Internet Res 2019 Sep 27;21(9):e13947 [FREE Full text] [doi: 10.2196/13947] [Medline: 31573903]
33. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
34. Demirci JR, Cohen SM, Parker M, Holmes A, Bogen DL. Access, use, and preferences for technology-based perinatal and breastfeeding support among childbearing women. J Perinat Educ 2016;25(1):29-36 [FREE Full text] [doi: 10.1891/1058-1243.25.1.29] [Medline: 26848248]
35. Lee K, Vasileiou K, Barnett J. 'Lonely within the mother': an exploratory study of first-time mothers' experiences of loneliness. J Health Psychol 2019 Sep;24(10):1334-1344. [doi: 10.1177/1359105317723451] [Medline: 28795604]
36. Bartick M, Reyes C. Las dos cosas: an analysis of attitudes of latina women on non-exclusive breastfeeding. Breastfeed Med 2012 Feb;7(1):19-24. [doi: 10.1089/bfm.2011.0039] [Medline: 22007765]
37. Gianni ML, Bettinelli ME, Manfra P, Sorrentino G, Bezze E, Plevani L, et al. Breastfeeding difficulties and risk for early breastfeeding cessation. Nutrients 2019 Sep 20;11(10):pii: E2266 [FREE Full text] [doi: 10.3390/nu11102266] [Medline: 31547061]
38. Li T, Guo N, Jiang H, Eldadah M. Breastfeeding self-efficacy among parturient women in Shanghai: a cross-sectional study. J Hum Lact 2019 Aug;35(3):583-591. [doi: 10.1177/0890334418812044] [Medline: 30517822]
39. Yilmaz F, Küçükoğlu S, Özdemir AA, Oğul T, Aşki N. The effect of kangaroo mother care, provided in the early postpartum period, on the breastfeeding self-efficacy level of mothers and the perceived insufficient milk supply. J Perinat Neonatal Nurs 2020;34(1):80-87. [doi: 10.1097/JPN.0000000000000434] [Medline: 31895303]

Abbreviations

EMA: ecological momentary assessment
IBCLC: international board-certified lactation consultant
MILK: Mobile, semiautomated text message–based Intervention to prevent perceived Low or insufficient milK supply
PIM: perceived insufficient milk

Edited by G Eysenbach; submitted 06.12.19; peer-reviewed by L Kair, D Leightley; comments to author 05.01.20; revised version received 24.01.20; accepted 26.01.20; published 29.04.20.

Please cite as:
Demirci JR, Suffoletto B, Doman J, Glasser M, Chang JC, Sereika SM, Bogen DL
The Development and Evaluation of a Text Message Program to Prevent Perceived Insufficient Milk Among First-Time Mothers: Retrospective Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e17328
URL: https://mhealth.jmir.org/2020/4/e17328
doi:10.2196/17328
PMID:32347815

Original Paper

A Mobile Health Intervention for Fetal Alcohol Spectrum Disorders (Families Moving Forward Connect): Development and Qualitative Evaluation of Design and Functionalities

Christie LM Petrenko[1], PhD; Jennifer Parr[1], MS; Carson Kautz[1], BA; Cristiano Tapparello[2], PhD; Heather Carmichael Olson[3,4], PhD

[1]Mt Hope Family Center, University of Rochester, Rochester, NY, United States

[2]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, United States

[3]Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, Seattle, NY, United States

[4]Center for Child Health, Behavior, and Development, Seattle Children's Research Institute, Seattle, WA, United States

Corresponding Author:
Christie LM Petrenko, PhD
Mt Hope Family Center
University of Rochester
187 Edinburgh St
Rochester, NY, 14608
United States
Phone: 1 585 275 2991
Email: christie.petrenko@rochester.edu

Abstract

Background: Fetal alcohol spectrum disorders (FASD) affect approximately 2% to 5% of the US population. However, most families are unable to access FASD-informed interventions. Barriers to care include the lack of a knowledgeable and skilled workforce and family-level barriers such as limited financial resources, inability to access childcare, and stigma. As a result, families often try peer-to-peer and self-help support strategies. However, they often take these strategies from disparate sources, which have quite variable intervention quality and empirical support.

Objective: This study aimed to initiate systematic development and evaluation of a mobile health intervention *(app)* for caregivers raising children with FASD. Focus groups were conducted to elicit participant perspectives on app design and functionalities to inform further app development.

Methods: The app, called *FMF Connect*, was derived from the scientifically validated *Families Moving Forward (FMF) Program*, a clinician-delivered behavioral consultation intervention. *FMF Connect* was intended for caregiver self-delivery and included five main components: (1) *Learning Modules*, (2) *Family Forum*, (3) *Library*, (4) *Notebook*, and (5) *Dashboard*. Focus group methods were used to solicit perspectives from diverse families during the early stages of app development. Questions were asked about interface design, relevance of components and content, and perceived barriers and facilitators of use. A total of 25 caregivers participated in 7 focus groups across 5 US cities. Data were analyzed thematically.

Results: Focus group participants were generally enthusiastic about the app interface design and components. Four global positive impression themes emerged, including (1) ease of access, (2) how the app guides and organizes information, (3) connection to other users and information, and (4) ability to share some content with others. Themes arose not only in discussions relating to positive app features but also when participants were asked about motivators for app use. Participants related how these positive global themes could address some system-level barriers, such as limited access to services, feeling isolated, and increased advocacy needs related to the societal lack of FASD knowledge. Participants identified many positive features about individual app components and functionalities. They also communicated potential barriers to use and raised important concerns and considerations relating to several app components. These included recognizability of the app based on the logo, and the balance of following the planned intervention sequence versus obtaining immediate answers. Also mentioned were privacy and dynamics within the *Family Forum*.

Conclusions: *FMF Connect* is a promising novel intervention with potential to reach many families in need and reduce significant barriers to care, resulting in a broader public health impact. Study findings will guide further app development both in terms of

content and technological advances to optimize intervention effects. *FMF Connect* app development provides useful directions for other apps aimed at changing parenting practices.

(JMIR Mhealth Uhealth 2020;8(4):e14721) doi:10.2196/14721

KEYWORDS

fetal alcohol spectrum disorders; fetal alcohol syndrome; parenting; children; mobile health; mHealth; treatment

Introduction

Background

Fetal alcohol spectrum disorders (FASD) affect an estimated 2% to 5% of the US general population [1] and occur at even higher rates in special populations, such as those served by child welfare and juvenile justice or in psychiatric care [2]. FASD are diagnosed when there is prenatal alcohol exposure and evidence of neurobehavioral impairment [3-5]. Additional symptoms may also be present. These include deficient brain growth or seizures, a pattern of characteristic facial features, and growth delays. Outcomes are variable and can be impacted by other factors such as maternal and fetal genetics, nutrition, immune functioning, and pre- and postnatal stress [6-8].

Unfortunately, research suggests that a majority of people affected by this condition are undiagnosed or misdiagnosed [9]. Children with FASD have high rates of health care and other service utilization [10,11], which should afford opportunities for detection and delivery of FASD-informed services and supports. However, significant system-level barriers exist that interfere with appropriate detection and service delivery [12,13]. One of the leading causes for system-level barriers is inadequate training in FASD provided to educators and health professionals. Surveys of trainees and professionals document that although many have heard about FASD, most do not feel competent to diagnose or effectively treat this condition [14-16]. As a result, there are very few providers in the community who regularly diagnose this condition or provide FASD-informed care. This problem is especially the case in rural or underserved communities.

Although an imperfect metric, an informal survey of resource directories from the National Organization on Fetal Alcohol Syndrome and their state affiliates show a median of 3 (range 0-13) diagnostic providers and 2 (range 0-86) mental health intervention services per state [17]. Clearly, existing resources are insufficient to meet the needs of this prevalent and often complex condition. It is also relevant to consider the demographics of families seeking FASD diagnostic and intervention services. Although services try to meet the needs of all family types, most families that receive FASD diagnostic services or participate in intervention research trials are foster parents, adoptive parents or relative caregivers (70%-98%) [18,19].

Families also experience many family-level barriers and stressors that adversely impact their ability to access FASD-informed care or even any care at all [13,20,21]. These include practical challenges such as finding appropriate childcare, financial strains, getting time off work, scheduling constraints, and juggling multiple demands and time-intensive services. Emotional stressors, such as feeling overwhelmed, isolated, and stigmatized, can impact the ability to access care. Parents raising children with FASD also report feeling discouraged by previous unsuccessful treatment outcomes, which can further reduce motivation to seek services.

Potential of Mobile Health Interventions for Fetal Alcohol Spectrum Disorders

Since the 1980s, FASD parent support and advocacy groups have seen robust international growth in response to families' urgent needs for help and support [22]. Families' natural social support networks have been found helpful in the broader developmental disabilities literature [22,23], and caregivers of children with FASD often try peer-to-peer and self-help support strategies. With advances in technology and improvements in internet access, parents are more often obtaining information and social support from Web-based sources [24]. Social media has surpassed previously popular listserv and email formats for resource sharing and support [23]. Unfortunately, studies document multiple problems with the quality, consistency, and readability of Web-based information for parents of children with developmental disabilities [24].

Mobile health (mHealth) interventions are likely well suited to delivery of peer-to-peer and self-help support strategies for families raising children with FASD. Most adults have smartphones and regularly use apps in their daily lives [25,26]. mHealth interventions can potentially address many of the family- and system-level barriers currently interfering with access to FASD-informed care. One notable advantage of mHealth interventions is the ability to scale up for a large number of users. Developing mHealth interventions for FASD could also augment current programmatic efforts toward more traditional intervention dissemination and provider training, while capitalizing on the clear interest in and attempted use of self-help by caregivers raising children with FASD.

There are currently over 318,000 health-related apps available in app stores [25]. Remarkably, so far very few have been subjected to empirical study or are derived from evidence-based principles of behavior change [27-29]. Fortunately, evaluations of existing apps are becoming more common and clinical trials on mHealth interventions are increasing in number and quality [25,28]. Most apps evaluated to date have utilized health behavior theory constructs, such as self-monitoring and goal setting, and have received high user acceptability ratings [29]. Although sample sizes have been modest in intervention trials, preliminary evidence supports the potential for mHealth apps to accomplish behavior change, and ultimately improve outcomes across varied conditions [29,30].

Evidence Base for Digital Parenting Interventions

Relatively few studies have explored the efficacy of mHealth interventions targeting parenting for preschool and school-aged children. However, a number of reviews and meta-analyses have summarized the growing evidence base for self-directed digital parenting interventions, which largely represent Web-based interventions [31-33]. The vast majority of self-directed interventions reviewed have been adapted from existing empirically validated interventions traditionally delivered by a clinician. In controlled trials, data suggest digital parenting interventions have similar or better retention (mean 84.8%) and adherence (mean 73.7% content completed) than in-person interventions [33]. Broadly, the evidence for digital parenting interventions is more consistent for interventions targeting externalizing behaviors than internalizing behaviors, given the very small number of such interventions focused on internalizing problems [33]. A meta-analysis of seven digital self-directed parenting interventions for children with externalizing behavior found overall small-medium effect sizes for child behavior (d=0.44), parent behavior (d=0.41), and parent confidence (d=0.36) [31]. Effects were larger for samples with clinically elevated behavior problems (d=0.61) than for nonclinical samples (d=0.21), and when interventions were interactive (d=0.82) versus noninteractive (d=0.36) [31].

Although smartphone technology is promising, few digital parenting interventions have capitalized on this method of delivery. Most interventions published to date have been developed for internet websites optimized for desktop or laptop computers. Yet, apps for smartphones and tablet computers have the advantage of greater ease of access for many parents during the course of a day. Importantly, integration of peer-to-peer support into mHealth and other digital parenting interventions could augment the power of these treatments. Integration could efficiently capitalize on benefits identified in previous research on social support. However, to date, the inclusion of peer-to-peer support within digital platforms has been surprisingly limited.

Families Moving Forward Connect: A Novel Mobile Health Intervention for Caregivers Raising Children With Fetal Alcohol Spectrum Disorders

This study presents data from an initial evaluation of the design and planned functionalities of a novel mHealth intervention for caregivers raising children with FASD, called *Families Moving Forward (FMF) Connect*. *FMF Connect* is based on the caregiver-focused *FMF Program* developed by Olson and her research team at the Seattle Children's Research Institute (SCRI)/University of Washington [19,22,34,35]. See Table 1 for the comparison of FMF and *FMF Connect*.

Table 1. Comparison of features of the standard Families Moving Forward (FMF) Program and FMF Connect mobile health intervention.

Features	Standard *FMF Program*	*FMF Connect*
Format	In-person; originally tested in families' homes (but can also be delivered in-clinic or through telehealth)	Mobile health app
Target	Parents and caregivers of children (aged 3-12) with FASD or prenatal alcohol exposure	Parents and caregivers of children (aged 3-12) with FASD or prenatal alcohol exposure
Materials and delivery	Materials provided by specially trained mental health or child development provider	Materials are self-directed by the caregiver
Content division	14-17 sessions (includes core + optional material)	12 *Learning Modules* + optional material
Duration	90-min sessions, every other week (can be 60-min weekly sessions)	Self-directed by the caregiver
Clinical techniques	*Caregiver-focused:* Integration of psychoeducation and support, positive behavior support, cognitive behavioral strategies, advocacy education, and motivational interviewing	*Caregiver-focused:* Integration of psychoeducation and support, positive behavior support, cognitive behavioral strategies, advocacy education, and motivational interviewing
Key treatment processes	*Reframing, accommodations, brainstorming*	*Reframing, accommodations, brainstorming*
Key outcomes	Improve positive cognitive appraisal of child, improve parenting sense of competence, meet unmet family needs, and improve child adaptive function (and reduce problem behaviors)	Improve positive cognitive appraisal of child, improve parenting sense of competence, meet unmet family needs, and improve child adaptive function (and reduce problem behaviors)
Routine outcomes monitoring	Progress checklist completed at the start of each session to rate child behavior, self-care, and service barriers	Daily notifications to rate self-care and support, weekly notifications to rate child behavior
Social support	Support provided by a specialist, linkages to community or Web support groups	*Family Forum* for peer support integrated into app, moderator supported by training and consultation

Structure, Theoretical Framework, and Outcomes of the Families Moving Forward Program

The standard, therapist-led *FMF Program* was designed for parents and caregivers of children (aged 3-12 years) with FASD. The *FMF Program* is traditionally implemented in families' homes every other week for 14 to 17 sessions, although in practice, it can also be delivered in clinic settings and in other patterns of session frequency or duration. The standard *FMF Program* was designed to fit with the highly diverse demographics of families raising children with FASD, including

all family structures, and a wide range of socioeconomic status and caregiver racial and ethnic background.

The *FMF Program* is grounded in developmental systems theory, and is informed by research on parenting, developmental disabilities, family systems, and treatment of child behavior problems. It is designed to modify specific parenting attitudes and responses to children's problem behaviors via integration of psychoeducation and support, positive behavior support, cognitive behavioral strategies, and motivational interviewing [19,22,34,35]. By helping caregivers interpret their children's behavior from a neurodevelopmental perspective (called *reframing* in the standard *FMF Program*), it is theorized they will develop a more positive and realistic cognitive appraisal of the child, use more effective antecedent-based behavioral strategies to promote adaptive child functioning and decrease challenging behavior, and feel more efficacious in their role as a parent.

Among other findings, studies have documented generally medium to large intervention effects on caregiver knowledge, family needs met, parenting efficacy, reported improvement in self-care, and child behavior immediately posttreatment [19,36]. Effects on *reframing* and targeted parenting practices have been in the small to medium ranges [19,36]. Given how difficult are the lives of these children and families, improving the positive trajectory in any measurable way is a significant (and vital) aim.

Deriving the Families Moving Forward Connect Mobile Health Intervention

Consistent with the adaptation approach advocated by Card et al [37], the core components and underlying theory of the standard *FMF Program* were first identified. These components and theorized mechanisms of change were linked with technological features that correspond to how users interact with technology versus a literal adaptation to a new delivery mode [38]. This process was facilitated by the use of a backward design process [39], evaluation of behavior change techniques [40], and consideration of ethical principles [41]. See Figure 1 for an illustration of the *FMF Connect* components.

Figure 1. Illustration of the five main components of the Families Moving Forward (FMF) Connect mobile health intervention and their primary functionalities.

Psychoeducation, attitude change, and skill-building content from the standard *FMF Program* is distilled into brief learning modules. Standard *FMF Program* materials (short fact sheets, simple worksheets, and brief videos) lend themselves well to mHealth adaptation. Although much of the content is preserved, the flow of content delivery differs somewhat in *FMF Connect* to be more amenable to self-direction by caregivers.

Although much of the content translated well, potential mismatches between the original program and new delivery context were considered [37]. Several examples of this intensive process are given here. Difficult concepts introduced early in the standard *FMF Program* were simplified and moved later in the *FMF Connect* app flow because of the absence of clinician support. Since it is self-directed, *FMF Connect* was designed

to rely much more heavily than does the standard *FMF Program* on video examples filmed with real families.

Derived from the standard *FMF Program* techniques, *FMF Connect* also prompts caregiver ratings of key outcomes (eg, child behavior, caregiver self-efficacy, and self-care) displayed on the app dashboard. These ratings capitalize on the confirmed benefits of *routine outcomes monitoring* for enhancing treatment effectiveness and tailoring intervention content [42].

Given the benefits of parent social support [22], an important innovation of *FMF Connect* is to integrate a peer-moderated *Family Forum* to aid in engaging families, promote implementation of new knowledge and skills, and provide secure online and high-quality support. The use of trained peer moderators promotes sustainability and surmounts workforce barriers, building on and enhancing what has naturally evolved in the *real world* of FASD self-help.

Other functionalities are also included, some of which were inspired by the standard *FMF Program*, and some unique to or transformed by the app format. These include a *Library* for additional optional content, a *Notebook* to organize completed exercises and tools, and weekly emails to engage users and highlight app features and tips.

This Study

The relatively small, but quickly growing, evidence base for mHealth interventions is promising and offers direction and guidance. This study represents a crucial step in the systematic development of *FMF Connect* with a focus on app design and functionalities. Using rigorous qualitative methods, the app design and planned functionalities were presented to groups of caregivers of children with FASD in multiple cities across the United States. The integration of key stakeholder feedback early and iteratively throughout the development and evaluation process is aimed to facilitate the acceptability and utility of the intervention [43,44].

Methods

Study Design

The aim of this study was to elicit feedback about the design and components of the *FMF Connect* mHealth intervention from targeted users: caregivers of children with FASD. Qualitative methodology is well suited to the aims of this study. Focus group methods were specifically chosen to elicit in-depth discussion among caregivers on aspects such as interface design, ease of use, relevance of components and content, and barriers and facilitators of use.

Recruitment

Given that the app is designed for use by families across the United States, caregiver perspectives were elicited across various geographical regions. Caregivers were eligible for the study if they were over the age of 18 years and a primary caregiver of a child (aged 3-17 years) with an FASD. Diagnosis was based on caregiver report, although most participants were recruited through well-established diagnostic clinics. Although *FMF Connect* is designed for caregivers of children aged 3 to 12 years, caregivers of adolescents (aged 13-17 years) were also included. These caregivers have the advantage of being able to reflect on their experiences parenting their child across the full age range targeted by the app. They can also offer a broader perspective on the types of features and content that would be helpful. Caregivers who had previously completed the standard *FMF Program* were also included in this study. They could reflect on previous lived experience learning and applying the content of the *FMF Program*. In addition, they could offer important insights on what it might be like to learn this content in self-directed manner through *FMF Connect*.

Caregivers were recruited through multiple mechanisms. These included existing FASD research registries, provider referrals, targeted flyers in parent support groups or Web newsletters, and conferences. Several principal investigators within the Collaborative Initiative on FASD (CIFASD) also offered to help with recruitment and logistics for holding focus groups at their sites.

A total of 25 caregivers participated in 7 focus groups across 5 US cities from December 2017 to June 2018. Each focus group included 3 to 4 caregivers. Focus groups were held in Rochester, NY (3); Atlanta, GA (1); Minneapolis, MN (1); San Diego, CA (1); and Los Angeles, CA (1).

The University of Rochester Institutional Review Board reviewed and approved all study procedures. Participants provided written informed consent before enrollment in the study.

Procedures

All focus group interviews were conducted in a private meeting room at each site. In preparation, participants completed a brief demographic questionnaire, including some metrics of their comfort with technology and smartphone usage.

A consistent research team conducted all focus group interviews. The first author, who is a clinical psychologist and researcher with 15 years of experience in the field of FASD, was the lead moderator of all focus groups. She has 7 years of experience using qualitative research methods and multiple published studies with this population [12,21,45]. The fourth author is a faculty member in computer engineering. He led the demonstrations of design mock-ups and app prototypes in focus groups. The second author is a doctoral student in counseling and counselor education and took detailed observational notes during all groups. No personnel apart from participants and researchers were present during the interviews.

Each focus group session began with research team introductions, a statement about the purpose of the focus group, and discussion of ground rules and expectations. To reduce positive response bias, participants were explicitly encouraged to share any concerns or negative feedback during focus groups. The research team emphasized they would rather hear these concerns during development when changes could more easily be made than later once the app was widely disseminated.

At the start of discussion, participants were provided with a handout giving brief bullet point descriptions about each component of the *FMF Connect* app. The research team, then, showed participants mock-ups or prototypes of the app design

(see Multimedia Appendix 1). They, then, elicited participant discussion component by component, following a semistructured interview guide. After reviewing individual app components, participants were asked about general impressions of the app interface and perceived motivators and barriers to use. The interview guide included open-ended questions (eg, "What do you think about the *Family Forum*?" "What might you improve or do differently?" "What do you think about the look and feel of the app?" and "What would motivate you to use an app like this?") and additional probes to elicit more in-depth responses, when needed.

After each focus group, the research team reflected on key themes discussed and further refined the interview guide. Novel ideas or considerations raised by participants during earlier groups were also posed to later groups for discussion. Examples included gradual access to subforums, seeking immediate answers versus learning module progression, and privacy concerns. All interviews were audio recorded with participants' consent. Participants were provided a US $20 cash incentive for their participation.

Data Analysis

Audio recordings from focus groups (average duration=91 min, range 79-109 min) were transcribed verbatim by the research team and rechecked for accuracy by the second author. Detailed observational notes recorded by the second author during focus groups were integrated within transcripts. For example, observational data included nonverbal gestures (eg, head nods), distracted or nonengaged behaviors (eg, looking at phone), and affect and tone of voice. Data were, then, imported into Atlas.ti (version 8.3.1, ATLAS.ti Scientific Software Development GmbH, Berlin, Germany) for coding and analysis.

A thematic analysis was undertaken to understand participants' perspectives on the app design and functionalities. The purpose was to inform further app development. Thematic analysis focuses on identifying patterns or themes within the data [46,47]. Three research team members conducted primary analyses. This included the first two authors (both involved in data collection), with the additional perspective of a graduate student in clinical psychology not involved in collecting original data (third author).

Consistent with the approach advocated by Miles et al [47], research team members each familiarized themselves with the data, iteratively reviewed each transcript, and independently assigned initial codes. The team, then, came together and discussed, operationalized, and refined each code. Transcripts were, then, recoded. Codes were refined through further discussion and consensus. The research team examined interrelationships and networks among codes, then built and refined the analytic model. Participant matrices [47] were utilized to examine variance in themes across participants and several key demographic features (eg, previous participation in FMF). Participant demographic variables were also imported into Atlas.ti and code co-occurrence tables were examined to assist with this process.

Results

Sample Demographics

Table 2 provides participant demographics. All participants were either adoptive parents or relatives of the child. In all, 80% (20/25) of the sample was female, with caregiver age ranging from 35 to 73 years. Although a wide income range was represented in the sample, over half of the sample had annual family incomes over $75,000. The sample resided mainly in suburban areas. Mean child age was 8.1 years. Intentionally, 40% (10/25) of caregivers had previously received the standard FMF Program. Although 40% (10/25) of participants rated themselves as *very comfortable* with technology, a wide range of perceived comfort is represented in this sample.

Table 2. Participant demographics.

Sample characteristics	Value
Caregiver type, n (%)	
Adoptive parent	18 (72)
Grandparent	5 (20)
Other relative	2 (8)
Caregiver gender, n (%)	
Female	20 (80)
Male	5 (20)
Caregiver age (years)	
Mean (SD)	51.36 (10.29)
Range	35-73
Caregiver race/ethnicity[a], n (%)	
White	23 (92)
Black/African American	2 (8)
Native American/Alaskan Native	2 (8)
Hispanic/Latinx	4 (16)
Caregiver education, n (%)	
High school diploma/ General Education Development	4 (16)
Some college/associates degree	7 (28)
Bachelor's degree	6 (24)
Master's degree	6 (24)
Doctoral/professional degree	2 (8)
Estimated annual family income (US $), n (%)	
Less than 25,000	1 (4)
25,000-34,999	4 (16)
35,000-49,999	1 (4)
50,000-74,999	4 (16)
75,000-99,999	3 (12)
Over 100,000	9 (36)
Did not answer	3 (12)
Type of Community, n (%)	
Rural	3 (12)
Suburban	20 (80)
Urban	2 (8)
Age of child(ren) (years), n=40[b]	
Mean (SD)	8.1 (3.96)
Range	1-17
Previous receipt of FMF[c], n (%)	
Yes	10 (40)
No	15 (60)
Comfort with technology	

Sample characteristics	Value
Mean (SD)	5.20 (1.92)
1 *I find it very difficult*, n (%)	1 (4)
2, n (%)	2 (8)
3, n (%)	2 (8)
4, n (%)	4 (16)
5, n (%)	3 (12)
6, n (%)	3 (12)
7 *I am very comfortable*, n (%)	10 (40)

[a]Nonexclusive categories. No participants identified as Asian, Native Hawaiian/Pacific Islander, or Other.

[b]Several parents also had younger children with a fetal alcohol spectrum disorder, in addition to a child within the study age range.

[c]FMF: Families Moving Forward.

Participant matrices were examined across themes to assess for differences based on demographic characteristics. Overall, themes identified were fairly consistent across focus groups and did not generally differ based on participant demographics. Caregivers who had previously received the standard *FMF Program* occasionally referenced the program when discussing a positive feature of FMF Connect. However, themes did not generally differ (with one exception noted in the Organizing/Guiding theme section) relative to those who had not completed the program. In addition, participants with older children or adolescents were more likely to raise the need for interventions for adolescents and adults. They otherwise communicated similar themes as did participants with younger children.

Global Impressions

Participants were generally enthusiastic about the app and had positive global impressions. Four global impression themes emerged in analysis, which include (1) ease of access or accessibility of the app; (2) the app's ability to guide and organize key information; (3) how the app connects users with information, resources, and other caregivers; and (4) the ability to share information with people outside the app. These themes arose not only in discussions relating to app positive features but also when participants were asked about motivators for using the app and unprompted discussions about system-level barriers (see later section on this topic: *How Families Moving Forward Connect Addresses System-Level Barriers*). Evidence for each of these themes is presented in the following sections. Table 3 is a participant matrix illustrating the high level of agreement among participants on these themes.

Table 3. Participant matrix for global impression themes. Codes in theme cells indicate when participants gave one or more extended utterances (EU) or simple agreement (SA) to comments related to each theme.

Focus group and participant identification number	Previous FMF[a]	Ease of access	Guiding/organizing	Connection	Share with others
Rochester 1					
FG001	Yes	SA[b]	EU[c]	EU	EU
FG002	No	EU	EU	EU	EU
FG003	Yes	SA	__[d]	SA	—
FG004	Yes	EU	EU	EU	EU
Rochester 2					
FG005	Yes	EU	EU	EU	EU
FG006[e]	Yes	SA	EU	EU	EU
FG007[e]	Yes	SA	SA	EU	SA
FG008	Yes	EU	EU	EU	SA
Atlanta					
FG009	No	SA	—	SA	SA
FG010	No	EU	—	EU	EU
FG011	No	—	—	EU	EU
Minneapolis					
FG012[e]	Yes	EU	EU	EU	EU
FG013	No	EU	EU	EU	SA
FG014	No	EU	EU	EU	EU
Rochester 3					
FG015	Yes	EU	EU	EU	EU
FG016	Yes	EU	SA	EU	EU
FG017	No	EU	EU	—	EU
San Diego					
FG018[e]	No	EU	—	SA	EU
FG019[e]	No	EU	—	EU	SA
FG020	No	SA	—	EU	EU
FG021	No	SA	—	EU	SA
Los Angeles					
FG022	No	EU	—	EU	SA
FG023	No	EU	SA	—	SA
FG024	No	EU	EU	EU	EU
FG025	No	SA	EU	—	SA

[a]FMF: standard Families Moving Forward Program.

[b]SA: Simple agreement, defined as at least one single word (eg, *yes*, *I agree*) or nonverbal nod relating to the theme.

[c]EU: Expanded utterance, defined as at least one multiple word phrase or sentence(s) reflecting the theme.

[d]Dashes indicate a participant did not provide a clear nonverbal or verbal response relating to this theme.

[e]Caregiver of an adolescent (aged 13-17 years).

Ease of Access

This theme includes two aspects. First, participants referenced app accessibility as a positive feature. They thought the accessibility of the app would allow more people to obtain needed information and strategies for raising a child with an FASD. For example, one participant said enthusiastically:

> *I think it's awesome because so many more people are going to have access to the Families Moving Forward.* [FG024]

Second, participants spoke of the benefits of having information in one place they could easily access. One caregiver said:

> *What I like about the app is you're not digging through all the papers you were dealing with before [whole group nods yes].* [FG005]

Participants described most needing this information in emotional or stressful moments. The app provides an easy way for parents to meet their needs in such moments. This idea is highlighted by the following quote:

> *...It sounds easy enough for me to navigate through…which is very, very good because… sometimes you're overwhelmed and you're frustrated and you wanna look at something quickly…I can then navigate through that with no problem and get the information.* [FG015]

Guiding and Organizing

This theme was discussed in five of the seven focus groups. The two groups where this theme was not discussed had no participants with previous FMF experience. Participants referenced the fact that the app guides users through step by step learning, starting with the basics and then building on that foundational knowledge. For example, a participant said:

> *Oh I like it so far, very much. And I'm looking at the different content and, I mean definitely it looks like you have built from one thing to the next to help us, … it's good to conquer what came before, before you go on to the next one.* [FG017]

Parents appreciated that the app offers a single place to aggregate and organize information about their child or children. The organization of this information allows parents to track patterns and changes in their children, which can often be difficult and time consuming. For example, one participant said:

> *That [tracking behavior] would be so amazing because I know when I go to the different doctors, or psychiatrists, or neurologists and they ask and I'm like, last week, last month, last day…it's all just one…That would be amazing… you're just able to pull that up and have it right there [FG025 nodding yes] or have a quick way to say like "oh, it's happening again."* [FG024]

Connection

Participants were enthusiastic about how the app connects users with key information, resources, strategies, and other caregivers in the Family Forum. The connection provided by the app was commonly regarded by participants as a motivator to use the app. This is revealed in the following interaction among three participants:

> *Well, I think the forum is going to keep me coming back too.* [FG008]
>
> *Yes, it's that social media part of it, that's going to keep me coming back.* [FG006]
>
> *And I think having access to…information on different services that are available.* [FG007]

Participants also emphasized the ability to share ideas with other users as a positive feature of the app. They were especially interested in having subforums within the Family Forum where they could connect with other users in their geographical area and share information on local resources. For example, a participant said:

> *If there's a way to identify other members that are in your geographical area… We're also always sharing resources. Like, who gets it? I'm so tired of signing up for therapy and the therapist doesn't have a clue what [FASD] is.* [FG012]

Share With Others

Participants recognized that many people who work with their child will not be using the app. They realized providers and school staff will not always have access to or be aware of key information presented in the app. Because of this, parents really liked the option to download fact sheets from the *Library* to share with providers, especially teachers. One parent shared:

> *And, I just think the idea of being able to pull little pieces of it here and there out and helping others understand is very exciting.* [FG017]

Similarly, another parent said:

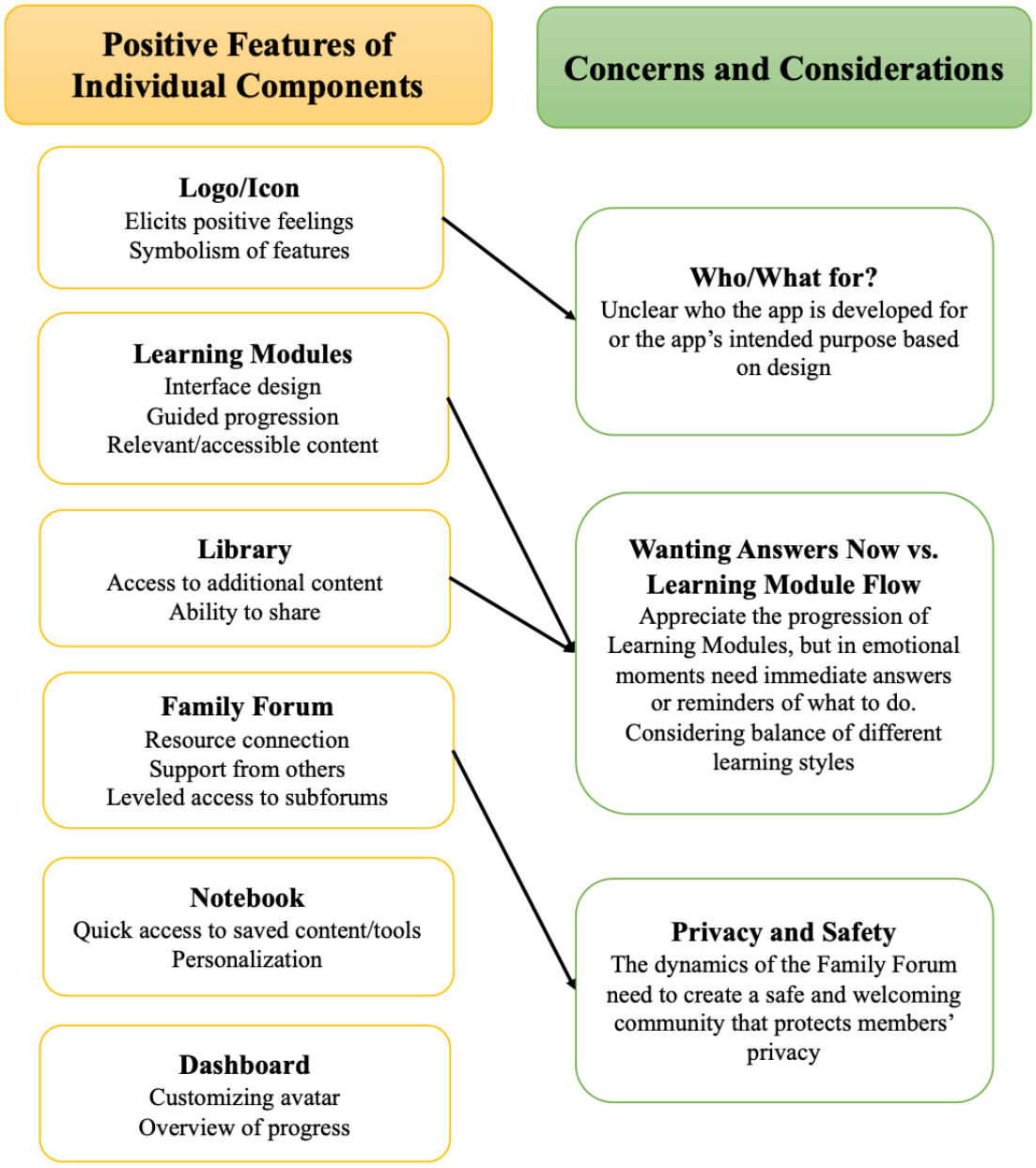
> *It would be great if there was a way to have that as one of the resources, a document you could share with teachers. Um, and maybe even something attached to the child specifically, based on your observations.* [FG014]

This theme generated a lot of enthusiasm in groups, especially in Atlanta and San Diego. As discussed below in the *Identified Remaining Needs* section, participants offered additional content and feature suggestions related to this theme.

Individual Components

Themes related to individual components fell into two categories: (1) positive features based on high enthusiasm and participant discussion, and (2) additional considerations or topics with which participants grappled. In reviewing the data, concerns and considerations tended to come up in relation to specific components. Opinions were mixed or evolved over the course of discussion. A number of constructive suggestions were offered about how to respond to these considerations. In the following sections, both positive features and concerns or considerations will be discussed in the context of individual components. Figure 2 illustrates these graphically.

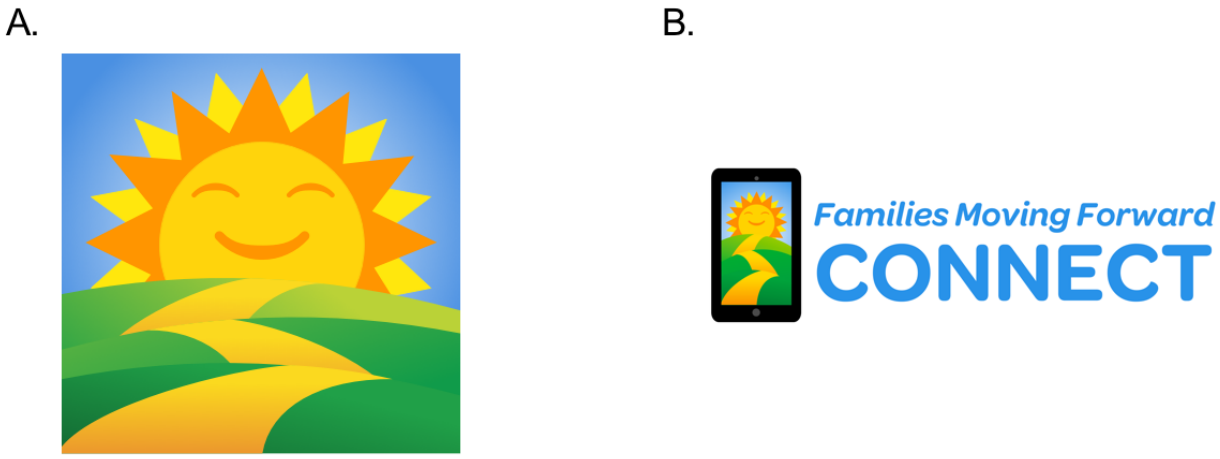
Figure 2. Primary themes relating to the main individual components of the Families Moving Forward (FMF) Connect mobile health intervention. Themes fell into two main categories: (1) positive features and (2) additional considerations and concerns. The considerations and concerns tended to come up in relation to specific components.



Families Moving Forward Connect Logo and Icon

Generally, the logo and icon did not generate a great deal of enthusiasm or engaged discussion (see Figure 3 for illustration). However, some consistent themes emerged. The logo and icon elicited positive feelings such as feeling hopeful, optimistic, and happy. Participants also commented on the symbolism of features. For example, one participant stated:

Figure 3. Part A of this figure is the icon shown to participants in focus groups. For the latter focus groups, this was shown on a phone simulator. Part B of the figure is the Families Moving Forward (FMF) Connect logo presented to participants.



> *I do like the crooked path, because it doesn't look overly simplistic to me. It does convey that this is a process and that there are steps to it, so I really like that.* [FG014]

The concern that was raised by participants for this component is a theme labeled *Who/What for?* Participants noted that it was unclear who exactly this app was developed for from the first glance at the logo or icon. Participants stated that the logo and icon felt somewhat childish and that they may mistake the app as a game for children. The following quote from a participant represents this dilemma

> *It does look aimed towards children, so I wouldn't know it was the right one.* [FG002]

Although the concern about recognizability was raised, participants felt that with time and exposure to the app, the logo and icon would become recognizable and they would associate the logo and icon with the app.

The face on the sun was identified as a key feature by multiple participants contributing to the childish nature of the logo and icon. For example, a participant shared that the icon "pushes this state of euphoria that none of us parents probably have ever experienced. So that piece to me is not reality. And then the sun makes it feel very childish" (FG018). Suggested changes mainly had to do with changing the face of the sun or trying to add features into the logo to make it more clearly related to FASD or brain-based disabilities.

Learning Modules and Library Components

Participants spoke positively of the interface design of the *Learning Modules*. In particular, they enjoyed the individual icons for each of the learning modules (see Figure 4 for the screenshots of *Learning Module* home screens). The primary critique relating to the design interface was that progressing through the *Learning Modules* from the bottom to the top of the screen was, at first, somewhat counterintuitive.

Figure 4. Screenshots of the Learning Modules interface. The first screen is the Learning Modules home page that shows the three levels: (1) Getting Started, (2) Taking Action, and (3) Taking it to the Next Level. The next three screens show the individual modules within each of these three levels.

Participants appreciated the guided nature and step-wise aspect of the learning content, allowing users to build on knowledge and skills from previous modules. They also thought the content areas were relevant. They especially liked having all of the information accessible in one app. This quote represents a caregiver expressing how valuable she felt this content would have been, early on, in raising her child:

> *I would've loved to have had those at the very beginning [whole group looks at FG019 and nods yes]…to have learned…what is FASD. Just like she was saying [gesturing towards FG018], there's so many places to go learn about it but if you're in one central place that is done by professors and people in the know, that would be awesome to be able to do that.* [FG019]

Participants often spoke of the *Library* in relation to the *Learning Modules*. They were positive about being able to extend their learning in the *Library* beyond the standard content offered in the *Learning Modules*. In addition, participants valued that the *Library* allows them to download fact sheets to share with other people in their children's lives.

Regarding the *Learning Modules* and the *Library*, participants grappled with one main concern. This was how to balance the guided progression of learning content with the acknowledged fact that people want content immediately available to navigate at their own pace and direction. Most participants recognized the importance in progressing through the learning content in a step-wise fashion. For example, one caregiver said:

> *You want the modules to work. You know, you want when they get to the end they've accomplished a great thing that will really be a resource for them to go back. It never helps when you're trying to do something that's nonfiction to jump to the end of the book or to start in the middle. [whole group nodding yes].* [FG017]

However, the point was raised that different people have various learning styles and may want access to all the material at once. For example, a participant commented:

> *I understand that you want sequence and building upon the knowledge that you've learned, but I still want more available in the library ahead of time.* [FG025]

This was a lively discussion during many of the focus groups. Participants predicted they would want answers immediately during times when they feel stressed or emotional, or when their children are struggling. The following quote provides an example of a participant who wrestled with understanding the importance of the *Learning Module* progression, while recognizing there are times that require immediate answers:

> *I could see myself, "I want this one, I need the calming strategies right now!" You know? So I understand that moving parents along on a continuum is important and getting them to do the education is important, but hopefully you're also going to find an audience of people that are fairly adept with some of this stuff and they really want to make good use out of specific components as well.* [FG012]

On the more extreme end of this continuum, one participant was clear that having to follow the *Learning Module* progression to get the information she wanted would be a barrier to her using the app:

> *If I can't get a resource within a reasonable amount of time and energy then apparently I don't need that resource… I wouldn't take the route [indicating learning module route] just because I don't know if getting to the end is going to give me enough of a benefit to be invested in it. So, I probably personally would not use it.* [FG018]

Family Forum

Participants were very excited about being able to share resources and ideas with other caregivers in the *Family Forum*. They were also excited about having a space to connect with and gain support from other caregivers. For example, one participant stated:

> *...because I love the fact that if someone had a great idea – not necessarily that it'd work for my kid but it's a possibility…I love the idea of exchanging ideas [FG006 nodding yes emphatically].* [FG005]

Another participant spoke to the value of getting support from other caregivers who understand their experience:

> *...You're in this one forum because other parents have been there or are going through it and it's not that you're wanting, 'oh, it'll be okay,' but you need a spot to be able to vent.* [FG010]

An idea was raised by participants in the first focus group to open certain subforums after the completion of particular *Learning Modules*. They suggested that by opening subforums based on module completion, users would be able to know they are talking with other parents that have the same level of familiarity with the content. That would mean that users could ask for support or advice based on more advanced concepts. This idea was proposed to each subsequent group and was viewed positively.

The biggest concern consistently raised in relation to the *Family Forum* was the importance of ensuring that forum dynamics fostered a safe and supportive community. Protecting one's privacy was often mentioned in these discussions. Participants frequently referenced their experiences in Facebook groups when discussing forum dynamics and safety or privacy concerns. Indeed, forum dynamics were described as a key factor for using the app. One participant stated:

> *One thing with Facebook, it keeps you coming back because there's interaction, there's back and forth, it's not just me and my app and...That's where the forum is going to be important. But the forum's got to be a safe place. [whole group nods emphatically].* [FG001]

In contrast, several participants indicated they would be less likely to use the *Family Forum*. Two participants (FG022 and FG023) stated that they were not social media users and did not think they would use this feature much. Another participant (FG012) thought the features of the *Family Forum* were duplicative of benefits she already receives through existing Facebook groups, but noted other parents might find it helpful.

Participants felt that the presence of trained peer moderators would address the concern of privacy and safety. They also felt it would be beneficial to have clear guidelines to help users provide and seek out productive support. Some participants also talked about a protection strategy of choosing to share different levels of information, depending on the nature of the group. For example, one participant stated:

> *On Facebook there's many different FASD groups and for me personally, how I handle them is I talk very generally on the more national ones… In our local FASD group, I share very intimately…But that's because it's a local, it's monitored, and we screen.* [FG004]

Notebook and Dashboard

The final two components of the app are the *Notebook* and the *Dashboard*. These components did not directly relate to any of the major concerns or considerations brought up by participants and were generally referenced in a positive manner. Participants appreciated that the *Notebook* provided quick access to saved content, as well as to various tools in the app. In addition, they liked that the *Notebook* could provide a personalized section and history of their child. For example, one participant stated:

> *...so that way you can make your own quick reference guide, you know what I mean, within your notebook....So that would keep me coming back.* [FG005]

The *Dashboard* was in the early stages of development when shown to focus groups; therefore, there was limited discussion surrounding this component. Generally, people liked that they could create and customize an avatar and that the dashboard would provide a quick overview of *Learning Module* progress.

Perceived Barriers to App Use

Participants described potential barriers to using *FMF Connect*. Possible barriers fell into four main themes. The first theme involves technological aspects such as the app loading slowly or crashing. For example, a participant (FG024) said that she would not use the app, "if it was 'buggy' and it was hard to like get to the things that I needed – it just like repetitively didn't work." The second theme involves whether information will be presented in ways that are overly complex and involve too much scrolling or navigation between screens. A participant described:

> *For me, scrolling I don't do scrolling. It gives me nausea [FG006 nods enthusiastically]. So, I would rather have chunks of information and then click next.* [FG008]

The third theme includes forum dynamics such as low user activity, negative tones, and overly judgmental posts or comments. For example, a participant said:

> *If people were very negative, very judgmental, I would probably not use it.* [FG003]

The fourth theme raised time and money as other potential barriers. Participants discussed how precious time is when raising children. Therefore, to make it worth their time, they felt the app should be accessible and easy to move through. For example, a participant said:

> *Obviously cost would be a factor. That if it was too costly, it wouldn't be worth my time. ...But again, I'm never opposed to paying...a reasonable amount for something that I'm getting value from.* [FG020]

How Families Moving Forward Connect Addresses System-Level Barriers

Although not specifically queried by the moderator, participants often raised their experience of system-level barriers, resulting

in some lively (and at times emotional) discussions based on shared experiences. The most common system-level barriers raised were (1) limited access to services, (2) feeling isolated, and (3) having to advocate for their children because of providers' and teachers' lack of knowledge about FASD. These barriers are especially notable given the relatively high educational and financial demographics of many families in this sample.

Most of the positive global impression themes were described by participants as helping to meet some needs relating to these barriers. For example, participants spoke passionately about not being able to access services and having very limited access to FASD-informed care. Participants felt the app addressed this barrier by providing a service that was easy to access and offered connections to other parents who may provide suggestions for additional local services. For example, one participant who had just completed the standard *FMF Program* after being on a waitlist for several months said:

> *So now you wouldn't have to wait so long… I had hiccups in my life that made it where I had to not go for a little while. Where if I had an app I could have done it at home.* [FG016]

Next, participants commonly spoke of feeling isolated in their experience of raising children with FASD. They commented that the support and connection provided in the app would help to alleviate or reduce that feeling of isolation. One participant said

> *And I think it's a very often isolating experience and...I really like...the idea that you can connect with other parents and learn from them.* [FG013]

Finally, participants felt the unmet need for advocacy support due to lack of knowledge among providers and teachers would be addressed. This is because users can easily share information from the app with others as a means of advocating for their child in different settings. These advocacy efforts were especially important to participants in reference to school settings. The following quote is fairly representative of participants' discussions of school advocacy and how the app might support their efforts:

> *Can you have handouts that we can give to teachers? Like this is FASD…this is how their brain works sometimes…that kind of stuff that's easy to give a teacher…like an overview of whoever you might be working with. Like a quick, simple this is mostly what you'll see in my kid.* [FG024]

Identified Remaining Needs

Although the app addresses some of the system barriers faced by families, participants identified additional needs that remain. Related to the theme of ease of access, caregivers asked whether app content could be accessed across multiple platforms, such as tablets or internet browsers on computers. They identified that using multiple platforms would make it easier to engage with different app components or different settings. For example, a participant said:

> *I would really like to access it on both [whole group nods yes]. You know for that fast access on the phone but, if I really want to spend time in the app I'd really much prefer it to be on the computer.* [FG008]

Participants also felt very strongly that they should have continued access to the app, even after finishing all *Learning Modules* or a certain length of time had passed. Continued access would allow caregivers to maintain connections with other users in the family forum. Continued access also allowed them to refresh knowledge and skills by reviewing key content. For example, one participant said

> *I definitely think that our children...they change so much. [Whole group nodding yes]...what their needs are, what medications they're on, changes a lot...And so I think being able to go back as if it was one of your favorite books.* [FG017]

A number of families were raising multiple children with FASD, each of whom had quite different needs. Participants wanted a way to consider the needs of multiple children in the app. For example, a participant emphasized the differences between his two daughters with FASD and stated:

> *If you don't address both of them [in the app] you're going to be lost as a parent.* [FG011]

Although participants appreciated the connection provided from the *Family Forum*, they would like it taken a step further by including a resource directory of FASD-informed providers and community resources. Participants also wanted additional features in the app to be used by their children, often related to calming strategies.

Finally, parents requested the development of adjunct or companion apps to aid in advocacy efforts with providers, teachers, respite workers, and other family members. Some participants were raising children older than those targeted for the app. Those parents highlighted the need for apps to be created to support adolescents, adults, and their caregivers.

Discussion

Principal Findings

This study represents an important step in the systematic development and evaluation of the *FMF Connect* mHealth intervention for caregivers raising children with FASD. Inclusion of key stakeholder feedback early in the app development process is a major strength of this process [43,44]. App-based interventions have the advantage of scalability and can potentially reach many in need and reduce significant barriers to care [12,48]. *FMF Connect* is one of the few parenting interventions of its kind. It is based on an intervention tailored for its diverse target population and is the first self-directed mHealth intervention for FASD.

Results from this study revealed that participants were largely enthusiastic about the app's initial design and functionalities. The positive global impression themes identified by participants (ease of access, guiding and organizing, connection, and share with others) are consistent with the functions for which mHealth apps are well suited [29,49]. Participants related how these

positive global themes could address some of the system-level barriers they encounter. Examples include limited access to services, feeling isolated, and high advocacy needs related to a lack of knowledge about FASD. The positive global themes were also primary factors identified by participants that would motivate them to use the app.

Participants evaluated many positive features about individual app components and functionalities. Yet, they also identified potential barriers to using the app, raising some important concerns and considerations relating to several app components. This knowledge will inform further refinements and evaluation of the app. For example, the suggestion by participants to open subforums based on module completion has already been implemented in a beta-testing version of the app. Similarly, feedback relating to behavior tracking and icon design has been taken into account in app refinement. Further, features supporting improved *Family Forum* dynamics have been added. This iterative and systematic approach to app development makes it more likely that the app will be acceptable and effective for families.

Limitations

FMF Connect is being developed for the US population. Although efforts were made to recruit a diverse sample of caregivers raising children with FASD, some subgroups are not well represented in the study sample. Multiple geographic areas of the United States were sampled. Given resource limitations and logistics, however, coverage did not reach all regions. Most notably, the study did not enroll any biological parents, despite multiple recruitment efforts designed to engage this important subgroup. It is possible that many biological parents may not be comfortable in group research settings because of stigma or other factors. Future studies could include alternate data collection methods, such as individual interviews, or identify other relationship-building strategies to engage this group. It is important to remember that birth parents comprised about 15% overall of the standard *FMF Program* research participants, so the original material was designed to be sensitive and useful with that subgroup. Videoconferencing may also work to include members of other geographic areas or underrepresented subgroups. The next stage of the evaluation of the *FMF Connect* app beta-testing will (1) integrate focus group and interview data collection methods and (2) explore videoconferencing as a method to increase sample diversity.

Although themes were similar across groups, the size of each focus group was smaller than anticipated. A total of 4 to 12 participants were scheduled for each group but, unfortunately, because of inclement weather, illness, and other unanticipated scheduling conflicts, participant turnout was lower than expected. A smaller group size did not seem to negatively impact the flow of discussion, but larger groups might have resulted in additional themes or different patterns of results.

Potential for possible biases should also be considered. The research team is developing the app, which could elicit a positive response bias from participants. In addition, some participants knew the moderator from other Rochester area FASD services, and other participants were recruited through CIFASD investigators at other sites. These participants may have come to focus groups with previous positive associations with the research team or their colleagues. Although the moderator encouraged negative feedback during focus groups and participants gave a range of critical and constructive feedback, it is impossible to completely rule out this source of bias.

Participants with strong, vocal opinions or different experiences (eg, previous FMF involvement) could also potentially influence themes in focus groups. The moderator made efforts to elicit feedback from all participants and seek differing opinions. Findings from this study do not suggest past FMF involvement influenced data, with the exception of possibly the guiding/organizing theme. Selection bias is also possible. Participants who had a favorable view of apps may have been more interested in participating in the study.

Broader Applicability of Findings

Findings from this study may inform other mHealth apps with families raising children with FASD or other special needs. This study engaged key stakeholders in the early development process of the *FMF Connect* intervention. Such engagement has been a critical process in the adaptation of other evidence-based interventions to a self-directed digital format [50]. This study showed the utility of stakeholder feedback, which yielded specific ideas that could be implemented immediately or used to guide further exploration.

For *FMF Connect*, caregivers thought the learning module content was relevant and gradually built on foundational knowledge. This input suggests that adaptations to the standard *FMF Program* sequence for the *FMF Connect* app make sense and are acceptable to stakeholders. However, participants grappled with retaining the step-wise progression of the intervention versus obtaining immediate answers and advice. Although several ideas were suggested, no clear solution was identified. This will be a key consideration to explore in next steps of prototype refinement with iterative feedback from stakeholders.

Participants were enthusiastic about being able to share ideas and connect with other caregivers in the *Family Forum*. They felt that connecting with other parents could reduce the feelings of isolation and help address barriers of finding and accessing services for their children and family. These findings are consistent with qualitative research assessing in-person and online peer support for parents of children with special needs. Findings from this research literature reveal the assets of shared experiences, mutual support, encouragement, and knowledge sharing [51,52]. Positive outcomes commonly described include themes of improved coping and assurance about child management strategies, less isolation, and ability to support other families. In this study, participants tempered their excitement about the *Family Forum* with concerns about privacy and safety. These concerns were often based on their past experiences with social media such as Facebook. These concerns have also been raised in other studies [53]. The attention to fostering a safe, nonjudgmental, and welcoming environment was stressed. Having clear guidelines and peer moderators were viewed as positive protections in the *Family Forum*. These findings are informative for other mHealth interventions

incorporating social media that are targeted to families of children with special needs.

Participants in this study placed positive emphasis on the accessibility and organization of content. The fact that content would be evidence based was viewed favorably. Previous reviews of Web-based information geared toward parents raising children with developmental disabilities document problems with the quality, consistency, and readability of information [24]. Participants in this study described gathering information from many different sources and sometimes struggling to find what they needed. They liked that the app would include evidence-based information all in one place. They were also enthusiastic about the *Notebook* component as a way to organize and individualize content for their child. These results suggest that deriving mHealth interventions from existing evidence-based treatments may be particularly well received by users, especially when apps tailor and personalize information.

Previous research has documented that self-directed digital interventions derived from well-studied therapist-led parenting programs are effective in improving child and parenting outcomes [31-33]. Effect sizes are larger for studies with clinical samples and interventions that use interactive content and formats to engage users [31]. In the field of FASD, two known studies have utilized Web-based intervention formats. Kable et al [54] found that a self-directed brief Web-based intervention for families raising children with FASD had multiple outcomes similar to a therapist-led workshop. The Strongest Families intervention, which integrates weekly Web-based content and coaching telephone calls with a trained coach, is also being tested with caregivers of children with FASD [55]. Although analysis is underway from the larger randomized controlled trial, early usability data with a small subsample found the website was easy to navigate and that content was written at a level that was understandable to families [56]. During early development, some challenges were identified with caregivers learning how to use the interface initially, having too much scrolling, and specificity of content to FASD, which were improved in subsequent testing. This body of research speaks to the potential of the *FMF Connect* mHealth app and the benefits of rigorous, systematic research. Much of the format, behavior change principles, and content of *FMF Connect* could be quite relevant for other parenting apps, especially those targeting parents of children with special needs.

Conclusions

FMF Connect builds on a solid foundation of empirical research on a tailored intervention designed for families raising children with FASD [19,36]. Capitalizing on the promise of the field of mHealth, *FMF Connect* has the potential to reach many families in need and reduce significant barriers to care. This can result in broader public health impact. This study's findings will guide further app development both in terms of content and technological advances to optimize intervention effects. Next steps will involve the completion of initial programming, iterative small-scale beta-testing and refinement, and larger feasibility testing. A large-scale randomized controlled trial is, then, planned to evaluate the efficacy with respect to caregiver and child outcomes. This rigorous development process can be an example in the field of mHealth and for the future of parenting interventions.

Acknowledgments

All or part of this work was done in conjunction with the CIFASD, which is funded by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Additional information about CIFASD can be found on their website [57]. Research reported in this publication was supported by the NIAAA of the National Institutes of Health under Award Number U01AA026104. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SCRI and the University of Washington are recognized for their contributions of intellectual property from the standard FMF Program. The standard *FMF Program* was developed, tested, and materials refined by a team led by author HO, based at and sponsored by these institutions (led by SCRI), with funding via multiple grants from the Centers for Disease Control and Prevention.

We appreciate all the families who made time in their busy schedules to participate in this study. This research would not have been possible without their valuable insights.

Authors' Contributions

All authors contributed to the preparation of this work. CP conceived of the study, led app content adaptation and development, obtained necessary approvals, conducted focus groups, and led data analysis and manuscript preparation. JP assisted with app content development, coordinated participant recruitment, collected observational data during focus groups, managed data quality, processing, and analysis, and assisted with manuscript preparation. CK assisted with app content revisions, data analysis, and manuscript revisions. CT led the technological development of the app, assisted with focus group moderation, and contributed to manuscript revisions. HO is the developer of the standard *FMF Program*, from which the app is derived. She made significant contributions to intellectual property exchange, app content adaptation and development, and assisted with manuscript revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mock-ups of the Families Moving Forward (FMF) Connect interface design shown to participants in focus groups. Screen mock-ups were shown to participants using an interactive interface in Invision or initial prototype (iOS). Components shown include: app icon, Dashboard, Learning Modules, Family Forum, Library, and Notebook.
[PNG File , 821 KB - mhealth_v8i4e14721_app1.png]

References

1. May PA, Chambers CD, Kalberg WO, Zellner J, Feldman H, Buckley D, et al. Prevalence of fetal alcohol spectrum disorders in 4 US communities. J Am Med Assoc 2018 Feb 6;319(5):474-482 [FREE Full text] [doi: 10.1001/jama.2017.21896] [Medline: 29411031]
2. Popova S, Lange S, Shield K, Burd L, Rehm J. Prevalence of fetal alcohol spectrum disorder among special subpopulations: a systematic review and meta-analysis. Addiction 2019 Jul;114(7):1150-1172 [FREE Full text] [doi: 10.1111/add.14598] [Medline: 30831001]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Washington DC: American Psychiatric Publishing; 2013.
4. Astley SJ. UW Departments Web Server. Seattle, WA: University of Washington; 2004. Diagnostic Guide for Fetal Alcohol Spectrum Disorders: The 4-Digit Diagnostic Code. Third Edition URL: https://depts.washington.edu/fasdpn/pdfs/guide2004.pdf [accessed 2020-02-11]
5. Hoyme HE, Kalberg WO, Elliott AJ, Blankenship J, Buckley D, Marais A, et al. Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. Pediatrics 2016 Aug;138(2):pii: e20154256 [FREE Full text] [doi: 10.1542/peds.2015-4256] [Medline: 27464676]
6. Bodnar TS, Raineki C, Wertelecki W, Yevtushok L, Plotka L, Zymak-Zakutnya N, Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD). Altered maternal immune networks are associated with adverse child neurodevelopment: Impact of alcohol consumption during pregnancy. Brain Behav Immun 2018 Oct;73:205-215 [FREE Full text] [doi: 10.1016/j.bbi.2018.05.004] [Medline: 29738852]
7. Gupta KK, Gupta VK, Shirasaka T. An update on fetal alcohol syndrome-pathogenesis, risks, and treatment. Alcohol Clin Exp Res 2016 Aug;40(8):1594-1602. [doi: 10.1111/acer.13135] [Medline: 27375266]
8. May PA, Gossage JP. Maternal risk factors for fetal alcohol spectrum disorders: not as simple as it might seem. Alcohol Res Health 2011;34(1):15-26 [FREE Full text] [Medline: 23580036]
9. Chasnoff IJ, Wells AM, King L. Misdiagnosis and missed diagnoses in foster and adopted children with prenatal alcohol exposure. Pediatrics 2015 Feb;135(2):264-270. [doi: 10.1542/peds.2014-2171] [Medline: 25583914]
10. Amendah DD, Grosse SD, Bertrand J. Medical expenditures of children in the United States with fetal alcohol syndrome. Neurotoxicol Teratol 2011;33(2):322-324. [doi: 10.1016/j.ntt.2010.10.008] [Medline: 21073947]
11. Brownell MD, de B Hanlon AC, Macwilliam LR, Chudley AE, Roos NP, Yallop LP, et al. Use of health, education, and social services by individuals with fetal alcohol spectrum disorder. J Popul Ther Clin Pharmacol 2013;20(2):e95-106. [Medline: 23648378]
12. Petrenko CL, Tahir N, Mahoney EC, Chin NP. Prevention of secondary conditions in fetal alcohol spectrum disorders: identification of systems-level barriers. Matern Child Health J 2014 Aug;18(6):1496-1505 [FREE Full text] [doi: 10.1007/s10995-013-1390-y] [Medline: 24178158]
13. Ryan DM, Bonnett DM, Gass CB. Sobering thoughts: town hall meetings on fetal alcohol spectrum disorders. Am J Public Health 2006 Dec;96(12):2098-2101. [doi: 10.2105/AJPH.2005.062729] [Medline: 17077397]
14. Brems C, Boschma-Wynn RV, Dewane SL, Edwards AE, Robinson RV. Training needs of healthcare providers related to Centers for Disease Control and Prevention core competencies for fetal alcohol spectrum disorders. J Popul Ther Clin Pharmacol 2010;17(3):e405-e417. [Medline: 21063036]
15. Gahagan S, Sharpe TT, Brimacombe M, Fry-Johnson Y, Levine R, Mengel M, et al. Pediatricians' knowledge, training, and experience in the care of children with fetal alcohol syndrome. Pediatrics 2006 Sep;118(3):e657-e668. [doi: 10.1542/peds.2005-0516] [Medline: 16950957]
16. Rojmahamongkol P, Cheema-Hasan A, Weitzman C. Do pediatricians recognize fetal alcohol spectrum disorders in children with developmental and behavioral problems? J Dev Behav Pediatr 2015 Apr;36(3):197-202. [doi: 10.1097/DBP.0000000000000146] [Medline: 25767932]
17. National Organization on Fetal Alcohol Syndrome (NOFAS). Resource Directory URL: https://www.nofas.org/resource-directory/ [accessed 2019-05-14]
18. Astley SJ. Profile of the first 1,400 patients receiving diagnostic evaluations for fetal alcohol spectrum disorder at the Washington State Fetal Alcohol Syndrome Diagnostic & Prevention Network. Can J Clin Pharmacol 2010;17(1):e132-e164. [Medline: 20335648]
19. Bertrand J, Interventions for Children with Fetal Alcohol Spectrum Disorders Research Consortium. Interventions for children with fetal alcohol spectrum disorders (FASDs): overview of findings for five innovative research projects. Res Dev Disabil 2009;30(5):986-1006. [doi: 10.1016/j.ridd.2009.02.003] [Medline: 19327965]

20. Bobbitt SA, Baugh LA, Andrew GH, Cook JL, Green CR, Pei JR, et al. Caregiver needs and stress in caring for individuals with fetal alcohol spectrum disorder. Res Dev Disabil 2016 Aug;55:100-113. [doi: 10.1016/j.ridd.2016.03.002] [Medline: 27058320]
21. Petrenko CL, Alto ME, Hart AR, Freeze SM, Cole LL. 'I'm Doing My Part, I Just Need Help From the Community': intervention implications of foster and adoptive parents' experiences raising children and young adults with FASD. J Fam Nurs 2019 May;25(2):314-347. [doi: 10.1177/1074840719847185] [Medline: 31079560]
22. Olson HC, Oti R, Gelo J, Beck S. 'Family matters:' fetal alcohol spectrum disorders and the family. Dev Disabil Res Rev 2009;15(3):235-249. [doi: 10.1002/ddrr.65] [Medline: 19731388]
23. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The role of online social support in supporting and educating parents of young children with special health care needs in the United States: a scoping review. J Med Internet Res 2016 Dec 22;18(12):e333 [FREE Full text] [doi: 10.2196/jmir.6722] [Medline: 28007689]
24. Hall C, Culler E, Frank-Webb A. Online dissemination of resources and services for parents of children with autism spectrum disorders (ASDs): a systematic review of evidence. Rev J Autism Dev Disord 2016;3(4):273-285 [FREE Full text] [doi: 10.1007/s40489-016-0083-z]
25. IQVIA Institute for Human Data Science.: IQVIA Institute for Human Data Science; 2017 Nov 7. The Growing Value of Digital Health: Evidence and Impact on Human Health and Healthcare System URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-05-01] [WebCite Cache ID 783WkuOMv]
26. Pew Research Center. 2019 Jul 12. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-05-01] [WebCite Cache ID 783PjeHCw]
27. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, 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]
28. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
29. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
30. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
31. Baumel A, Pawar A, Kane JM, Correll CU. Digital parent training for children with disruptive behaviors: systematic review and meta-analysis of randomized trials. J Child Adolesc Psychopharmacol 2016 Oct;26(8):740-749. [doi: 10.1089/cap.2016.0048] [Medline: 27286325]
32. Corralejo SM, Rodríguez MM. Technology in parenting programs: a systematic review of existing interventions. J Child Fam Stud 2018;27(9):2717-2731 [FREE Full text] [doi: 10.1007/s10826-018-1117-1]
33. Hansen A, Broomfield G, Yap MB. A systematic review of technology‐assisted parenting programs for mental health problems in youth aged 0–18 years: applicability to underserved Australian communities. Aust J Psychol 2019;71(4):433-462 [FREE Full text] [doi: 10.1111/ajpy.12250]
34. UW Departments Web Server. Families Moving Forward Program URL: http://depts.washington.edu/fmffasd/home [accessed 2019-11-27]
35. Olson HC, Montague RA. An innovative look at early intervention for children affected by prenatal alcohol exposure. In: Adubato SA, Cohen DE, editors. Prenatal Alcohol Use and Fetal Alcohol Spectrum Disorders: Diagnosis, Assessment and New Directions in Research and Multimodal Treatment. Potomac, MD: Bentham Science Publishers; 2011:64-107.
36. Petrenko CL, Pandolfino ME, Robinson LK. Findings from the families on track intervention pilot trial for children with Fetal Alcohol Spectrum Disorders and their families. Alcohol Clin Exp Res 2017 Jul;41(7):1340-1351 [FREE Full text] [doi: 10.1111/acer.13408] [Medline: 28440861]
37. Card JJ, Solomon J, Cunningham SD. How to adapt effective programs for use in new contexts. Health Promot Pract 2011 Jan;12(1):25-35. [doi: 10.1177/1524839909348592] [Medline: 19858321]
38. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
39. Wiggins G, McTighe J. What is backwards design? In: Understanding By Design. First Edition. Upper Saddle River, NJ: Merrill, Prentice Hall; 1998:7-19.
40. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]

41. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. Prof Psychol Res Pr 2016;47(2):155-162. [doi: 10.1037/pro0000069]
42. Gondek D, Edbrooke-Childs J, Fink E, Deighton J, Wolpert M. Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: a systematic review. Adm Policy Ment Health 2016 May;43(3):325-343 [FREE Full text] [doi: 10.1007/s10488-015-0710-5] [Medline: 26744316]
43. Eborall H, Morton K. Use of focus groups in developing behavioural mHealth interventions: a critical review. In: Barbour R, Morgan D, editors. A New Era in Focus Group Research: Challenges, Innovation and Practice. London: Palgrave Macmillan; 2017:109-127.
44. Baek E, Cagiltay K, Boling E, Frick T. User-centered design and development. In: Spector J, Merrill MD, van Merriënboer J, Driscoll MP, editors. Handbook of Research on Educational Communications and Technology. Abingdon: Routledge; 2008:660-668.
45. Petrenko CL, Pandolfino ME, Roddenbery R. The association between parental attributions of misbehavior and parenting practices in caregivers raising children with prenatal alcohol exposure: A mixed-methods study. Res Dev Disabil 2016 Dec;59:255-267 [FREE Full text] [doi: 10.1016/j.ridd.2016.09.005] [Medline: 27662038]
46. Liamputtong P. Qualitative data analysis: conceptual and practical considerations. Health Promot J Austr 2009 Aug;20(2):133-139. [doi: 10.1071/he09133] [Medline: 19642962]
47. Miles MB, Huberman AM, Saldana J. Qualitative Data Analysis: A Methods Sourcebook. Third Edition. Los Angeles, CA: SAGE Publications Inc; 2014.
48. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
49. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
50. Breitenstein SM, Shane J, Julion W, Gross D. Developing the eCPP: adapting an evidence-based parent training program for digital delivery in primary care settings. Worldviews Evid Based Nurs 2015 Feb;12(1):31-40. [doi: 10.1111/wvn.12074] [Medline: 25588505]
51. Niela-Vilén H, Axelin A, Salanterä S, Melender H. Internet-based peer support for parents: a systematic integrative review. Int J Nurs Stud 2014 Nov;51(11):1524-1537. [doi: 10.1016/j.ijnurstu.2014.06.009] [Medline: 24998788]
52. Law M, King S, Stewart D, King G. The perceived effects of parent-led support groups for parents of children with disabilities. Phys Occup Ther Pediatr 2001;21(2-3):29-48. [Medline: 12029852]
53. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
54. Kable JA, Coles CD, Strickland D, Taddeo E. Comparing the effectiveness of on-line versus in-person caregiver education and training for behavioral regulation in families of children with FASD. Int J Ment Health Addict 2012 Dec;10(6):791-803 [FREE Full text] [doi: 10.1007/s11469-012-9376-3] [Medline: 31024223]
55. Turner K, Reynolds JN, McGrath P, Lingley-Pottie P, Huguet A, Hewitt A, et al. Guided internet-based parent training for challenging behavior in children with fetal alcohol spectrum disorder (Strongest Families FASD): study protocol for a randomized controlled trial. JMIR Res Protoc 2015 Oct 13;4(4):e112 [FREE Full text] [doi: 10.2196/resprot.4723] [Medline: 26462968]
56. Hundert AS, Huguet A, Green CR, Hewitt AJ, Mushquash CJ, Muhajarine N, et al. Usability testing of guided internet-based parent training for challenging behavior in children with fetal alcohol spectrum disorder (Strongest Families FASD). J Popul Ther Clin Pharmacol 2016;23(1):e60-e76. [Medline: 27115205]
57. Collaborative Initiative on Fetal Alcohol Spectrum Disorders. URL: https://cifasd.org/ [accessed 2020-02-11]

Abbreviations

CIFASD: Collaborative Initiative on Fetal Alcohol Spectrum Disorder
FASD: fetal alcohol spectrum disorder
FMF: Families Moving Forward
mHealth: mobile health
NIAAA: National Institute on Alcohol Abuse and Alcoholism
SCRI: Seattle Children's Research Institute

Edited by G Eysenbach; submitted 17.05.19; peer-reviewed by A Yin, K Mateo, C Jacob, S Van Stee; comments to author 03.10.19; revised version received 27.11.19; accepted 24.01.20; published 06.04.20.

Please cite as:
Petrenko CLM, Parr J, Kautz C, Tapparello C, Olson HC
A Mobile Health Intervention for Fetal Alcohol Spectrum Disorders (Families Moving Forward Connect): Development and Qualitative Evaluation of Design and Functionalities
JMIR Mhealth Uhealth 2020;8(4):e14721
URL: https://mhealth.jmir.org/2020/4/e14721
doi:10.2196/14721
PMID:32250274

Original Paper

Proposal for the User-Centered Design Approach for Health Apps Based on Successful Experiences: Integrative Review

Guillermo Molina-Recio[1], RN, MSc, PhD; Rafael Molina-Luque[1], RN, MSc, PhD; Ana M. Jiménez-García[1], RN; Pedro E Ventura-Puertos[1], RN, MSc, PhD; Alberto Hernández-Reyes[1], RD, MSc; Manuel Romero-Saldaña[2], RN, MSc, PhD

[1]Department of Nursing, University of Córdoba, Córdoba, Spain

[2]Department of Occupational Safety and Health, Córdoba City Hall, Córdoba, Spain

Corresponding Author:
Rafael Molina-Luque, RN, MSc, PhD
Department of Nursing
University of Córdoba
Av. Menéndez Pidal 7
Córdoba
Spain
Phone: 34 957 218 101
Email: rafael.moluq@gmail.com

Abstract

Background: Different strategies encompassed within mHealth have shown themselves to be effective for maintaining good health or controlling certain diseases. However, there is usually a very high rate of abandonment of health apps. Therefore, it would seem obvious that there is a need for involving the end users (whether they are health professionals, patients, or both) in the design process from the early stages in order to enable their needs and characteristics to be identified. In this sense, it is common knowledge that focusing on the user permits the consideration of valuable details aimed at making the correct adjustment between the patient, the technology, and the organization of attention.

Objective: The goal of the research was to propose a methodology based on the review of previous successful user experiences in setting up health apps by using qualitative techniques (focus groups and discussion groups) that includes the participation of information technology and health professionals and the patients themselves.

Methods: An integrative review was made of studies in which a qualitative methodology was employed mainly through focus and/or discussion groups for the design and development of health apps, consulting diverse databases (PubMed, Scopus, and Proquest) with the following search strategy: “mHealth AND apps AND focus group OR discussion group.” A total of 69 papers were included in the review.

Results: A proposal structured in 4 sessions of variable duration was made in which information technology and health professionals and patients take part: composing, preparing, and organizing contents (session 1); testing structure and usability (session 2); does the app fit the needs of end users? (session 3); and last testing—keep on improving (session 4). Throughout the sessions, we propose studying aspects like previous user experiences in mHealth, barriers to the adoption of mHealth, interface contents, management and browsability, usability, perceived quality, security and privacy, capacity to self-manage disease with the app, ergonomics, and glanceability, etc. Specific tools that have proved useful in previous research for measuring these aspects are presented.

Conclusions: These work sessions would be based on predominantly qualitative methodologies although, as they evolve, validated questionnaires permitting the assessment of the objectivity of certain technical aspects could be incorporated. With this proposal, a project centered on end users could be effected, responding to their needs. However, this requires validation that will be made via implementation in the development of health apps, with the subsequent measurement of results in terms of adherence and improvement in the clinical variables of the end users.

(JMIR Mhealth Uhealth 2020;8(4):e14376) doi:10.2196/14376

KEYWORDS

mHealth; user-centered development; focus groups; discussion groups; interdisciplinarity

Introduction

Mobile technology already forms part of our daily lives, and its presence is increasing rapidly. It is estimated that in 2019 there were over 2.7 billion smartphone users and around 1.4 billion tablet owners worldwide [1]. In addition, technical improvements in mobile devices, including larger screens, higher resolution, increase in browsing speeds, and development of many thousands of mobile apps with a multitude of new functionalities [2], have meant a genuine social and cultural revolution reaching all strata of society. As a result, the incorporation of mobile technology into our daily tasks has triggered changes in the way in which we live, work, communicate, and relate to each other socially [3].

According to the Global System Mobile Association, there are more devices connected to the network than people in the world. In 2017, 7.42 billion mobile connections were identified, whereas the population census in the world was of 7.23 billion [4]. Another relevant fact that helps to size up the magnitude of this technological trend is that, in 2014, for the first time, the number of accesses and the browsing time on the Web through mobile devices exceeded those made with desktop computers [3,5-7]—so much so that the future of technology and that of the mobile telephone are considered to be equal and it is very difficult to distinguish between one from the other. Thus, it is thought that within a few years, we shall be able to dispense with the adjective mobile when referring to technologies, as they will all have that characteristic [3].

The health care field has not been alien to this revolution. The term mHealth (mobile health) was used and defined for the first time in the year 2000 [8]. This concept was subsequently employed at the mHealth Summit 2010 of the Foundation for National Health Institutes to refer to “the rendering of medical attention services by means of mobile communication devices” [9], and nowadays this is understood globally as being medical practice and public health based on the use of mobile devices [10]. Since then and up to the present, around 40% of the more than 300,000 apps available in the different apps stores are related to health topics, with those focusing on disease monitoring and management standing out [11]. Different strategies included in mHealth, from simple phone calls or sending of texts (short message service, or SMS) to the use of apps as a support for clinical decision making or telemedicine, have shown themselves to be effective in the communication between patients and health professionals, in changes toward healthy lifestyles (giving up smoking or increasing physical activity), in the improvement of disease management (in diabetes or asthma, for example), and in the increase in adherence to treatments [12-15].

Further, it has been demonstrated that certain functionalities that are implicit in the habitual use of smartphones, like dissemination of information, possibility of self-monitoring with easy and intuitive record systems, interaction between users or using gamification strategies, also have positive effects on the state of the users’ health [16].

We must not forget that the popularity, mobility, and technical capacity of these devices mean that, as many people have them and are never separated from them, it is possible for synchronization between the health professionals and the patients without needing the former’s physical presence, either to give them special care or to warn them about any risks or changes in their health that require more urgent attention [15].

In this respect, a national survey of habitual users of mobile apps in the United States demonstrated that 58.23% of them had installed at least one health app, nutrition and physical activity ones being prominent. However, many had given up using them or had uninstalled them, mainly due to lack of time for entering the data; lack of interest; because after downloading the app for free, there were hidden costs that only appeared after a trial period (freemium models); difficulties in using them, or because of data being shared in the social networks or among groups of friends that they did not want to divulge [17].

In essence, and as reported in some other works [3,16,17], the abandonment rate with these apps is usually higher when a user has a bad experience. More than half of the customers who either uninstalled apps or did not have any continuity in their use claimed this reason, despite the apps being indispensable for their health care, especially in the case of chronic diseases. Therefore, as recommended by Alonso-Arévalo and Mirón-Canelo [3], any public or private entity involved in the design, development, and implementation of an app related to the health field should take into account all of these aspects and highlight their functionality, their being easy to use, their compatibility, performance, and safety.

In fact, it would be fitting to involve the end users of the apps (whether they be health professionals or patients, or both) in the process of designing them during the early stages in order to identify their needs and characteristics. In this regard, it is known that a user-centered approach permits the contemplation of useful details aimed at forging an adequate relationship between the patient, the technology, and health care organizations [18]. This participation of end users and health professionals in all stages of app development could result in an increase in their commitment and an improvement in integration, self-management, and health results, since most apps in which end users and health professionals did not participate in development have been seen to fail [19]. Besides, there is a consensus on the suitability of the use of qualitative techniques that permit the inclusion of all the actors implicated [20], in which the focus and discussion groups stand out [21]. However, although certain conceptual frameworks that could serve as a guide for setting up health apps have been proposed [22], there is no clear evidence-based proposal for the sequence and contents of each of the sessions.

The main objective of this study was, therefore, to suggest a methodology or guidelines (including tools for assessment) based on the review of previous successful user-centered experiences for the development of health apps by means of qualitative techniques (focus groups and discussion groups) that include the participation of information technology and health professionals and patients themselves.

Methods

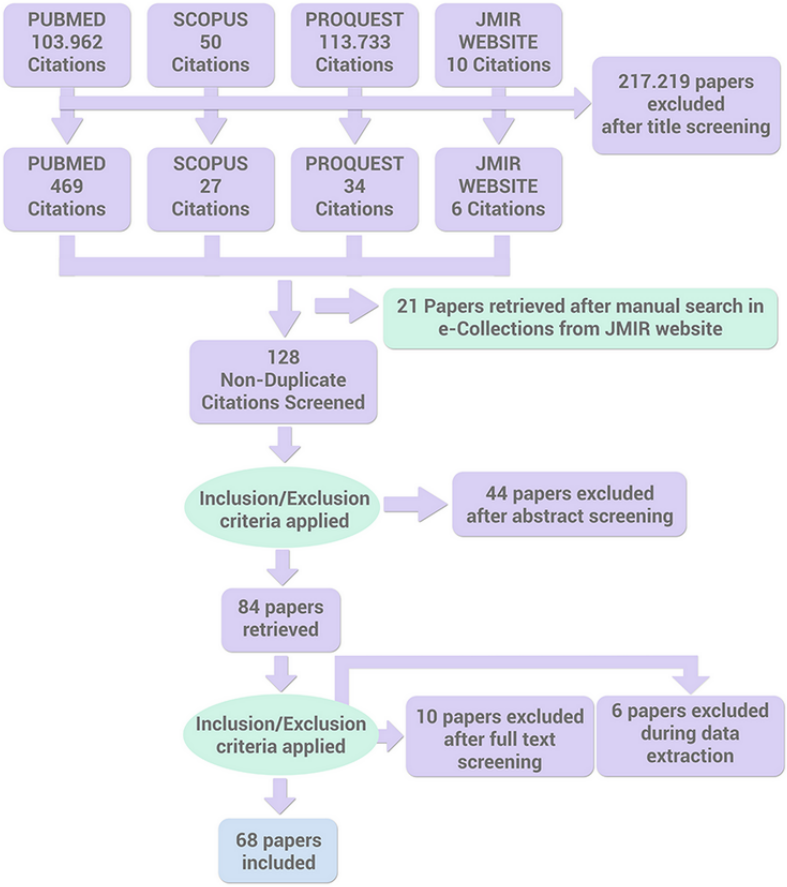
Study Design

An integrative review was completed of studies in which a qualitative methodology was employed (ie, focus and/or discussion groups) for the design and development of health apps. The electronic databases consulted were PubMed, Scopus, and Proquest, with the terms "mHealth," "apps," "focus group," and "discussion group" forming the search algorithm "mHealth AND apps AND focus group OR discussion group." To locate papers not indexed in these databases, a manual search in JMIR journals (especially in e-collections) was performed. Analysis and selection of the manuscripts was performed by two experts in epidemiology and quantitative research (who assessed the scientific quality of the manuscripts) and one researcher with extensive experience in qualitative research (who assessed the quality of the information provided on the techniques used and the evolution of the sessions). The manuscripts were also evaluated by two researchers with previous experience in the development and evaluation of health apps.

Eligibility Criteria

The publications were located and selected between January 2000 and June 2018. Only those written in English or Spanish that were available as a complete text were considered. A reverse investigation was also made to prevent zones of silence. Articles referred to by the studies reviewed, which a priori would not have been found in the databases consulted, were chosen. Finally, 69 articles were included in the review (Figure 1).

Figure 1. Paper selection flowchart.



The following information was extracted from the selected papers: type of study, number of participants, duration of the intervention, methodology used for the design of the apps, number of sessions held, main results, and questionnaires and other validated tools used in the sessions. If the manuscript included specific information on the questions used in the focus or discussion groups, it were recorded verbatim. With regard to the type of information used to make the proposal, the technique employed in developing the mobile app was followed: what type of information and how the latter was obtained for the apps design by means of the use of focus and/or discussion groups, number of participants recommended for training, duration of the sessions, how to evaluate the results obtained by the researchers, and, especially, which qualitative techniques contributed to the final state of the app.

This information was synthesized in order to establish a methodological proposal that included all sections considered for the design, development, and start-up of an app that attends to the needs of the end users and can serve as a guide for researchers and software developers with a view to offering products more in accord with the real demand on the apps health care market.

The decision on the type of session to be held for designing user-centered health apps (focus or discussion groups) and their

duration was made following the recommendations of Lane et al [23], Sáez et al [24], and Savin-Bandenet al [25].

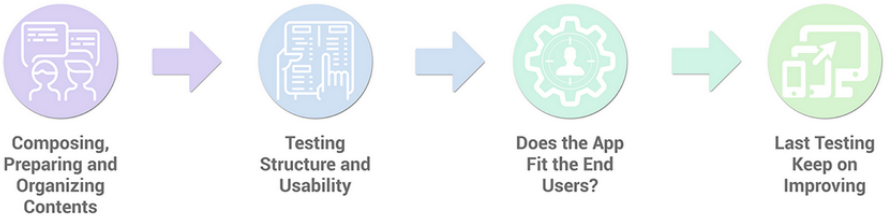
Results

Four Sessions for Implementing User-Centered Health App Design

Based on the articles reviewed, a method was proposed for the development of health care apps through qualitative techniques (focus groups and discussion groups) structured over 4 sessions of variable durations that included information technology and health professionals and patients. Diverse works with successful results have indicated that the fundamental objective of the inclusion of all these actors is to produce a design centered on end users that permits the detection of their needs, tests new behavior change concepts [18], increases their adherence to the developing app, and obtains positive behavior changes in health [26]. This practice permits one to unite the health care and information technology professionals' technical knowledge, forming interdisciplinary teams that provide better results and a greater research impact [9], all of which are more positively valued by the users [27].

A basic outline of the structuring of the sessions and the fundamental topics dealt within them is shown in Figure 2. In Multimedia Appendix 1, readers can find verbatim quotations from participants (patients, stakeholders, or health care professionals) when qualitative tools (focus or discussion groups) were applied.

Figure 2. Steps in the design of user-centered health apps.



First Session: Composing, Preparing, and Organizing Contents

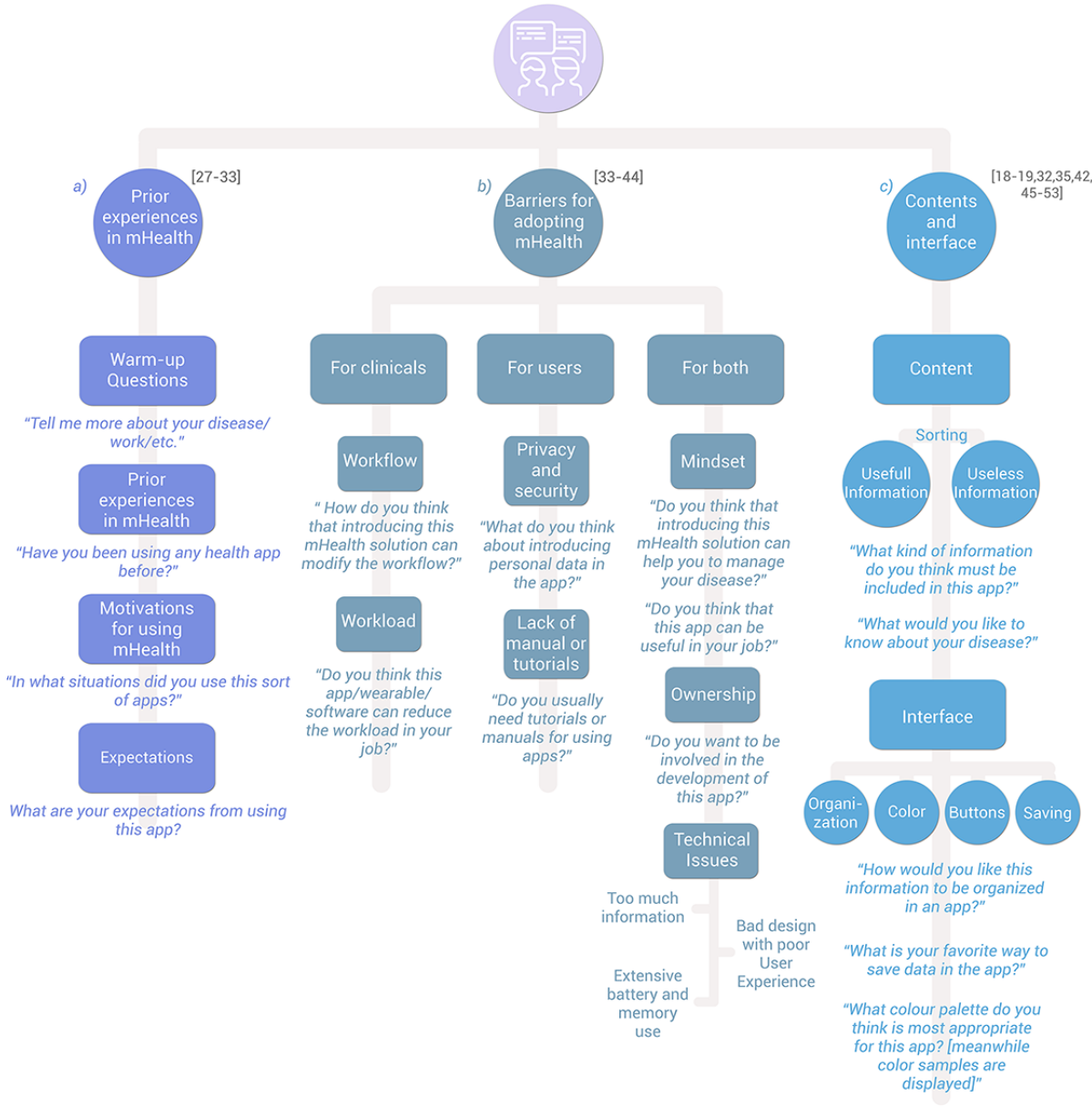
Session Overview

Given the generalist characteristics of the introductory themes to be addressed in the first session and the importance of its being an open debate so that the largest amount of opinions and attitudes are aired about the development of a certain health app, a discussion group is proposed with health professionals, software developers, and patients or end users. This is considered to be more appropriate than a focus group because it encourages the participants to work together, thus obviating the influence of a moderator [23,24]. The discussion group will have approximately 10 participants, last between 60 and 90 minutes, and address topics on contents and mobile health.

Previous User Experiences in mHealth

Previous works have obtained positive results by assessing the previous experiences of the targeted end users and showing how they can influence the use, browsing, and capacity to apply knowledge offered to health care management before beginning to elaborate a technological solution. For example, the work done by Greenfield et al [28] included the study of the previous experiences, motivations, and expectations for the use of wearables as a mechanism for promoting health in truck drivers. Other similar initiatives were put into practice by Pulman et al [27] for the development of mobile apps for adolescents with type 1 diabetes, Mirkovic et al [29] in the design and contents of an app for improving cancer patient care, and Cox et al [30] on the use of apps to enter anthropometric data or food intake. This strategy is very important in both elderly [31] and younger people [32]. In any case, it is important to include all final users, especially health care professionals and patients. In this regard, it is worth highlighting the experience of Lyles et al [33] in which a tablet app was designed for being used in the waiting room of a primary care consultation for complex patients that allowed prioritizing of the issues to be treated in these visits [33]. A proposal of the questions to ask, the topics to be addressed, and their sequence order is shown in Figure 3. See Multimedia Appendix 1 for quotes from participants in previous experiences.

Figure 3. Session 1: Composing, preparing and organizing contents.



Barriers to the Adoption of mHealth

Another, closely related aspect that can be addressed in this first session refers to the difficulties or impediments that end users may find in incorporating this type of technology into the health care field. Some user experiences have emphasized the need to collect this information in order to increase the effectiveness of the interventions based on mobile technology. In a pilot study performed by Hao et al [34], participants expressed their discontent with the reception of sample results in a laboratory via an SMS system. Their complaints arose from their lack of participation in the design, since they were only involved in it after its implementation; they expressed their wish to be informed early on about the project and wanted to have their opinions considered. This situation triggered a lack of motivation for performing that intervention and, more important, to give it continuity in time. Problems were observed deriving from the workload and flow and security and privacy of the data. Other studies highlighted more technical aspects as barriers, such as the lack of an appropriate instructions manual, too much information, an unattractive design, and excessive battery and internal memory consumption [33-41]. Last, it is commonly perceived that the patients do not really have any power to make decisions on their processes [41-44]. These perceptions can end up being obstacles (both for the professionals and the patients) that ought to be resolved starting from the first sessions. An extensive list of questions on obstacles and facilitator elements for the adoption of mHealth has been described in the research of Giunti et al [41]. A summary of the way in which to approach these issues in the first session is shown in Figure 3.

Contents and Interface

Without a doubt, one of the most important decisions in beginning to set up a health care app is organization of contents. Discussing these aspects with the end users is, therefore, fundamental, and there is ample evidence of its importance and

positive effects. Shishido et al [45] developed a mobile app for compiling and reporting instructions for the evaluation of cardiometabolic risk that used graphic components like radio buttons and pull-down lists, giving rise to a standardization of frequently used data input to make it easy to complete forms. This proposal was made on the basis of suggestions from 5 nurses and a nutritionist, who gathered this type of information habitually in their work, and it was highly valued by all users. Casillas et al [46] also used qualitative methodologies through which they improved the contents and interface of a Web- and SMS-based system to guarantee access to comprehensive quality care for young adults surviving childhood cancer. Also, Kok et al [42], in their intervention on preventive treatment against the recurrence of calculi by means of a mobile app, reported that the use of striking colors and easy data input encouraged a greater adherence to the app. Many other research works have explored the contents and interface characteristics prior to the development of health apps for prevention, monitoring, or treatment of different diseases or creation of lifestyle changes: HIV [35], cancer and other chronic conditions [47], gout [48], multiple sclerosis [49], weight management [50], increasing physical activity [32], cardiovascular diseases [51], idiopathic arthritis [52] or chronic pain [53]. In view of these and other user experiences [18,19,26,33], some questions that could be included in this first session have been proposed and are shown in Figure 3. Some participant answers are also shown in Multimedia Appendix 1.

Second Session: Testing Structure and Usability

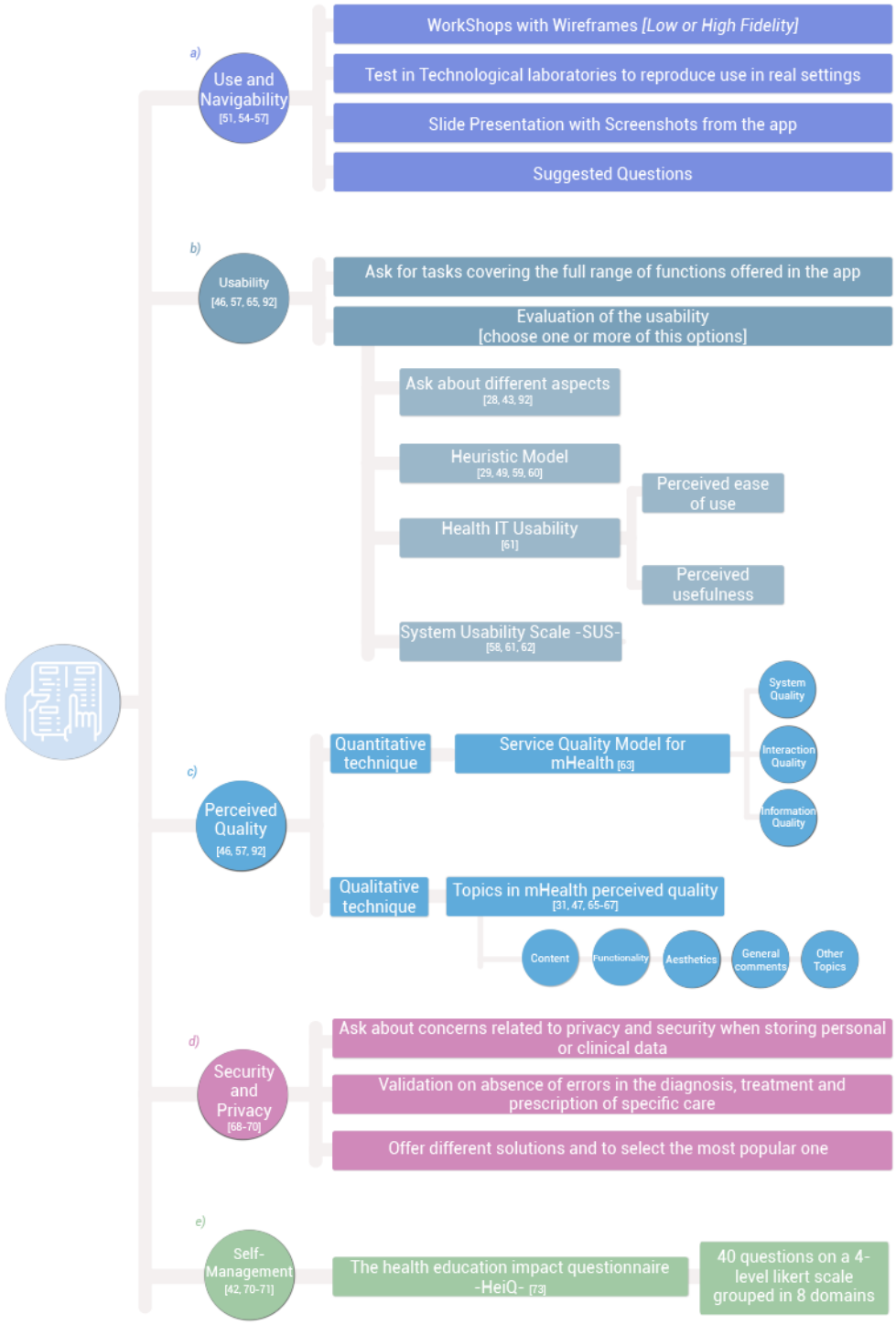
Session Overview

After a first design and implementation of the app (that can be presented with a viable minimum product, mock-ups, or high- or low-fidelity wireframes), it is helpful to have a second session using focus groups so that the end users can give their opinions on the structuring of the information and the characteristics of its use. In this case, the group will put itself in the hands of the moderator as it needs to be directed in order to assess some specific aspects or to execute some tasks. The focus group will have approximately 10 participants, last between 60 and 90 minutes, and address topics on structure and usability.

Management and Browsing

The importance of easy management and navigation (browsability) in adherence to the use of health apps is evident [54]. The experience of Jakobsen et al [55] in the development of My Osteoporosis Journey, an app for the control of osteoporosis in recently diagnosed women, stands out. Different workshops were run in which wireframes were shown with different degrees of details (from low to high fidelity) and with a test in technological laboratories, where it was attempted to reproduce real conditions of use. Health professionals and women affected by the pathology reported a high degree of satisfaction with the app. In this same regard, some user experiences in the development of other apps for the self-management of illnesses or in their prevention have also resulted in products being well accepted by end users, when those aspects have been previously discussed with them [49,51,56,57]. A summary of the proposal for evaluating these aspects is shown in Figure 4 (more detailed information in Multimedia Appendix 2).

Figure 4. Session 2: Testing structure and usability.



Usability

This factor represents possibly the greatest obstacle for the adoption of information and communication technologies in the health care settings. Generically, it refers to being easy to use, and its assessment is based on methods for identifying specific problems, centering on the user-app interaction and the study of the degree to which technology can be satisfactorily integrated into the task envisaged [46,49,51,58]. The latter requires device systems and characteristics to include facility of use, intuitive design and interoperability by means of esthetic issues [43], screen resolution, recharging time (for wearables), and the relevance of the data supplied [28]. It has been demonstrated that taking all of this into account in the early stages of development improves the predictability of the products and is time and cost saving [49].

Some specific methodologies that assess and improve the usability of a mobile app have been employed with good results

in mHealth. For example, Mirkovic et al [29] used a heuristic usability model with high-fidelity prototypes in the design of an app to help cancer patients by contemplating the evaluation of 8 aspects that detected improvement areas [59]. This heuristic model, with small variations, was also used to assess the usability of apps developed for telerehabilitation [49] and increasing physical activity [60] of patients with multiple sclerosis. Likewise, Brown et al [61] created the Health Information Technology Usability Evaluation Model (Health-ITUEM) scale, which evaluates 9 aspects and gauges, basically, simplicity of use and perceived usefulness. These authors have also identified two possible reasons why the usability of the health apps has traditionally been reduced that should be addressed in group sessions: (1) small screen with low resolution and no keyboard or mouse available and (2) connectivity problems. Last, another option is the one used by Ribu et al [62], who applied an instrument developed in the 1990s for the self-management of diabetes using 10 Likert-type items, the System Usability Scale, to measure app usability [63]. This scale was also used by Vilardaga et al [58] to evaluate the usability of a mobile app to quit smoking in people with serious mental illness. These types of assessment (which are basically quantitative) can also be completed with a qualitative analysis through open-ended questions on the difficulty of performing different tasks assigned to the participants. A clear example of the use of this methodology is the research conducted by Mann et al [64] on the evaluation of the usability of an app aimed to improve iron intake and bioavailability in premenopausal women. Consult Figure 4 and Multimedia Appendix 2 for a summary of the way to explore usability in the design of health apps.

Perceived Quality

In a study by Akter et al [65], perceived quality is defined as the user impression of the excellence of the mHealth service. These authors developed an instrument to measure it in health apps with a scale comprising 22 Likert-type items grouped into 3 primary dimensions: system quality (user perception with respect to technical level of communication), interaction quality (between the service supplier and the user), and information quality (benefits of the service's processes or what consumers receive as a result of their interaction with a supplier). These 3 dimensions comprise 8 other subdimensions (Multimedia Appendix 2). Other works based on the use of interactive wireframes and mock-ups [31,47,66,67] or apps already available in app stores [68] have measured quality by means of qualitative methods, asking about general aspects like contents, functionalities, or esthetics, leaving a margin for general comments or for any other item related to mHealth. Both approaches have been seen to be adequate for the measurement of perceived quality.

Security and Privacy

The inclusion of clinical experts in app design and considering the opinions of end users increases the security of the system objectively and ensures that the security is perceived, which increases use adherence. For example, Hilliard et al [19] involved patients with cystic fibrosis in the development of their app using semistructured in depth interviews and discussing design problems, among them those relating to security, and disclosed that most people's concerns revolved around the storage of their clinical and personal data. Offering different alternatives (customization of the mode and type of data stored, inclusion of privacy options, etc) and listening to participant opinions permitted the development of an app that was perceived as being safer. Privacy has also been successfully explored in the development of apps for reminders of taking medication in HIV patients [69] and for stress management in cancer survivors [67], showing how for the former it is a fundamental aspect ("The 'Did you take your medicine notification?' is a problem. Did you take your medication? Anybody in their right mind is going be, 'What do you mean you take medication?' It lets them know you're sick. Be hiding it from your family") while for the latter is an unimportant topic ("My life is not that exciting" and "I have nothing to hide"). In addition, security in mHealth is also related to the diminution of errors in the transmission of information and the advice or health care given, so that, if necessary, this aspect should also be evaluated. Research projects like the ones conducted by the team of Surka et al [69] to improve the detection of cardiovascular diseases, Holmen et al [70] focused on the management of type 2 diabetes, and Jibb et al [71] centered on developing an app for treating pain in adolescents with cancer all constitute good examples of how to address these important aspects. Except for some cases [67], there is general agreement from all end users on the importance of the use of passwords to regulate access to their data [28,72].

Self-Management

One of the desirable characteristics offered by health apps (especially those directed toward controlling chronic diseases) caters to the possibility of empowering users to make decisions on their process (coping skills, target setting, self-monitoring, environment modification, etc) instead of merely providing the care prescribed by the experts [42]. Thus, in this second session, it is necessary to estimate the characteristics and functionalities the app should have to facilitate the self-management of one disease in particular [71]. Holmen et al [70] designed a low-intensity self-management intervention for patients with type 2 diabetes using a mobile app (Few Touch) and successfully used the Health Education Impact Questionnaire [73] to measure the impact of the app on the self-management of this disease. This questionnaire comprises 40 Likert-type items with 4 response levels grouped into 8 domains (Multimedia Appendix 2), and it could be employed in this second session to estimate the effect of the app and monitor its development, although it would be of interest to collect these data again after prolonged use of the definitive version.

Third Session: Does the App Fit the Needs of End Users?

Session Overview

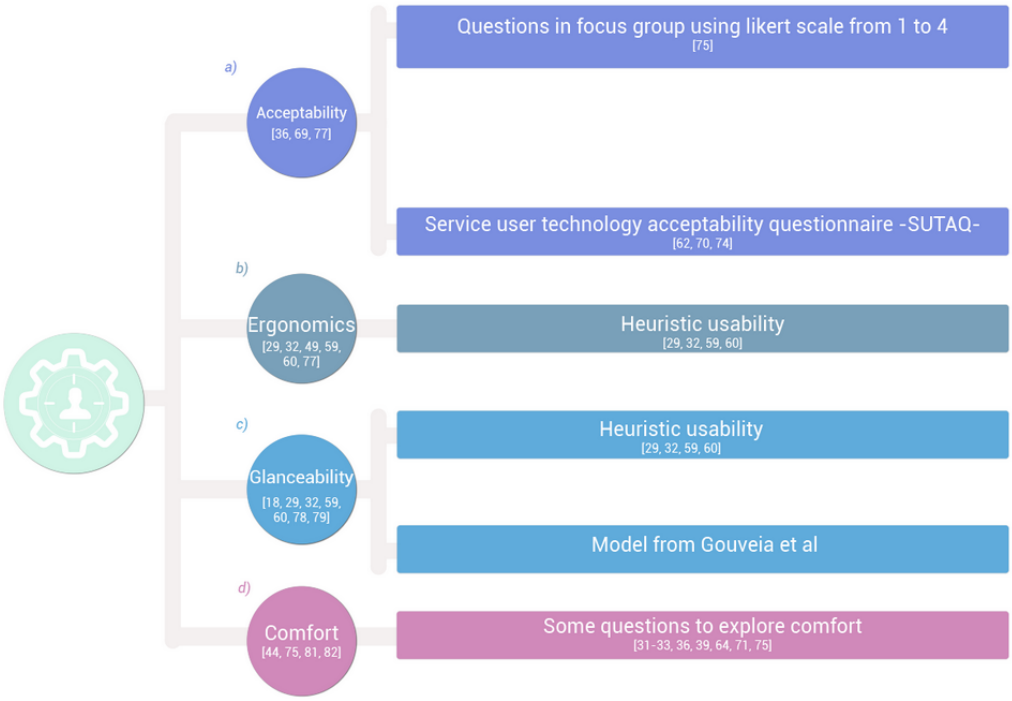
This will be completed several weeks after access to the version to be assessed so that end users can make an adequate evaluation of the app. This time period will vary according to the number of functionalities included in the app and its estimated daily use time and should be decided by the participants in the previous session. A focus group is recommended, with an estimated duration of 90 minutes and a similar number of attendees. Many

of the topics to be addressed in this session are closely related to usability (or they have been seen in the previous session as part of it), but once the app's development status has advanced, some of its characteristics should be revised individually. This would allow us to make a more exhaustive study of the different functionalities and the way to present them in the interface. It is recommended to gather information on the needs and desires of the user.

Acceptability

Acceptability is an intimate concept directly related to usability, and it refers to the extent to which the patients are satisfied with a service and are willing to use it [62]. In the low-intensity intervention for type 2 diabetes patients of Ribu et al [62,70], the Service User Technology Acceptability Questionnaire was employed at a pilot phase with good results. This questionnaire contains 22 questions (that can be applied to satisfaction with an app or with any other technological solution) in the following domains: enhanced care, increased accessibility, privacy and discomfort, personal care concerns, kit as substitution (when the mHealth solution includes some device or wearable kit), and satisfaction [74]. Another alternative for the study of acceptability was demonstrated by Eisenhauer et al [75] in which, during a focus group, they used surveys based on a Likert-type scale and administered by the researcher to evaluate the satisfaction and use guidelines of an app for the self-monitoring of food and physical activity patterns. Also using focus groups, Dworkin et al [69] evaluated the acceptability of an app that showed an avatar for reminding HIV-positive men who have sex with men to take antiretroviral medication ("He looks so real, and he's a nice attractive man, and I'm going to ask him a lot of questions about medication! This is genius idea!"), while Duff et al [36] measured the degree of acceptance of the Medfit app for the improvement of cardiovascular disease self-management through changes in lifestyles ("I found the progress part very useful. I got a reality check when I saw what I was doing and thought I was more active than I am"). Similarly, Van der Weegen et al [18] investigated the requirements of users of a mobile device for stimulating physical activity in primary care, including the end users in the design process from an early stage. These authors demonstrated that centering on the users improves the relationship between them, the technology, and health care organizations, and, therefore, the acceptability of the tool. Other researchers expressed the same perspective when studying the health of war veterans [66], people wishing to lose weight [76], and those wanting to give up smoking [77]. Also, as Mirkovic et al [29] concluded, the acceptability of an app is influenced by the phase of the disease, since needs change and they influence the perception of the utility and acceptability of the properties of the system. Therefore, using different qualitative techniques, the characteristics of the app must be defined in chronic pathologies that include a progressive deterioration [29,36,69], and this could be relevant in this session or, in the case of this evolution being expected habitually, in earlier phases when interface and content aspects are addressed (first session). More details about tools to be used for evaluation of acceptability of a health app and some answers from focus group participants are shown in Figure 5 (with more detailed information in Multimedia Appendix 3) and Multimedia Appendix 1, respectively.

Figure 5. Session 3: Does the app fit the needs of final users?



Ergonomics

With this characteristic, the extent to which the mobile app adapts itself to the needs of a patient is evaluated. One way of doing so is by means of the heuristic usability questionnaire [29,49,59,60], already mentioned, specifically in its fourth dimension, which refers to using with parsimony the resources available on the principal screens and of the state of the app [77] and limiting the use of pop-ups and other notifications [29]. Ergonomics ensures clarity and simplicity of use, so it can also be assessed through qualitative techniques such as focus groups. For example, Simon et al [32] showed that clarity and ease of use was well valued by users of an app promoting an active lifestyle ("I think the app is easy and very clear. And there is not too much in it. With some apps you are like 'Where do I find that again?' but that is not a problem with this one"). At this point, the work of the graphic designer and software developers should be centered on guaranteeing the adaptability of the app to different mobile devices (responsive features) and to the special needs of some end users.

Glanceability

This characteristic of mobile apps refers to the information being comprehensible at first sight or with occasional glances, requiring a minimum of attention and effort to understand it [78]. That is to say, this refers to the perception and interpretation of the information after users have given their attention to the interface, which can be measured on the basis of the speed and ease with which the messages from the app can transmit the information after being seen [79]. For instance, the inclusion of demonstration videos, images, and other audiovisual items and the use of widgets and showing results in graphics can increase it [18]. It can also be evaluated by the heuristic usability questionnaire [29,49,59,60], so its evaluation would be simultaneous to that of ergonomy. In addition, Gouveia et al [80] created a system with 6 glanceability aspects used to select the best of 21 physical activity trackers, and we believe this system could also be employed in making health apps. The 6 aspects proposed by these authors are as follows:

- Data summary (abstract): measure in which the data shown appear as being already processed and permit users to process, quickly become aware of, and reflect on their health behaviors
- Integration with existing activities: the degree to which the most relevant information for the user and the app is integrated into places frequently accessed and, therefore, commonly seen by users
- Comparison to target and norms: this aspect is employed to assess whether the app provides comments on the user's progress in such a way that they are easily processed and evaluated by the users themselves, providing clear feedback
- Being actionable: another desirable quality as a part of glanceability is that the interface would offer effective feedback and information but also trigger short actions to fulfill the health goals proposed
- Leading checking habits: this refers to the results being presented not only as being user-friendly but that they should urge the users into acquiring the habit of verifying them (ie, systematically examining the screen of the app) and consulting their progress. For that reason, it is important that this information can be checked at a single, quick glance [18]. Novelty (that the app continually presents different types of data) and scarcity (when the behavior feedback is shown for only a limited time) are strategies which have proved to be effective for this purpose
- Proxy to further engagements: it is known that, with the passing of time, users stop consulting their data. An app with a high degree of glanceability will trigger moments not expected by the user that would act as signals and increase their commitment with the use and consultation of the data that the app offers. One strategy could be to present information that generates questions instead of giving replies; another could offer ideas that surprise the user

See Figure 5 (with more detailed information in Multimedia Appendix 3) for a summary of the tools shown to be effective for evaluating glanceability.

Comfort

With this term we refer to the capacity of the health app to make disease management easier, due to the employment of mobile technology assistance in the collection, processing, and analysis of health information [44,75,81]. It also means that the app under development, if it is for professionals, will be a more efficient and faster method for disseminating knowledge within the scientific community [37,45]. In this regard, Zanetti [82] recognizes the need to find strategies that strengthen scientific creativity, with research based on the setting up of new technologies. This aspect can be assessed in a focus group to define the characteristics the app should have to make management of disease easier [31-33,36,58,64]. For example, in the development and evaluation of an app for diabetic foot care [31], it was found that the content of the interface was too small for these types of patient, who usually also have retinopathy ("...that's too small an interface for my eyes because I've had retinopathy, I've had laser surgery on both eyes, I've had cataracts removed off both eyes"). Given that there are no objective tools for evaluating comfort quantitatively in the use of a health app, this theme can be addressed in this session once aspects like the interface and usability, to which it is directly related, have been defined. Some questions used in prior studies [39,71,75] that evaluated this characteristic can be posed in this session and are found in Figure 5 (and Multimedia Appendix 3). See, also, some answers from participants in Multimedia Appendix 1.

Fourth Session: Last Testing. Keep on Improving

Session Overview

The fourth and last session (with a similar duration and number of participants) should be held 3 weeks after the third session so that end users will have had time to test the last versions of the app. The fundamental goal of the session will be to define all characteristics and improvements and future development of the app, with the aim of obtaining a product that could be validated in real settings and with a larger number of patients. During this time, based on the user experience, participants will

be able to reflect on the topics addressed previously and discuss the future of the app.

Proposals for Improvement

The aim of this block will be to explore whether any difficulties have appeared or any individual needs have been detected and assess the response offered by the app. In this way, the possibility of introducing improvements that allow us to give specific answers to the largest possible group of users will be considered.

As various works have shown, customization is important to users to satisfy individual preferences and disease management goals [19,56,80,83,84]. Similarly, a more formal or clinical language should be incorporated for some functions (description of pathologies) but more informal language for others (evaluation of conducts) if the users wish. As some authors have pointed out [42,72], it is important to strike a balance between the mobile app being attractive and amusing but not so much so that it discredits the sense of authority. In this sense, Mirkovic et al [29], using the heuristic usability model [49,59,60] during the development of an app for assistance in the management of their illness to cancer patients, found that the users demanded the possibility of having a configuration menu to select the visibility of the principal functions. The work of Koskinen and Salminen [85] also stands out, in which the elaboration of an app for increasing healthy living habits highlighted the importance of the configuration of menus to augment use adherence. Some of the parameters recommended (some for new users and others for advanced ones) were to (1) enable or disable health parameters, (2) aggregate new properties to an existing parameter, (3) modify the presentation of data, (4) add new parameters, (5) change parameters and existing properties, and (6) aggregate or modify units (only through the XML edition). Open-ended questions in focus groups have also proved useful in assessing the need for customization for end users. Thus, for example, in the development of the app for taking antiretroviral medication discussed above, one of the participants commented on the need to customize the alarms and physical characteristics of the avatar [69]. Also interesting is the contribution of some users who participated in the design and evaluation of an app to facilitate self-monitoring and management of mood symptoms in young people, by revealing that it could be important to let each person choose colors to better define the moods which participants could experience [86]. Other works also valued positively personalization for elaborating contents that could be sent by mail to the doctor [48], further adapting the contents of the messages according to the achievements recorded [87], etc.

People's needs change throughout the health-disease process in which they are immersed. For that reason, a health app, especially for end users who have chronic conditions, should be capable of accommodating their preferences and objectives in managing their illness in terms of the stage at which they find themselves [19,43,88,89]. Flexibility has been successfully measured as part of the study of heuristic usability (component 6) [29,49,59,60] and of the model Health-ITUEM [61].

Usefulness

This characteristic can be assessed by surveys (original and/or specific, for evaluating aspects closely related to the health theme addressed or validated and focused on improvement in the quality of life, adherence to treatment, or advances in clinical parameters) [29,90]. Besides, there is sound evidence that the utility perceived in health apps maintains a direct relationship with the continuity in their use [91]. In any case, this should refer to the usefulness to illness management of monitoring and accessing information sources [19,34,54,72]. The usability study model Health-ITUEM [61] includes assessment of utility perceived as part of user satisfaction with the app; estimation of how easy it is to learn to use; perception of the skill needed to perform tasks (the extent to which the users trust their ability to do tasks using the system); speed of task completion; and flexibility or capacity to personalize the app. The usefulness of the developed apps can also be evaluated through specific questions in the focus groups [31,32,36,64]. For example, an app for the control of gout was considered helpful because it allowed the patient to become aware of what food could cause the attacks ("...so as I'm putting in an attack I can access the relevant triggers that have caused me issues in the past") [48]. Another that offered telerehabilitation in patients with multiple sclerosis was considered suitable for presenting videos with different types of exercises [49] ("I think the exercise videos are good because a lot of the movements are what you do in therapy. So, this is along that line to get you moving more"), while users of another app to increase physical activity in patients with type 2 diabetes mellitus assessed the app's ability to motivate ("It made me feel motivated.... I would (exercise) because I was afraid they were going to say, 'Hey! Get off that sofa!'") [87].

Hardware Limitations

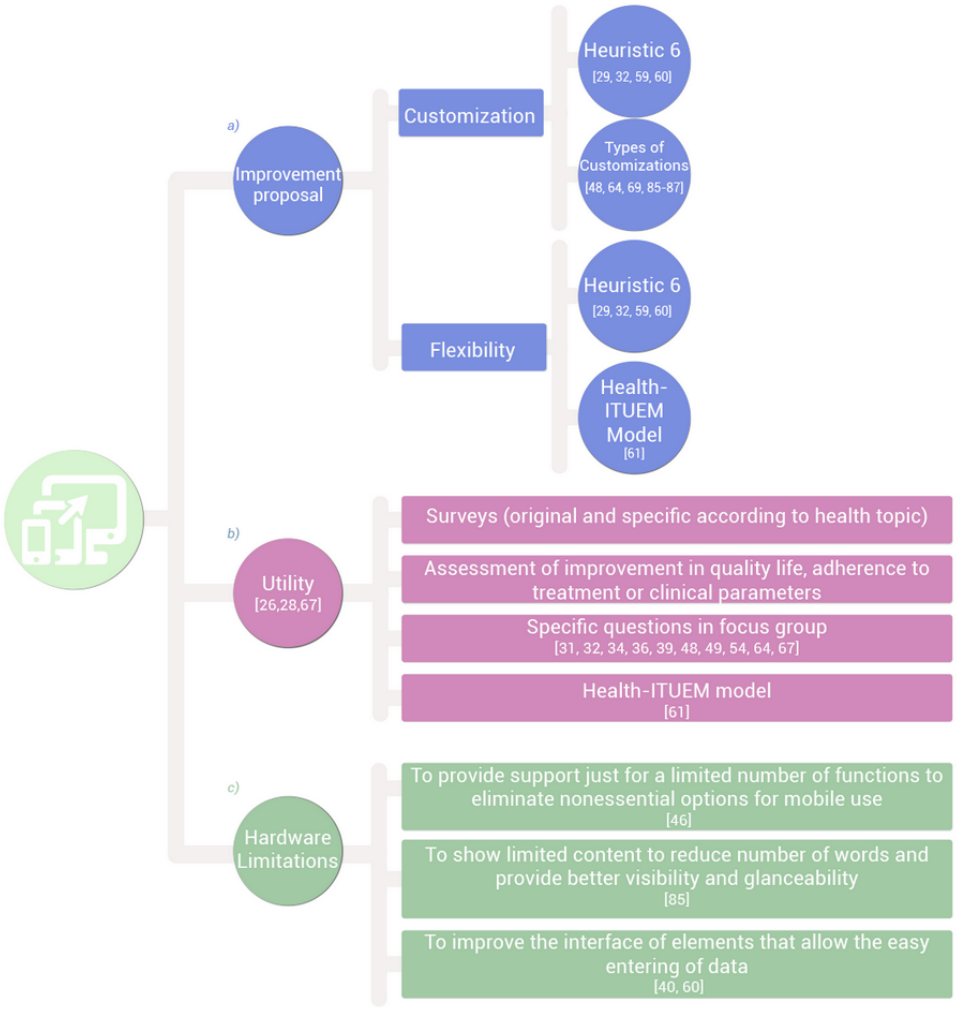
One aspect of reality makes it difficult to develop health apps. Technology advances rapidly but the test equipment for the end user and software move slowly [59]. For example, one problem that has been highlighted in previous user experiences occurs during an attempt to migrate a Web-based system to a mobile platform (a fairly frequent practice in mHealth), a challenge for software developers and designers due to the limitation in screen size and input capacity of some devices [29]. General measures toward solving hardware limitations that have proved effective and whose application can be agreed upon in this session include the following:

- Provide support only to a limited number of functions in order to eliminate the variety of options not fundamental to mobile use [46]
- Show limited contents to reduce word count and facilitate better visibility and glanceability [85]
- Improve the interface with elements that permit easy data input [40,60]

In view of the characteristics of the topics to be dealt with in this session and the continuity it maintains with the previous one, the focus group is also considered as being the most suitable qualitative technique, with a similar duration to the previous ones [25]. It is recommended that the same participants as before participate in it. A structured summary of this fourth session

can be seen in Figure 6 (with more detailed information in Multimedia Appendix 4) while verbatim quotations from participants in the focus group are shown in Multimedia Appendix 1.

Figure 6. Session 4: Last testing. Keep on improving.



Discussion

Principal Findings

This is a proposal to implement the design and development of health apps using a user-centered approach. For this purpose and based on previous experiences of other researchers who obtained positive results, a sequence of 4 group sessions has been proposed. Scientific evidence has demonstrated the effectiveness of exploring in these sessions the needs, previous experiences, and difficulties experienced by the users of health apps, as well as improvements in usefulness and usability. These and other aspects will be essential to achieve adherence to healthy lifestyles or the adequate management of chronic health conditions.

Limitations

The principal limitation of this work is that, despite the proposal being based on evidence brought by previous successful user experiences in the field of mHealth, the method proposed requires a specific validation to determine its efficacy. However, we have attempted to offer a logical sequence with specific tools and methodologies for each aspect to be evaluated in order to be flexible and open to improvements that permit the creation of a valid development framework for setting up health apps. Although this proposal has managed to encompass the best and most recent scientific evidence available in the mHealth field, other propositions including different tools or sequences could also be valid.

Other limitations are directly linked to the fact that we based the development of a health app on qualitative techniques. As other works have implied [29], by proposing to work with a reduced group of end users, it is necessary to consider that these users may not be an adequate representation of the target population. However, the size of the sample is more or less consistent with the general recommendations that have established that most usability problems can be identified with a smaller number of participants.

In addition, although qualitative methods have habitually used compilation of data in the development and validation of technology-based solutions, there are certain demands in data

analysis that cannot be dealt with from this perspective: identification of the analysis unit, elaboration of constructs, mitigation of the effect of the dynamics within the group, variations between groups, inconsistency of data, etc. To alleviate these problems, the research team should train the focus group leaders in moderation techniques in order to enable them to focus the sessions on the topics related to the issue being addressed in the research [25,43]. However, employing validated questionnaires to measure certain aspects linked to the efficacy of the app developed could also remedy this lack.

As others have pointed out [92], these techniques can be subject to group biases. Nevertheless, they generate a natural open discussion, providing fruitful feedback on usability. For that reason, a design is proposed in which the participants are offered different iterations that are focused on concrete functionalities during the sessions. In any case, this is a natural consequence of agile iterative development.

It is also important to mention that the qualitative sessions take place in meeting rooms behind closed doors and not in real settings, which might modify the behavior of some of the participants. For that reason, as already mentioned, we recommend holding group sessions only once with each participant (except between the third and fourth session so improvements of the app versions can be evaluated). Also, the data collection techniques should only include fieldwork observations, follow-up surveys, and information sessions with moderators [25].

Finally, we would like to emphasize that this research does not aim to evaluate the effectiveness of apps, the methodology of their development, or the practical application. It would be difficult to achieve that objective, given that there is no protocol, proposal, or guideline that can be taken as a reference for the evaluation of the methodologies developed in the reviewed papers. For this reason, this manuscript aimed to provide a structured, user-centered scheme for health app design and development with effectiveness assessed in subsequent mHealth investigations. Nor did we try to identify which user experiences are vital and necessary to the effectiveness of the app itself (which could differ according to population, intervention, or pathology). For this reason, we have proposed a methodology in which these vital elements could be explored.

Despite its limitations, as reported by Peng et al [92], this type of research adds important qualitative evidence in the setting up of mHealth, since it permits access to important information for researchers and app designers, making development of an app with potential for the adoption of healthy lifestyles and improvements in self-care more likely.

Conclusion

This work has proposed a 4-session methodology for the development of health apps, in which aspects such as the difficulties in adopting behavior based on mHealth, acceptability, browsability (ie, the ability to easily browse or navigate through the information offered by the app), usability, and interface study could be studied. These work sessions would be based on predominantly qualitative methodologies (focus and discussion groups); although, in their elaboration, they include validated questionnaires that permit the assessment of objectivity of certain technical aspects. There would be around 10 participants in the groups, with information technology, graphic design, and health care professionals and patients represented. Prior evidence tells us that, in this way, the app's design will be focused on end users, attracting and responding to their needs and, therefore, increasing their adherence to using the app. This would result in positive changes in their attitude toward their health and an increase in their commitment and self-management of the health-disease processes. This proposal requires validation with subsequent measurement of results in terms of adherence and improvement in the clinical variables of the end users, either professionals or patients.

Acknowledgments

GMR participated in all the stages of the research, especially writing the article introduction, extracting results of selected papers, and designing figures. PVP, AJG, and AHR performed the search strategy, paper selection, and writing the material and methods section. MRS and RML reviewed and supervised the entire paper and participated in writing all sections.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sessions and themes with associated quotes.
[DOCX File , 17 KB - mhealth_v8i4e14376_app1.docx]

Multimedia Appendix 2
Session 2: Testing structure and usability.
[PDF File (Adobe PDF File), 1482 KB - mhealth_v8i4e14376_app2.pdf]

Multimedia Appendix 3
Session 3: Does the app fit the needs of final users?

[PDF File (Adobe PDF File), 797 KB - mhealth_v8i4e14376_app3.pdf]

Multimedia Appendix 4
Session 4: Last testing. Keep on improving.
[PDF File (Adobe PDF File), 1022 KB - mhealth_v8i4e14376_app4.pdf]

References

1. Porter J, Huggins CE, Truby H, Collins J. The effect of using mobile technology-based methods that record food or nutrient intake on diabetes control and nutrition outcomes: a systematic review. Nutrients 2016 Dec 17;8(12):815 [FREE Full text] [doi: 10.3390/nu8120815] [Medline: 27999302]
2. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
3. Alonso-Arévalo J, Mirón-Canelo J. Aplicaciones móviles en salud: potencial, normativa de seguridad y regulación. Rev Cuba Inf Cienc Salud 2017;28(3):1-13.
4. Gazdecki A. Bizness Apps. 2017. 9 Mobile Technology Trends For 2017 (Infographic) URL: https://www.biznessapps.com/blog/mobile-technology-trends/ [accessed 2019-04-17]
5. Eurostat. Digital economy and society statistics—households and individuals URL: https://tinyurl.com/ybek6ccg [accessed 2019-04-17]
6. Eurostat. 2016. Mobile connection to internet URL: https://tinyurl.com/rkqxoo5 [accessed 2019-04-17]
7. Eurostat. 2016. Internet use by individuals URL: https://tinyurl.com/j6yvg3r [accessed 2019-04-17]
8. Laxminarayan S, Istepanian RS. UNWIRED E-MED: the next generation of wireless and internet telemedicine systems. IEEE Trans Inf Technol Biomed 2000 Sep;4(3):189-193. [doi: 10.1109/titb.2000.5956074] [Medline: 11026588]
9. Molina Recio G, García-Hernández L, Molina Luque R, Salas-Morera L. The role of interdisciplinary research team in the impact of health apps in health and computer science publications: a systematic review. Biomed Eng Online 2016 Jul 15;15 Suppl 1:77 [FREE Full text] [doi: 10.1186/s12938-016-0185-y] [Medline: 27454164]
10. Torgan C. The mHealth summit: local & global converged. 2009. URL: https://caroltorgan.com/mhealth-summit/ [accessed 2019-04-17]
11. Business Insider. 2017. IQVIA Institute for Human Data Science Study: Impact of digital health grows as innovation, evidence and adoption of mobile health apps accelerate URL: https://tinyurl.com/y7qamjat [accessed 2019-04-17]
12. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
13. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
14. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
15. 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 Jan;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
16. 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]
17. 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]
18. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
19. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
20. Holloway I, Wheeler S. Qualitative Research in Nursing and Healthcare. 3rd Edition. Hoboken: Wiley-Blackwell; 2010.
21. Krueger R, Casey M. Focus Groups: A Practical Guide for Applied Research. 4th Edition. Thousand Oaks: SAGE Publications, Inc; 2008.
22. Shah S, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
23. Lane L, Woloshyn V, Taber N. Tangles, tears and messy conversations: using a media discussion group to explore notions of strong women. J Gender Studies 2016 Oct 12;27(5):497-508. [doi: 10.1080/09589236.2016.1243044]

24. Aliaga FA, Basulto OF, Cabrera J. Rev Investigación Soc. 2012 Dec 31. The discussion group: elements for the investigation around social imaginaries URL: http://www.isdfundacion.org/publicaciones/revista/numeros/9/secciones/tematica/06-grupo-discusion.html [accessed 2020-03-15]
25. Savin-Baden M, Howell-Major C. Focus group interviews. In: Savin-Baden M, Howell-Major C, editors. Qualitative Research: The Essential Guide to Theory and Practice. London: Routledge; 2012.
26. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
27. Pulman A, Taylor J, Galvin K, Masding M. Ideas and enhancements related to mobile applications to support type 1 diabetes. JMIR Mhealth Uhealth 2013;1(2):e12 [FREE Full text] [doi: 10.2196/mhealth.2567] [Medline: 25100684]
28. Greenfield R, Busink E, Wong C, Riboli-Sasco E, Greenfield G, Majeed A, et al. Truck drivers' perceptions on wearable devices and health promotion: a qualitative study. BMC Public Health 2016 Jul 30;16:677 [FREE Full text] [doi: 10.1186/s12889-016-3323-3] [Medline: 27475984]
29. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
30. Cox AM, McKinney P, Goodale P. Food logging: an information literacy perspective. Aslib J Info Manag 2017 Mar 20;69(2):184-200. [doi: 10.1108/AJIM-12-2016-0208]
31. Ploderer B, Brown R, Seng LSD, Lazzarini PA, van Netten JJ. Promoting self-care of diabetic foot ulcers through a mobile phone app: user-centered design and evaluation. JMIR Diabetes 2018 Oct 10;3(4):e10105 [FREE Full text] [doi: 10.2196/10105] [Medline: 30305266]
32. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. A smartphone app to promote an active lifestyle in lower-educated working young adults: development, usability, acceptability, and feasibility study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e44 [FREE Full text] [doi: 10.2196/mhealth.8287] [Medline: 29463491]
33. Lyles CR, Altschuler A, Chawla N, Kowalski C, McQuillan D, Bayliss E, et al. User-centered design of a tablet waiting room tool for complex patients to prioritize discussion topics for primary care visits. JMIR Mhealth Uhealth 2016 Sep 14;4(3):e108 [FREE Full text] [doi: 10.2196/mhealth.6187] [Medline: 27627965]
34. Hao W, Hsu Y, Chen K, Li H, Iqbal U, Nguyen P, et al. LabPush: a pilot study of providing remote clinics with laboratory results via short message service (SMS) in Swaziland, Africa—a qualitative study. Comput Methods Programs Biomed 2015 Jan;118(1):77-83. [doi: 10.1016/j.cmpb.2014.10.005] [Medline: 25453385]
35. Marent B, Henwood F, Darking M. Development of an mHealth platform for HIV care: gathering user perspectives through co-design workshops and interviews. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e184. [doi: 10.2196/mhealth.9856]
36. Duff O, Walsh D, Malone S, McDermott L, Furlong B, O'Connor N, et al. MedFit App, a behavior-changing, theoretically informed mobile app for patient self-management of cardiovascular disease: user-centered development. JMIR Form Res 2018 Apr 27;2(1):e8 [FREE Full text] [doi: 10.2196/formative.9550] [Medline: 30684426]
37. Freyne J, Bradford D, Pocock C, Silvera-Tawil D, Harrap K, Brinkmann S. Developing digital facilitation of assessments in the absence of an interpreter: participatory design and feasibility evaluation with allied health groups. JMIR Form Res 2018 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/formative.8032] [Medline: 30684405]
38. Reblin M, Wu YP, Pok J, Kane L, Colman H, Cohen AL, et al. Development of the electronic social network assessment program using the center for ehealth and wellbeing research roadmap. JMIR Hum Factors 2017 Aug 30;4(3):e23 [FREE Full text] [doi: 10.2196/humanfactors.7845] [Medline: 28855149]
39. Nguyen NH, Hadgraft NT, Moore MM, Rosenberg DE, Lynch C, Reeves MM, et al. A qualitative evaluation of breast cancer survivors' acceptance of and preferences for consumer wearable technology activity trackers. Support Care Cancer 2017 Dec;25(11):3375-3384. [doi: 10.1007/s00520-017-3756-y] [Medline: 28540402]
40. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
41. Velu AV, van Beukering MD, Schaafsma FG, Frings-Dresen MH, Mol BW, van der Post JA, et al. Barriers and facilitators for the use of a medical mobile app to prevent work-related risks in pregnancy: a qualitative analysis. JMIR Res Protoc 2017 Aug 22;6(8):e163 [FREE Full text] [doi: 10.2196/resprot.7224] [Medline: 28830851]
42. Kok DJ. The preventive treatment of recurrent stone-formation: how can we improve compliance in the treatment of patients with recurrent stone disease? Urolithiasis 2016 Feb;44(1):83-90 [FREE Full text] [doi: 10.1007/s00240-015-0842-9] [Medline: 26667826]
43. Campling NC, Pitts DG, Knight PV, Aspinall R. A qualitative analysis of the effectiveness of telehealthcare devices (i) are they meeting the needs of end-users? BMC Health Serv Res 2017 Jul 04;17(1):455 [FREE Full text] [doi: 10.1186/s12913-017-2408-8] [Medline: 28676054]
44. Giunti G, Kool J, Rivera RO, Dorronzoro ZE. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]

45. Shishido HY, Alves da Cruz de Andrade R, Eler GJ. mHealth data collector: an application to collect and report indicators for assessment of cardiometabolic risk. Stud Health Technol Inform 2014;201:425-432. [Medline: 24943577]
46. Casillas J, Goyal A, Bryman J, Alquaddoomi F, Ganz PA, Lidington E, et al. Development of a text messaging system to improve receipt of survivorship care in adolescent and young adult survivors of childhood cancer. J Cancer Surviv 2017 Aug;11(4):505-516 [FREE Full text] [doi: 10.1007/s11764-017-0609-0] [Medline: 28364263]
47. Backonja U, Haynes SC, Kim KK. Data visualizations to support health practitioners' provision of personalized care for patients with cancer and multiple chronic conditions: user-centered design study. JMIR Hum Factors 2018 Oct 16;5(4):e11826 [FREE Full text] [doi: 10.2196/11826] [Medline: 30327290]
48. Nguyen AD, Frensham LJ, Wong MX, Meslin SM, Martin P, Lau AY, et al. mHealth app patient testing and review of educational materials designed for self-management of gout patients: descriptive qualitative studies. JMIR Mhealth Uhealth 2018 Oct 15;6(10):e182 [FREE Full text] [doi: 10.2196/mhealth.9811] [Medline: 30322835]
49. Thirumalai M, Rimmer JH, Johnson G, Wilroy J, Young H, Mehta T, et al. TEAMS (tele-exercise and multiple sclerosis), a tailored telerehabilitation mhealth app: participant-centered development and usability study. JMIR Mhealth Uhealth 2018 May 24;6(5):e10181 [FREE Full text] [doi: 10.2196/10181] [Medline: 29798832]
50. Rivera J, McPherson AC, Hamilton J, Birken C, Coons M, Peters M, et al. User-centered design of a mobile app for weight and health management in adolescents with complex health needs: qualitative study. JMIR Formativ Res 2018 Apr 04;2(1):e7. [doi: 10.2196/formative.8248] [Medline: 30181109]
51. Baek H, Suh J, Kang S, Kang S, Lim TH, Hwang H, et al. Enhancing user experience through user study: design of an mhealth tool for self-management and care engagement of cardiovascular disease patients. JMIR Cardio 2018 Feb 09;2(1):e3 [FREE Full text] [doi: 10.2196/cardio.9000] [Medline: 31758783]
52. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
53. Fledderus M, Schreurs KM, Bohlmeijer ET, Vollenbroek-Hutten MM. Development and pilot evaluation of an online relapse-prevention program based on acceptance and commitment therapy for chronic pain patients. JMIR Hum Factors 2015 Jan 05;2(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.3302] [Medline: 27025550]
54. 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]
55. Ravn Jakobsen P, Hermann A, Søndergaard J, Wiil U, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018 Feb 13;15(2):330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]
56. Bendixen RM, Fairman AD, Karavolis M, Sullivan C, Parmanto B. A user-centered approach: understanding client and caregiver needs and preferences in the development of mhealth apps for self-management. JMIR Mhealth Uhealth 2017 Sep 26;5(9):e141 [FREE Full text] [doi: 10.2196/mhealth.7136] [Medline: 28951378]
57. Smith SA, Whitehead MS, Sheats J, Mastromonico J, Yoo W, Coughlin SS. A community-engaged approach to developing a mobile cancer prevention app: the mCPA study protocol. JMIR Res Protoc 2016 Mar 02;5(1):e34 [FREE Full text] [doi: 10.2196/resprot.5290] [Medline: 26935995]
58. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
59. Bertini E, Gabrielli S, Kimani S, Catarci T, Santucci G. Appropriating and assessing heuristics for mobile computing. Proc Working Conf Adv Visual Interfaces 2006:119-126. [doi: 10.1145/1133265.1133291]
60. Giunti G, Mylonopoulou V, Rivera Romero O. More stamina, a gamified mhealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e51 [FREE Full text] [doi: 10.2196/mhealth.9437] [Medline: 29500159]
61. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
62. Ribu L, Holmen H, Torbjørnsen A, Wahl AK, Grøttland A, Småstuen MC, et al. Low-intensity self-management intervention for persons with type 2 diabetes using a mobile phone-based diabetes diary, with and without health counseling and motivational interviewing: protocol for a randomized controlled trial. JMIR Res Protoc 2013 Aug 26;2(2):e34 [FREE Full text] [doi: 10.2196/resprot.2768] [Medline: 23978690]
63. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996:189-194.
64. Mann D, Riddell L, Lim K, Byrne LK, Nowson C, Rigo M, et al. Mobile phone app aimed at improving iron intake and bioavailability in premenopausal women: a qualitative evaluation. JMIR Mhealth Uhealth 2015 Sep 28;3(3):e92 [FREE Full text] [doi: 10.2196/mhealth.4300] [Medline: 26416479]

65. Akter S, D'Ambra J, Ray P. Development and validation of an instrument to measure user perceived service quality of mHealth. Inf Manag 2013 Jun;50(4):181-195. [doi: 10.1016/j.im.2013.03.001]
66. Haun JN, Chavez M, Nazi K, Antinori N, Melillo C, Cotner BA, et al. Veterans' preferences for exchanging information using veterans affairs health information technologies: focus group results and modeling simulations. J Med Internet Res 2017 Oct 23;19(10):e359 [FREE Full text] [doi: 10.2196/jmir.8614] [Medline: 29061553]
67. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018 Sep 06;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
68. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]
69. Surka S, Edirippulige S, Steyn K, Gaziano T, Puoane T, Levitt N. Evaluating the use of mobile phone technology to enhance cardiovascular disease screening by community health workers. Int J Med Inform 2014 Sep;83(9):648-654 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.008] [Medline: 25002305]
70. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
71. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Stinson JN. A smartphone-based pain management app for adolescents with cancer: establishing system requirements and a pain care algorithm based on literature review, interviews, and consensus. JMIR Res Protoc 2014;3(1):e15 [FREE Full text] [doi: 10.2196/resprot.3041] [Medline: 24646454]
72. 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]
73. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
74. Hirani SP, Rixon L, Beynon M, Cartwright M, Cleanthous S, Selva A, et al. Quantifying beliefs regarding telehealth: development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. J Telemed Telecare 2017 May;23(4):460-469. [doi: 10.1177/1357633X16649531] [Medline: 27224997]
75. Eisenhauer C, Hageman P, Rowland S, Becker B, Barnason S, Pullen C. Acceptability of mHealth technology for self-monitoring eating and activity among rural men. Public Health Nurs 2017 Mar;34(2):138-146 [FREE Full text] [doi: 10.1111/phn.12297] [Medline: 27757986]
76. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
77. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
78. Matthews T. Designing and evaluating glanceable peripheral displays. Proc 6th Conf Designing Interact Syst 2016 Jun:343-345. [doi: 10.1145/1142405.1142457]
79. Consolvo S. Designing for healthy lifestyles: design considerations for mobile technologies to encourage consumer health and wellness. Hum Comput Interact 2012;6(3–4):167-315. [doi: 10.1561/1100000040]
80. Gouveia R, Pereira F, Karapanos E, Munson S, Hassenzahl M. Exploring the design space of glanceable feedback for physical activity trackers. 2016 Presented at: Proc ACM Int Joint Conf Pervasive Ubiquitous Comput; 2016; Heidelberg. [doi: 10.1145/2971648.2971754]
81. Wright CJC, Dietze PM, Crockett B, Lim MSC. Participatory development of MIDY (Mobile Intervention for Drinking in Young people). BMC Public Health 2016;16:184 [FREE Full text] [doi: 10.1186/s12889-016-2876-5] [Medline: 26911299]
82. Morla E. Las enfermedades crónicas no transmisibles. Rev Lat-Am Enfermagem 2020 Mar 03;4(1):3. [doi: 10.22206/cysa.2020.v4i1.pp3-3]
83. Van Dijk MR, Koster MP, Rosman AN, Steegers-Theunissen RP. Opportunities of mHealth in preconception care: preferences and experiences of patients and health care providers and other involved professionals. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e123 [FREE Full text] [doi: 10.2196/mhealth.7834] [Medline: 28818820]
84. Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, Van Rooyen V, Swallow V. Young people's, parents', and professionals' views on required components of mobile apps to support self-management of juvenile arthritis: qualitative study. JMIR Mhealth Uhealth 2018 Jan 19;6(1):e25 [FREE Full text] [doi: 10.2196/mhealth.9179] [Medline: 29351898]
85. Koskinen E, Salminen J. A customizable mobile tool for supporting health behavior interventions. Conf Proc IEEE Eng Med Biol Soc 2007:5908-5911. [doi: 10.1109/IEMBS.2007.4353692] [Medline: 18003358]

86. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
87. Horner GN, Agboola S, Jethwani K, Tan-McGrory A, Lopez L. Designing patient-centered text messaging interventions for increasing physical activity among participants with type 2 diabetes: qualitative results from the text to move intervention. JMIR Mhealth Uhealth 2017 Apr 24;5(4):e54 [FREE Full text] [doi: 10.2196/mhealth.6666] [Medline: 28438728]
88. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
89. Harder H, Holroyd P, Burkinshaw L, Watten P, Zammit C, Harris PR, et al. A user-centred approach to developing bWell, a mobile app for arm and shoulder exercises after breast cancer treatment. J Cancer Surviv 2017 Dec;11(6):732-742 [FREE Full text] [doi: 10.1007/s11764-017-0630-3] [Medline: 28741202]
90. Whealin JM, Jenchura EC, Wong AC, Zulman DM. How veterans with post-traumatic stress disorder and comorbid health conditions utilize ehealth to manage their health care needs: a mixed-methods analysis. J Med Internet Res 2016 Oct 26;18(10):e280 [FREE Full text] [doi: 10.2196/jmir.5594] [Medline: 27784650]
91. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
92. Peng W, Kanthawala S, Yuan S, Hussain S. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

Abbreviations

Health-ITUEM: Health Information Technology Usability Evaluation Model
mHealth: mobile health
SMS: short message service

Edited by G Eysenbach; submitted 22.04.19; peer-reviewed by K Agarwal, M Reynolds, A Nguyen; comments to author 14.09.19; revised version received 05.11.19; accepted 26.01.20; published 22.04.20.

Please cite as:
Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M
Proposal for the User-Centered Design Approach for Health Apps Based on Successful Experiences: Integrative Review
JMIR Mhealth Uhealth 2020;8(4):e14376
URL: http://mhealth.jmir.org/2020/4/e14376/
doi:10.2196/14376
PMID:32319965

Original Paper

The Development of a Mobile Health App for Breast Cancer Self-Management Support in Taiwan: Design Thinking Approach

I-Ching Hou[1*], RN, PhD; Min-Fang Lan[2,3*], RN, MSc; Shan-Hsiang Shen[4], PhD; Pei Yu Tsai[1,3], MSc; King Jen Chang[5], MD, PhD; Hao-Chih Tai[6,7], MD, PhD; Ay-Jen Tsai[5], MSc; Polun Chang[2], PhD; Tze-Fang Wang[1], RN, MSN; Shuh-Jen Sheu[1], RN, PhD; Patricia C Dykes[8,9], RN, PhD

[1]School of Nursing, National Yang-Ming University, Taipei, Taiwan

[2]Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

[3]Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan

[4]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

[5]Taiwan Breast Cancer Foundation, Taipei, Taiwan

[6]Division of Plastic Surgery, Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan

[7]Department of Surgery, College of Medicine, National Taiwan University, Taipei, Taiwan

[8]Center for Patient Safety Research and Practice, Brigham and Women's Hospital, Boston, MA, United States

[9]Department of Medicine, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
I-Ching Hou, RN, PhD
School of Nursing
National Yang-Ming University
No 155, Sec 2
Li-Nong Street, Beitou
Taipei, 11221
Taiwan
Phone: 886 28267000 ext 7315
Email: evita@ym.edu.tw

Abstract

Background: Evidence has shown that breast cancer self-management support from mobile health (mHealth) apps can improve the quality of life of survivors. Although many breast cancer self-management support apps exist, few papers have documented the procedure for the development of a user-friendly app from the patient's perspective.

Objective: This study aimed to investigate the information needs of Taiwanese women with breast cancer to inform the development of a self-management support mHealth app.

Methods: A 5-step design thinking approach, comprising empathy, define, ideate, prototype, and test steps, was used in the focus groups and individual interviews conducted to collect information on the requirements and expectations of Taiwanese women with breast cancer with respect to the app. A thematic analysis was used to identify information needs.

Results: A total of 8 major themes including treatment, physical activity, diet, emotional support, health records, social resources, experience sharing, and expert consultation were identified. Minor themes included the desire to use the app under professional supervision and a trustworthy app manager to ensure the credibility of information.

Conclusions: The strengths of the design thinking approach were user-centered design and cultural sensitivity. The results retrieved from each step contributed to the development of the app and reduction of the gap between end users and developers. An mHealth app that addresses these 8 main themes can facilitate disease self-management for Taiwanese women with breast cancer.

(JMIR Mhealth Uhealth 2020;8(4):e15780) doi:10.2196/15780

KEYWORDS

breast cancer; mobile health application; self-management; design thinking

Introduction

Background

The term *breast cancer* refers to a malignant tumor that develops from cells in the breast. Its stage is usually expressed on a scale of 0 through IV with stage 0 describing noninvasive cancers that remain within their original location and stage IV describing invasive cancers that spread outside the breast to other parts of the body [1]. According to the World Health Organization, the estimated number of newly diagnosed breast cancer cases worldwide was 2,088,849 in 2018 [2], which was the highest among women's cancers. Over 10,000 women are diagnosed with breast cancer every year in Taiwan. Most women are between the ages of 40 and 65 years at the time of diagnosis and usually have breast cancer that is between stages 0 and II [3]. Treatments for breast cancer include surgery, chemotherapy, immunotherapy, radiation therapy, hormonal therapy, and targeted therapy; these are often accompanied by side effects such as pain, fever, diarrhea, lymphedema, dry skin, fatigue, and insomnia [4]. Research indicates that there is compelling evidence of an increased risk of anxiety, depression, suicide, and neurocognitive and sexual dysfunctions among breast cancer survivors when compared with women with no prior incidence of cancer [5-7]. However, when breast cancer is treated appropriately, the 5-year survival rate approaches 88% [8]. Therefore, breast cancer is now treated as a chronic disease and self-management of symptoms is necessary for women to live well with breast cancer. Self-management is "the ability of the individual, in conjunction with family, community, and health care professionals, to manage symptoms, treatments, lifestyle changes, and psychosocial, cultural, and spiritual consequences of health conditions" [9]. Evidence suggests that self-management interventions improve the quality of life of women with breast cancer by enabling self-care [10], mindfulness-based stress reduction [11], and improved management of medical, emotional, and role-based tasks [12].

Today, mobile phones are popular and convenient, with an open architecture that allows third parties to develop mobile health (mHealth) apps. mHealth apps are highly relevant to the future of disease and health management [13]. Research indicates that using mHealth apps to support patients with breast cancer has many advantages including enhanced knowledge, increased physical activity, eliciting short-term reductions in weight, decreased anxiety, improved self-confidence, emotional well-being, and improved quality of life [14-17]. According to one study, there were 599 breast cancer–related apps in the iOS and Android markets in the United States in February 2016. Breast cancer–related apps most commonly addressed disease and treatment information (29.22%), addressed disease management (19.03%), and increased awareness (15.03%) [18].

Objectives

To our knowledge, there are no breast cancer self-management support apps targeting Taiwanese women, and this creates several barriers in the use of existing apps. First, most apps do not provide educational content in traditional Chinese language. The noneffective use of language decreases app acceptance [19]. Second, when software is used to support instant translation, the accuracy and validity of the translation is unknown [20]. Third, the content of most existing apps lacks academic sources and references [18], leading to a lack of trust among Taiwanese women [21]. In addition to these barriers, several cross-cultural comparison studies have indicated that there are important differences in the depressive symptoms and quality of life between eastern and western breast cancer survivors [22-24]. Therefore, the contents of disease self-management support apps should consider cultural differences related to medical resources, food intake, personal beliefs, and support systems of their end users. To address this gap, our research team consisting of health care professionals (IC, MF, and PY), technology experts (SS and PC), and professionals at the Taiwan Breast Cancer Foundation (TBCF; KJ, HC, and IJ) partnered with Taiwanese women with breast cancer to investigate their information needs and to develop an app to meet their needs.

Methods

Study Design

In this study, our team used the design thinking approach. Design thinking is a user-centered problem-solving methodology that starts with assessing people's needs and then seeks innovative solutions to address the range of issues identified. The 5-step design thinking model comprises empathy, define (the problem), ideate, prototype, and test steps and was proposed by the Hasso-Plattner Institute of Design at Stanford (d.school) [25,26]. The concepts underpinning each step and the methodologies we used are summarized in Table 1. The study was approved by the Institutional Review Board (IRB) at National Yang-Ming University (IRB no: YM106005E).

We conducted this research in 2 phases using the 5-step design thinking methodology (Table 1). In phase one, we designed three concurrent 3-hour focus groups in the multifunction room of the TBCF, which included steps 1 through 3. We then entered the second phase of prototyping (step 4) followed by developing and testing (step 5). The research group conducted a data analysis by triangulating the qualitative interview data with our literature review and team discussion to iteratively develop the prototype. The original participants were invited to use the developed prototype, and it was validated further through in-depth interviews.

Table 1. Description of concepts and methodologies of the 5-step design thinking process.

Steps	Description of concepts	Methodologies
1. Empathy	To understand the way women with breast cancer do things and why, their physical and emotional needs, and what is meaningful to them	• Seek to understand • Icebreaker game • Empathy by asking "needs for cancer fighting journey"
2. Define	To bring clarity and focus to the design. The goal is to craft a meaningful and actionable problem statement	• Brainstorm by asking "the needs for the cancer journey" • Write needs in post-it notes • Name and prioritize the needs
3. Ideate	To concentrate on idea generation and get innovative solutions for women with breast cancer	• Sketch the mockups of the app on mobile phone cardboards • Demonstrate how to use the app • Choose the favorite mockups
4. Prototype	To generate the demonstrative solution that can talk to women with breast cancer without investing a lot of time and money	• Prototype system analysis • Design the simulation app with computer-aided software • Install in Android mobile phone
5. Test	To get feedback on the prototype from women with breast cancer and then find the right level of optimization of the prototype and solution	• Provide the simulation app for trial use • Feedback by asking "Tell me your feel to the app" statement

Sampling and Data Collection

We used purposive sampling to enroll Taiwanese women with breast cancer in our study. As suggested by others, we planned to include 5 to 8 members in each focus group [27-29]. Therefore, we planned to recruit 15 to 24 women with breast cancer for 3 concurrent focus groups. The recruitment was posted on TBCF's social networking sites (eg, Facebook and Line) and interested participants contacted us through online registration. A total of 20 women with breast cancer were registered but 5 dropped out because of illness or work obligations that coincided with our study date. The demographic data of the remaining 15 participants are shown in Table 2. Their average age was 55 years, and most women received routine outpatient care for follow-up or long-term hormone therapy. In the beginning of the focus group discussion, study investigators introduced the study, secured informed consent, and encouraged participants to share their ideas with us. In total, 2 female authors (IC and MF) moderated and 3 nurse graduate students joined each subgroup as a facilitator. In addition, 1 nurse graduate student and 2 TBCF staff observed, took photos, and recorded notes. Audio and visual recorders were used to collect data during the focus group discussion.

All participants in phase one were invited to participate in individual interviews. A total of 13 focus group subjects agreed to be interviewed individually. Two subjects did not enroll because of a transportation problem (ID: A5) and one felt that her disease was not severe enough to share her experience (ID: B4). The in-person interviews were held for 1.5 to 2 hours by a study investigator (MF) at a place near the subjects' house or office, or at a café. An audio recorder was used to collect the interview data. All subjects were given US $30 after participating in the research.

Table 2. Demographic data of subjects.

ID	Test step	Age (years)	Education	Marital status	Number of children	Occupation	Cancer stage	Treatment	Current treatment stage
A1	Yes	63	High school	Married	2	Housewife	III	M[a]+C[b]+R[c]+H[d]+T[e]	F/U[f]
A2	Yes	47	High school	Married	1	Electrical industry	I	M+H	F/U with H
A3	Yes	56	Junior college	Married	2	Babysitter retirement	I	M+C+H	F/U with H
A4	Yes	45	Graduate	Married	2	Culture and education industry	II	M+C+T	Under treatment
A5	No	52	High school	Widowed	2	Unemployed	II	M+C+R+H+T	Under treatment
A6	Yes	58	High school	Married	1	Textile industry	III	M+C+R	F/U
B1	Yes	48	College	Widowed	0	Financial industry retirement	III	M+C+R+H	F/U with H
B2	Yes	41	Junior college	Married	1	Culture and education industry	III	M+C+R+H	F/U with H
B3	Yes	44	College	Single	0	Trade	II	M+C+R+H+T	F/U with H
B4	No	59	Junior college	Married	2	Housewife	0	B[g]+R	F/U
B5	Yes	63	College	Married	2	Culture and education industry	II	B+C+R+H+T	F/U with H
C1	Yes	68	College	Married	2	Retired	II	M+C+H	F/U with H
C2	Yes	61	College	Widowed	3	Department store clerk	III	M+C+R+H	F/U
C3	Yes	62	Junior college	Married	2	Retired	II	M+C+R	F/U
C4	Yes	63	Junior college	Married	1	Housewife	III	M+C+R+H	F/U with H

[a]M: modified radical mastectomy.

[b]C: chemotherapy.

[c]R: radiotherapy.

[d]H: hormone therapy.

[e]T: target therapy.

[f]F/U: follow-up.

[g]B: breast-conserving surgery.

Procedure: Concepts and Study Activities

The concepts and study activities that were completed over 2 phases using the 5-step design thinking methodology are described below.

Phase One: Focus Group Approach

Step 1: Empathy

Concept

Empathy involves the work to understand and gain insight into people's thoughts and needs, within the context of the design challenge. The designers' goal is to understand how people do things and why they do so, their physical and emotional needs, how they think about their world, and what is meaningful to them [25].

Study Activity

To understand breast cancer and the characteristics of Taiwanese women with breast cancer better, our team conducted a review of the literature and existing breast cancer apps. We also conducted interviews with the TBCF staff. In the beginning of the focus group discussion, each subject was asked to illustrate their own appearance and write their name, nickname, date of diagnosis, current treatment status, and their current mood on the card (Figure 1). The subjects then used the card to introduce themselves within each subgroup. This activity provided an opportunity for investigators to observe, watch, and listen to the subjects' experiences at different stages of breast cancer treatment.

Figure 1. Self-instruction card for icebreaker. Demo card (left) and a participant's card (right).

Step 2: Define

Concept

The define step in the design process brings clarity and focus to the design space. The goal is to craft a meaningful and actionable problem statement [25].

Study Activity

After a debriefing, subjects were asked to brainstorm on one question: "How may we use the mHealth app to support you through your cancer fighting journey?" Each need was written on a post-it note, and then notes with similar needs were arranged into the same column on a large poster. To ideate in response to information needs, each subgroup was asked to name the needs with notes similar to those created in the define step. Then, each participant of the subgroup was asked to rank the importance of their information needs using Arabic numbers (eg, 1 was the most important, 2 was the second most important, and so on). The facilitator of each group summarized the total score of each need type and its relative rank of importance (Figure 2).

Figure 2. Names and prioritization of needs from subgroup B.

Step 3: Ideate

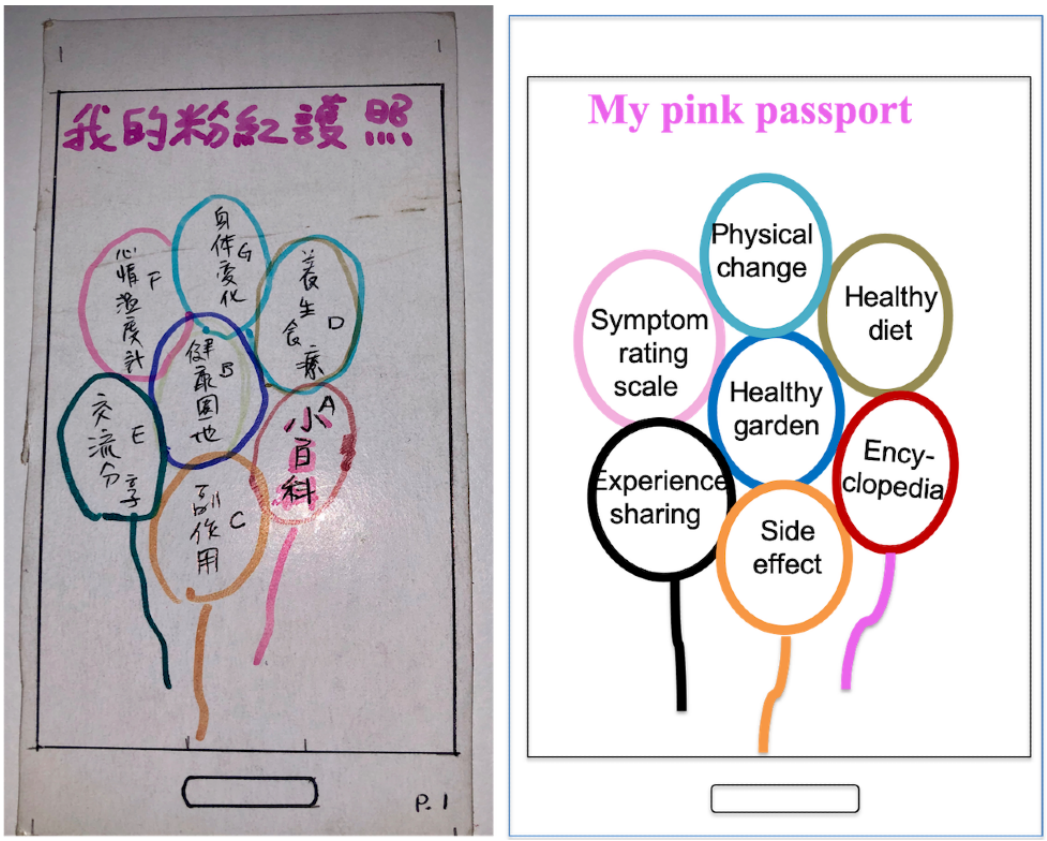
Concept

The ideate step in the design process concentrates on idea generation. Ideation provides the fuel and is the source for building prototypes and putting innovative solutions in the hands of users [25].

Study Activity

According to the prioritization determined in step 2, each subgroup was asked to sketch an mHealth app interface on 6.1-inch smartphone cardboard mockups using colored pens (Figure 3). At the end of the focus group discussion, the 3 subgroups were invited to demonstrate how they would use these mockups for disease self-management. After the demonstration, the participants were asked to choose their favorites. The mockups of subgroup A were preferred by most participants and were therefore used as a reference in the prototype step (step 4).

Figure 3. Mockups of the app from subgroup A (partial example). The original mockup menu of the app (left) and the same redrawn in English (right).



Phase Two: Individual Operation and Interview

Step 4: Prototype

Concept

The prototype step involves the iterative generation of artifacts intended to answer questions that get closer to a final solution. Through prototyping, designers can talk to users, resolve differences, reduce poor communication, and test ideas without investing a lot of time and money in programming [25].

Study Activity

According to the results from steps 1 through 3, nursing informatics graduate student investigators (MF and PY) with basic mobile app programming skills, used the JustInMind Prototyper tool (JustInMind) to design the mHealth app simulation in the 6.1-inch, 1280×720-pixel touch screen (L×W×H: 161.5×84.5×9.3 mm) and an Android 4.2.2 Jelly Bean (Google LLC) mobile phone to be used as the tool in step 5 (Figure 4).

Figure 4. Prototype of the simulation app (partial example).

Step 5: Test

Concept

The test step aims to solicit feedback through the demonstration of the prototype to end users. This helps identify opportunities for the improvement and optimization of the prototype and solution [25].

Study Activity

To obtain participants' feedback, individual interviews were held to evaluate the simulated breast cancer mHealth app. A study investigator (MF) showed the simulation to the participants and encouraged them to use it for about 15 min. The open-ended question, "Tell me your recommendations for each function in this simulated app," was asked to collect feedback from each participant. An audio recorder was used during individual interviews, and the recordings were transcribed within 48 hours. A guideline for transcription was used to avoid inconsistency in transcript styles. The transcripts were emailed to each participant to confirm content accuracy.

Data Analysis

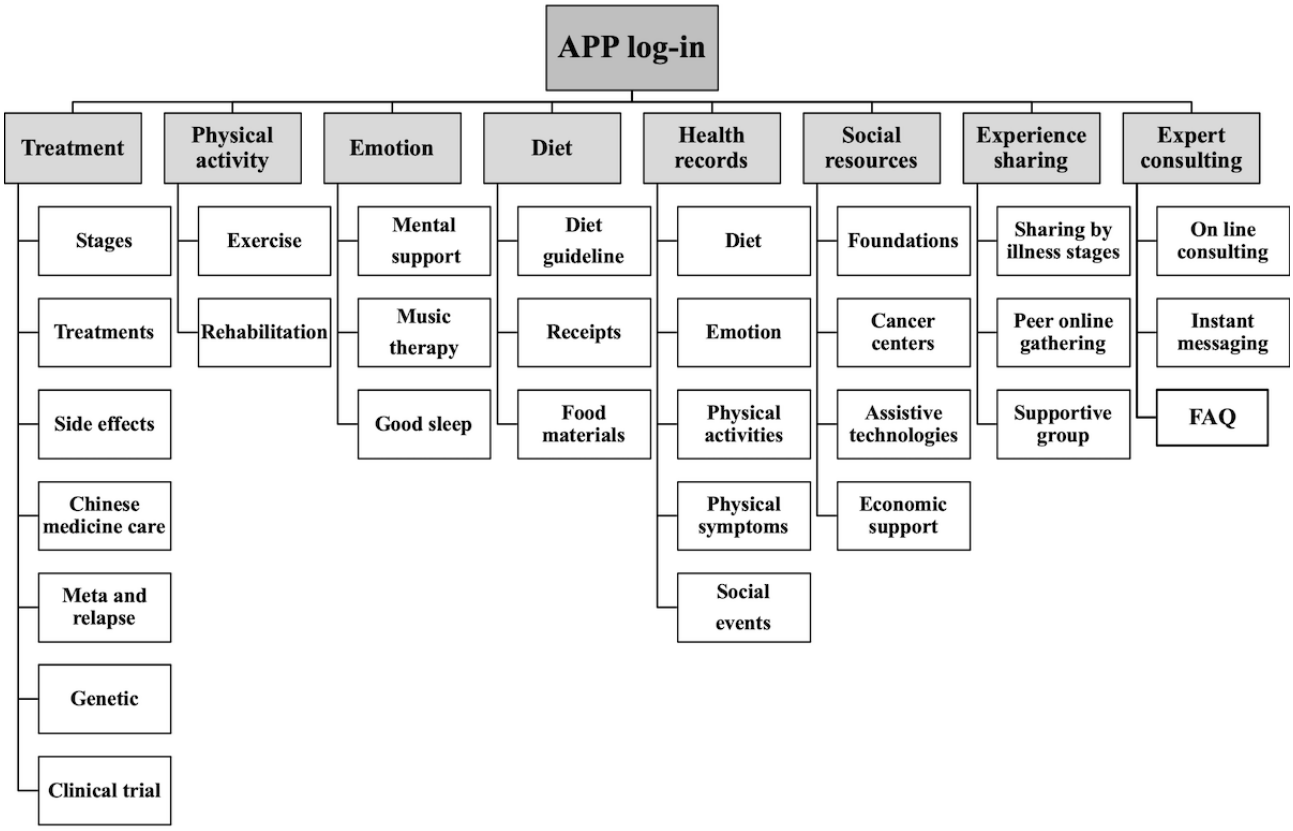
We adopted Norwell's 6-phase methodology for a trustworthy thematic analysis to summarize our subjects' information needs [30]. The first involved becoming familiar with the data. Subjects' notes, naming of information needs, prototype cards, and individual interview transcripts were transcribed into an Excel spreadsheet. The text was read several times by 2 coders (MF and IC) to familiarize themselves with the data and to confirm accuracy. The second was generating initial codes. As the mockups from subgroup A were most preferred by participants, their naming of information needs was used as the initial codes to help guide the analysis. Texts that were similar to the initial codes were merged with those codes. Texts that could not be categorized into existing codes were formed into new ones based on their meanings. Two coders checked and finalized all the codes according to the semantic meaning to validate them, and to eliminate redundant ones. The third was searching for themes. The final codes were extracted for each theme to consider whether they formed a coherent pattern. The fourth involved reviewing themes. All themes were vetted during research team meetings. The fifth entailed defining and naming themes. The consensus among research team members was used to define and name each theme. The sixth involved producing the report. The Consolidated Criteria for Reporting Qualitative Research reporting guidelines [31] were used to produce this report. This process continued until the saturation of main themes was attained [32,33]. No new themes were found both after using this process and after 30 research team discussions were conducted over a year. Multimedia Appendix 1 shows an example of "diet" as a theme and the related codes that emerged from the thematic analysis in this study [34].

Results

Thematic Analysis

After the thematic analysis, 8 major themes and 30 codes were identified. The framework is shown in Figure 5. The definitions of each theme and their codes are described below.

Figure 5. Framework of Breast Cancer Self-Management Support Mobile Health app.



Theme 1: Treatment

Treatment was defined as information needs related to breast cancer treatment. A total of 31 notes (eg, "treatment protocol" and "the side effects of each treatment") from the define step and 10 recommendations from the test step were categorized into the treatment theme. The codes included treatment stages, side effects, Chinese medicine care, metastases, and relapse, and genetic and clinical drug trials. Our subjects wanted more information on the treatment and side effects through the app:

> *We have one list, there are many (chemotherapy)drugs on it. (It tells) which one would make you lose hair and which one will not. I think it was very helpful and I had posted this list on our group four to five times.* [B5, 63-year-old, stage II]

The Chinese medication care code received the most feedback from the test step. Most subjects wanted to know which Chinese herbs were safe, when to receive Chinese medication support and care, and information on physicians qualified in Chinese medicine. A few subjects did not think that Chinese medicine care should be included in the app because they believed that it may interfere with formal treatment:

> *About the Chinese medication, I felt it was helpful, they can help you to relieve the uncomfortable symptoms. But you need to determine which Chinese medicine physician was good for you.* [C4, 63-year-old, stage III]

> *I believe in the Chinese medication, but my oncology physician told me not to take them because he thought the herbs might interfere with the hormone therapy.* [A6, 58-year-old, stage III]

Theme 2: Physical Activity

Physical activity was defined as the appropriate body movement and intensity for post–breast cancer treatment. A total of 24 notes (eg, "exercise," "dancing," "rehabilitation," and "Qi Gong") from the define step and 20 recommendations from the test step were categorized into the physical activity theme. The codes included exercise and rehabilitation. Most subjects identified preferred exercises and societies (eg, TBCF, Formosa Cancer Foundation, and Taiwan Cancer New Life Association) or private fitness firms (eg, Curves) for these exercises. Some recommended contacting a Taiwanese physical therapist to get his demonstration video for breast cancer exercises for use through the app.

We noted that some subjects described schedule conflicts with desired exercise activities. Others said that they would not go out during chemotherapy for fear of infection:

> *I would like to participant (Infinite Youth), but the time was too early to participant.* [A6, 58-year-old, stage III]

> *I would not want to participant in the course because I am afraid of being infected from the close space.* [B1, 48-year-old, stage III]

Theme 3: Emotion

Emotion was defined as psychological support during breast cancer treatment. A total of 20 notes (eg, "emotion," "mental

support," and "entertainment") from the define step and 11 recommendations from the test step were categorized into the emotion theme. The codes included mental support, music therapy, and sleeping well. For mental support, there were some short encouraging statements added to the simulation app. Some subjects expressed that they wanted to keep these statements in their mobile phone for easy access and use them for spiritual sustenance:

> *Can I download it here? Then I will save as my (mobile phone) desktop.* [A2, 47-year-old, stage I]

All participants believed that music in the simulation app could relieve stress related to their disease. Some subjects were choristers of TBCF. They recommended using their songs in the future app to help other patients who could not routinely participate in the chorus because of their geographical location:

> *The songs of the TBCF chorus could be provided (in this App) and let other patients listen to them. In fact, we hope that the patients during the treatment can come out (eg, participate the chorus). Some of sisters lived too far away to participant (the chorus). You cannot ask the them to come only for 1-hour chorus activity because it takes 2 to 3 hours in traffic to get there (eg, TBCF multifunction room).* [C2, 61-year-old, stage III]

To sleep well, the subjects shared that mindfulness, acupuncture point massage, and the pressure release activity were helpful:

> *I download the mindfulness activities in my phone. They were very helpful for me to practice.* [C3, 62-year-old, stage II]

Theme 4: Diet

Diet was defined as nutrition support and the norm for healthy foods for post–breast cancer treatment. A total of 18 notes (eg, "diet," "diet guideline," "gourmet food," and "food and nutrition") from the define step and 14 recommendations from the test step were categorized into the diet theme. The codes included diet guidelines, recipes, and ingredients. Most subjects reported that they could not prepare food by themselves during cancer therapy because of physical weakness. They were also concerned about what they can or cannot eat during treatment. Therefore, they recommended the inclusion of recipes from specific publications and recommended a recipe-sharing function on the app. The subjects also requested a convenient and reliable food ordering and delivery function on the app:

> *My sister prepared food for me. I was very weak during treatment.* [A2, 47-year-old, stage I]
>
> *Can I eat the beef? Some sister told me not to eat (beef).* [A6, 58-year-old, stage III]
>
> *It is good to share their own recipes.* [B1, 48-year-old, stage III]
>
> *It was convenient that the food can be delivered home when it is not convenient to go out during chemotherapy.* [A1, 63-year-old, stage III]

Theme 5: Health Records

Health records were defined as a diary of self-reported information related to breast cancer treatment and self-management. A total of 18 notes (eg, "body weight record," "body temperature record," and "arm circumference") from the define step and 25 recommendations from the test step were categorized into the health records theme. The codes included the food diary, emotions, physical activity, physical symptoms, and social events. For the food diary, some subjects mentioned that they had participated in the nutrition training course. The nutritionist had asked them to record their daily diet. Therefore, it was helpful for them to have these functions on the app. However, they looked forward to feedback from the nutritionist after recording their diet and uploading photos:

> *It was ok to upload the photo (to the nutritionist) but you need to have nutritionist.* [A1, 63-year-old, stage III]

The subjects shared that they felt negative emotions during their treatment and that they wanted to record their moods on the app. However, some believed that writing about their mood without reflection was insufficient. Others expressed that by recording their emotions on the app, patients could reflect on their negative thoughts, recover from bad moods, and remain focused on positive thinking:

> *Only record the emotion is not enough. You do not make the reflection. You shouldpractice self-conversation while entry the emotion records. [C3, 62-year-old, stage II]*

As for physical activity records (eg, heart rate, time, frequency, reminders), half the subjects felt that it would be difficult to remember to record their data without an auto-recording device (eg, an intelligent wearable device with a heart rate monitor). They also thought that it would be helpful for the app to remind them to engage in some physical activity if they had not done so for a while:

> *Generally, I can't know so many (data) without using auto (monitor)device to record.* [A2, 47-year-old, stage I]
>
> *I hope the App could remind me if I do not exercise for a long time. Sometime people would be lazy when stay at home.* [A4, 45-year-old, stage II]

As for the physical symptoms record (eg, body weight with BMI, temperature, upper arm circumference, symptoms, defecation, sleeping time, and menstrual period), our subjects reported that they had already recorded their body weight every day and appreciated help in the calculation of their auto BMI. One of the subjects hoped that the app could help them control their body weight. Some reported checking their body temperature at home for fear of fever. None of the participants were used to measuring their upper arm circumference. Participants also reported that they had other symptoms and had asked for app functionality that allowed them to record other symptoms that they were having (eg, diarrhea and paronychia). Recording sleep duration was not perceived as helpful in measuring the quality of sleep. Some subjects who were experiencing menopause felt that the menstrual period

record was not necessary for them. However, younger patients felt that it was helpful to show their physician this record with the notes, as it helped predict the next date of menstruation:

> *I am afraid that my BW was increased after checking it. Hopefully, it can have the function to tell us how to control BW such as eating fewer starchy foods.* [A1, 63-year-old, stage III]
>
> *Sleep record was not necessary. Some will have more anxiety because they are not sleeping well.* [C4, 63-year-old, stage III]
>
> *To predict and record the next menstruation so we don't need to use another application.* [B3, 44-year-old, stage II]

As for social event records (eg, date of treatment, clinic, examination, and social events), some subjects requested that we expand the social record to include a wider range of social events (eg, dancing and gardening society membership).

Theme 6: Social Resources

Social resources was defined as social or economic support during breast cancer treatment. A total of 13 notes (eg, "resource," "underwear tailor," and "cancer foundation") from the define step and 11 recommendations from the test step were categorized into the social resources theme. The codes included foundations, cancer centers, assistive technologies, and economic support. Most subjects had experience joining breast cancer–related foundations in Taiwan. Some of them reported serving as volunteers and getting more information from these foundations. Only a few reported having experience contacting the cancer center at a hospital. When it came to assistive technologies, they also shared information such as how to get a free wig, hat, and compression sleeve. Three subjects also mentioned information on economic support. They expressed their desire to have information on how to access these types of resources:

> *I volunteer at three foundations. Being a volunteer made me better understand their resources.* [A3, 56-year-old, stage I]
>
> *Are there cancer centers in the north, middle, south, east Taiwan?* [B1, 46-year-old, stage III]
>
> *You can add the information (eg, to the app) for where to borrow a wig. There are many wigs at the Hope Foundation, and they have many different stylish wigs.* [C4, 63-year-old, stage III]
>
> *When one of the parents had cancer, their family could apply for the financial support. However, no patient knows this information.* [A3, 56-year-old, stage I]

Theme 7: Experience Sharing

Experience sharing was defined as peer support among women with breast cancer. A total of 10 notes (eg, "experience sharing," "peer online gathering," and "peer support") from the define step and 3 recommendations from the test step were categorized into the experience sharing theme. The codes included experience sharing by illness stage, peer online gathering, and support groups. Every subject supported the function of experience sharing videos or online gathering on the app. For the experience sharing video, they preferred an audio recording rather than an actual one. For the online gathering, they used existing social network apps (eg, Line and Facebook). They identified management issues that should be addressed, such as preventing negative or false information. For the support group, all subjects agreed that they wanted to offer encouragement to other patients:

> *You can record the voice of your emotion when diagnosed with breast cancer without showing your face.* [A1, 63-year-old, stage III]
>
> *Everyone is using Line, Facebook, it will not be ok to use on line gathering only in this App except you can link to Line, Facebook. In addition, the on line gathering without management should avoid for someone had negative emotion or false information.* [A4, 45-year-old, stage II]

Theme 8: Expert Consulting

Expert consulting was defined as expert support during breast cancer treatment. A total of 5 notes (eg, "consulting," "online clinic for wound," and "professional team online response") from the define step and 7 recommendations from the test step were categorized into the social resources theme.

The codes included professional online consulting, instant messaging, and frequently asked questions. For professional online consulting, they believed that questions could not be answered in real time because of the workload of the professionals. Some subjects felt that it would decrease their workload if the questions were categorized as frequently asked questions:

> *I am wondering the professionals behind the App will answer the same question frequently? If frequently asked questions could be read somewhere, it will decrease the percentage of repeated questions and also decrease the workload of the professionals.* [A4, 45-year-old, stage II]

Discussion

Principal Findings and the Differences From Previous Studies

According to our 5-step design thinking approach and thematic analysis, a total of 8 themes that included treatment, physical activity, emotions, diet, health records, social resources, experience sharing, and expert consulting were retrieved. In each theme, there were multiple codes that consisted of the information needs of the end users of the app. Previous research has indicated that the information needs of women with breast cancer included knowledge of the disease, the impact of breast cancer on the body, cancer metastasis or relapse, understanding and preparing for treatment, preventing and facing side effects and risk, survival prediction, survival rate, consultations, follow-up schedule, diet suggestion and restrictions, body image changes, self-management, and mental support [35-38]. Four of our main themes and the codes (eg, treatment, diet, emotional support, and health records) matched findings from prior

research. In this study, our subjects did not mention the information needs of survival prediction and survival rate. It may have been that our subjects were mostly current breast cancer victims and were focused on healthy lifestyles to prevent cancer recurrence [39].

Previous research has also indicated that physical activity including exercise and physical therapy was helpful in lymphedema prevention, postoperative pain relief, and body weight control to prevent cancer relapse [39,40]. Our subjects expressed a positive attitude toward participation in group physical activities. However, in some cases, low immune tolerance and meeting schedules limited participation. Therefore, including some e-learning exercise courses on the app to enhance flexibility should be considered.

With regard to the theme of diet, our subjects' perceptions were consistent with findings from previous studies that suggested that women with breast cancer were concerned about healthy eating, food, and nutrition-related side effects of chemotherapy [41-43]. The incorporation of culturally appropriate Taiwanese recipes for women with breast cancer and related nutritional information in the simulation app was positively received by most subjects. However, subjects also felt that it would be difficult to cook by themselves during treatment. Thus, offering some convenient and reliable food ordering and delivery functions on the app was also considered helpful. Diet and physical activity recommendations could help those who had completed treatment control their body weight. Participants requested more advanced and useful body weight control functions within the app such as food calorie or energy burning calculators.

We found that some subjects kept their own health records (eg, recoding body weight and temperature). The simulation app offered many structured health record options but based on the subjects' feedback, it was difficult for us to assess whether Taiwanese women with breast cancer wanted to keep these records on their own in the future. According to a previous study based on a survey of oncology patients on app-assisted cancer care, the introduction of mobile apps needs to follow different strategies depending on the patients' attitudes and apps could support clinic visits, document adverse effects, and provide reminders of treatment dates or medication schedules [44].

In summary, 8 main themes were identified by the research team. The information needs based on these main themes and their codes were used to form the framework for a breast cancer self-management support mHealth app. Most themes were consistent with previous studies but some new themes specific to Chinese culture were also uncovered. We believe that the design thinking approach is a strong method for identifying the information needs for Taiwanese women with breast cancer for the development of mHealth technology.

Minor Findings

From the analysis, two minor findings emerged. The first was that the app should be used under the supervision of a professional. Taking physical activity for example, we provided three levels of rehabilitation actions to help women with breast cancer recover from surgery and to prevent lymphedema. However, they did not understand when to do these actions. Such patients would like to consult a professional rather than use an app. They also wanted the nutritionist to review their daily food records and prescribe the correct diet based on their assessment. We concluded that professional support while using the app would make end users more confident in the benefits of using the mHealth app.

The second minor finding was related to the management of the app. Expert consulting was seen as important for supporting women with breast cancer; however, such services will require a significant time commitment from professionals. Our subjects were concerned about the workload of these professionals and recommended that we offer frequently asked questions instead. Moreover, women with breast cancer wanted to share their experiences on online forums and also recommended using an app manager to ensure the credibility of information. These two minor findings can facilitate end user trust in the app.

Design Thinking Approach in Developing a Comprehensive and Culturally Sensitive App

The 5-step design thinking approach allowed users to brainstorm, design mockups, test, and provide feedback before the actual breast cancer–related self-management support app was developed. Through this approach, the app developer can identify new app requirements and compare end user requirements with those addressed in existing apps. In the ideate step, our participants sketched several balloons with different function labels as the main menu of the app (Figure 3) on the smartphone prototype. In eastern culture, people use a circle to represent the notion that everything is copasetic. Women with breast cancer may use the notion to pray for a smooth process during their cancer treatment. The design thinking approach and use of the illustration activity helped discover the user's interface preferences and increased the programmer's sensitivity to their needs.

In this study, the women with breast cancer had some information needs related to traditional Chinese medicine (TCM). The role of TCM was viewed as supportive medicine in the treatment of breast cancer. TCM is commonly used in eastern countries (eg, in Chinese culture, people often use Chinese herbs such as *Angelica sinensis*, *Fructus lycii*, pilosulae, tuceahoe, and shitake to enhance immunity) [45,46]. We were unable to find references or existing breast cancer–related self-management support apps that provide evidence of the effects of TCM treatment. Consistent with previous Chinese studies [45,46], most of our subjects wanted to know how TCM could help them during treatment. However, most women with breast cancer in Taiwan are undergoing treatment in hospitals (eg, the western medicine approach) and most of them are told by their physicians that TCM may interfere with western medicine treatment (eg, cause drug-drug interactions). To prevent confusion, a few subjects rejected the idea of including TCM information in the app. For such information needs, our team plans to consult with professional TCM physicians in the future. Our study did show that the design thinking approach is adequate to develop a user-centered and culturally sensitive support app for patients with cancer in a different social context.

Limitations

One limitation of this study was that the sample size was inadequate in offering perspectives from women across all stages of breast cancer (eg, initial diagnosis through stage IV). Our study was conducted in an urban area with more medical resources than are typically available in the more rural areas of the country, and this may result in a geographical bias. Our testing was done using a simulation app rather than the real app, which may result in different feedback.

Summary and Conclusions

Our team used the design thinking process to retrieve information needs related to the use of an mHealth app by women with breast cancer in Taiwan. The needs were retrieved using one focus group with three subgroups that included brainstorming, discussion, and validation. The interactive simulation app coupled with individual interviews helped us identify content and begin prototyping before designing the real app. A total of 8 themes with multiple codes consisting of common and culturally specific information needs provided the framework for the self-management support mHealth app that we developed for Taiwanese women with breast cancer.

Acknowledgments

This study was funded by the National Taiwan University of Science and Technology and National Yang Ming University Joint Research Program (NTUST-NYMU; 2017-2018: Developing personal mobile apps for women with breast cancer, NTUST-NYMU-106-04). The authors thank all team members engaged in this research and the TBCF.

Authors' Contributions

IC and MF contributed to the study design and implantation, analysis and interpretation of the findings, and preparation of the manuscript. SH, PY, KJ, HC, and AJ contributed toward the implantation of the study. PL, TF, and SJ contributed toward the study recommendations. PC contributed toward editing and revising the English version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis of information needs framework (example).
[DOCX File , 16 KB - mhealth_v8i4e15780_app1.docx]

References

1. Breast Cancer Information and Support. What Is Breast Cancer? URL: https://www.breastcancer.org/symptoms/understand_bc/what_is_bc [accessed 2018-08-27]
2. World Health Organization. Cancer Today. 2018. Estimated Number of New Cases in 2018, Worldwide, Both Sexes, All Ages URL: https://tinyurl.com/ugemlbs [accessed 2019-07-08]
3. Health Promotion Administration Ministry of Health and Welfare. Cancer Registration Report URL: https://www.hpa.gov.tw/Pages/List.aspx?nodeid=119 [accessed 2019-10-11]
4. Breast Cancer Information and Support. Treatment Side Effects URL: https://www.breastcancer.org/treatment/side_effects [accessed 2020-02-14]
5. Carreira H, Williams R, Müller M, Harewood R, Stanway S, Bhaskaran K. Associations between breast cancer survivorship and adverse mental health outcomes: a systematic review. J Natl Cancer Inst 2018 Dec 1;110(12):1311-1327 [FREE Full text] [doi: 10.1093/jnci/djy177] [Medline: 30403799]
6. Jacob L, Kalder M, Kostev K. Incidence of depression and anxiety among women newly diagnosed with breast or genital organ cancer in Germany. Psychooncology 2017 Oct;26(10):1535-1540. [doi: 10.1002/pon.4328] [Medline: 27897353]
7. Patoo M, Moradi AR, Payandeh M. P0134 Relationship between depression, anxiety, and quality of life in women with breast cancer. Eur J Cancer 2014;50:e47. [doi: 10.1016/j.ejca.2014.03.178]
8. Taiwan Cancer Registry. Five-year Relative Survival Rate for Cancer (Including Invasive Cancer Only) URL: http://tcr.cph.ntu.edu.tw/main.php?Page=A5B3 [accessed 2019-08-21]
9. Richard AA, Shea K. Delineation of self-care and associated concepts. J Nurs Scholarsh 2011 Sep;43(3):255-264. [doi: 10.1111/j.1547-5069.2011.01404.x] [Medline: 21884371]
10. van Dijck S, Nelissen P, Verbelen H, Tjalma W, Gebruers N. The effects of physical self-management on quality of life in breast cancer patients: A systematic review. Breast 2016 Aug;28:20-28. [doi: 10.1016/j.breast.2016.04.010] [Medline: 27173849]
11. Eyles C, Leydon GM, Hoffman CJ, Copson ER, Prescott P, Chorozoglou M, et al. Mindfulness for the self-management of fatigue, anxiety, and depression in women with metastatic breast cancer: a mixed methods feasibility study. Integr Cancer Ther 2015 Jan;14(1):42-56 [FREE Full text] [doi: 10.1177/1534735414546567] [Medline: 25161198]

12. Loh SY, Packer T, Chinna K, Quek KF. Effectiveness of a patient self-management programme for breast cancer as a chronic illness: a non-randomised controlled clinical trial. J Cancer Surviv 2013 Sep;7(3):331-342. [doi: 10.1007/s11764-013-0274-x] [Medline: 23519654]
13. Yu LP. The development of the mobile medical industry and the opportunities in Taiwan - from the US mobile medical application guidelines. Reg Med News 2014;42:9-22 [FREE Full text]
14. Zhu J, Ebert L, Guo D, Yang S, Han Q, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 1): qualitative study of women's perceptions. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e85 [FREE Full text] [doi: 10.2196/mhealth.9311] [Medline: 29643056]
15. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
16. Pope Z, Lee JE, Zeng N, Lee HY, Gao Z. Feasibility of smartphone application and social media intervention on breast cancer survivors' health outcomes. Transl Behav Med 2019 Jan 1;9(1):11-22. [doi: 10.1093/tbm/iby002] [Medline: 29471477]
17. McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]
18. Giunti G, Giunta D, Guisado-Fernandez E, Bender J, Fernandez-Luque L. A biopsy of Breast Cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
19. mHIMSS App Usability Work Group. Amazon Web Services. 2012 Jul. Selecting a Mobile App: Evaluating the Usability of Medical Applications URL: http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSguidetoappusabilityv1mHIMSS.pdf [accessed 2019-12-17]
20. Moshi MR, Tooher R, Merlin T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: a systematic review. Int J Technol Assess Health Care 2018 Jan;34(5):464-475. [doi: 10.1017/S026646231800051X] [Medline: 30201060]
21. Team HON. Health On the Net. 2020 Feb. The 8 Principles of the HONCODE APP Certification URL: https://www.hon.ch/en/certification/app-certification-en.html [accessed 2020-02-14]
22. Milbury K, Kavanagh A, Meng Z, Chen Z, Chandwani KD, Garcia K, et al. Depressive symptoms and positive affect in Chinese and United States breast cancer survivors: a cross-cultural comparison. Support Care Cancer 2017 Jul;25(7):2103-2109 [FREE Full text] [doi: 10.1007/s00520-017-3612-0] [Medline: 28233122]
23. Wen K, Fang CY, Ma GX. Breast cancer experience and survivorship among Asian Americans: a systematic review. J Cancer Surviv 2014 Mar;8(1):94-107 [FREE Full text] [doi: 10.1007/s11764-013-0320-8] [Medline: 24214498]
24. Fischer M, Inoue K, Matsuda A, Kroep JR, Nagai S, Tozuka K, et al. Cross-cultural comparison of breast cancer patients' Quality of Life in the Netherlands and Japan. Breast Cancer Res Treat 2017 Nov;166(2):459-471 [FREE Full text] [doi: 10.1007/s10549-017-4417-z] [Medline: 28762012]
25. Stanford d.school. 2010. An Introduction to Design Thinking PROCESS GUIDE URL: https://dschool-old.stanford.edu/sandbox/groups/designresources/wiki/36873/attachments/74b3d/ModeGuideBOOTCAMP2010L.pdf [accessed 2018-12-16]
26. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
27. Morgan DL, McCracken G. Focus Groups As Qualitative Research. Thousand Oaks, CA: SAGE Publications, Inc; 1996.
28. Krueger RA. Semantic Scholar. 2002 Oct. Designing and Conducting Focus Group Interviews URL: https://pdfs.semanticscholar.org/eb74/99b9e559c969d0a8e7bceafd36adc4578eaf.pdf?_ga=2.125119498.1602236450.1581740714-110518158.1581740714 [accessed 2019-12-01]
29. Lee T, Sheu S, Chang H, Hung Y, Tseng L, Chou S, et al. Developing a web-based comic for newly diagnosed women with breast cancer: an action research approach. J Med Internet Res 2019 Feb 4;21(2):e10716 [FREE Full text] [doi: 10.2196/10716] [Medline: 30714947]
30. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct;16(1). [doi: 10.1177/1609406917733847]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Ando H, Cousins R, Young C. Achieving saturation in thematic analysis: development and refinement of a codebook. Compr Psychol 2014;3:03.CP.3.4. [doi: 10.2466/03.cp.3.4]
33. Glaser BG, Strauss AL. Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Routledge; 1967.
34. Forman HJ, Davies KJ, Ursini F. How do nutritional antioxidants really work: nucleophilic tone and para-hormesis versus free radical scavenging in vivo. Free Radic Biol Med 2014 Jan;66:24-35 [FREE Full text] [doi: 10.1016/j.freeradbiomed.2013.05.045] [Medline: 23747930]

35. Alba-Ruiz R, Bermúdez-Tamayo C, Pernett JJ, Garcia-Gutierrez JF, Cózar-Olmo JM, Valero-Aguilera B. Adapting the content of cancer web sites to the information needs of patients: reliability and readability. Telemed J E Health 2013 Dec;19(12):956-966 [FREE Full text] [doi: 10.1089/tmj.2013.0050] [Medline: 24073899]
36. Lei CP, Har YC, Abdullah KL. Informational needs of breast cancer patients on chemotherapy: differences between patients' and nurses' perceptions. Asian Pac J Cancer Prev 2011;12(3):797-802 [FREE Full text] [Medline: 21627386]
37. Liou Y, Zhao TT, Yuan CJ. Qualitative study on information needs of different treatment stages for breast cancer patients. Nursing Journal Of Chinese Peoples Liberation Army 2016 Mar;33(6):31-34. [doi: 10.3969/j.issn.1008-9993.2016.06.007]
38. Jeang SR, Chi MT, Wang PH, Ku YC. Process of Seeking Treatment for Newly Diagnosed Breast Cancer. VGH Nursing 2015 Mar;32(1):10-20. [doi: 10.6142/vghn.32.1.10]
39. Hayes SC, Johansson K, Stout NL, Prosnitz R, Armer JM, Gabram S, et al. Upper-body morbidity after breast cancer: incidence and evidence for evaluation, prevention, and management within a prospective surveillance model of care. Cancer 2012 Apr 15;118(8 Suppl):2237-2249 [FREE Full text] [doi: 10.1002/cncr.27467] [Medline: 22488698]
40. Merchant SJ, Chen SL. Prevention and management of lymphedema after breast cancer treatment. Breast J 2015;21(3):276-284. [doi: 10.1111/tbj.12391] [Medline: 25772311]
41. Kwok A, Palermo C, Boltong A. Dietary experiences and support needs of women who gain weight following chemotherapy for breast cancer. Support Care Cancer 2015 Jun;23(6):1561-1568. [doi: 10.1007/s00520-014-2496-5] [Medline: 25398359]
42. de Cicco P, Catani MV, Gasperi V, Sibilano M, Quaglietta M, Savini I. Nutrition and breast cancer: a literature review on prevention, treatment and recurrence. Nutrients 2019 Jul 3;11(7):pii: E1514 [FREE Full text] [doi: 10.3390/nu11071514] [Medline: 31277273]
43. Hou R, Wei J, Hu Y, Zhang X, Sun X, Chandrasekar EK, et al. Healthy dietary patterns and risk and survival of breast cancer: a meta-analysis of cohort studies. Cancer Causes Control 2019 Aug;30(8):835-846. [doi: 10.1007/s10552-019-01193-z] [Medline: 31165965]
44. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile health in oncology: a patient survey about app-assisted cancer care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
45. Ching S, Mok E. Adoption of healthy lifestyles among Chinese cancer survivors during the first five years after completion of treatment. Ethn Health 2019 Jun 25:1-20. [doi: 10.1080/13557858.2019.1634182] [Medline: 31238712]
46. Cheng H, Sit JW, Cheng KK. A qualitative insight into self-management experience among Chinese breast cancer survivors. Psychooncology 2017 Jul;26(7):1044-1049. [doi: 10.1002/pon.4279] [Medline: 27620347]

Abbreviations

IRB: Institutional Review Board
mHealth: mobile health
NTUST-NYMU: National Taiwan University of Science and Technology and National Yang-Ming University Joint Research Program
TBCF: Taiwan Breast Cancer Foundation
TCM: traditional Chinese medicine

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by FH Hsiao, K Drusany Starič, JH Lim; comments to author 04.11.19; revised version received 16.12.19; accepted 21.12.19; published 30.04.20.

Please cite as:
Hou IC, Lan MF, Shen SH, Tsai PY, Chang KJ, Tai HC, Tsai AJ, Chang P, Wang TF, Sheu SJ, Dykes PC
The Development of a Mobile Health App for Breast Cancer Self-Management Support in Taiwan: Design Thinking Approach
JMIR Mhealth Uhealth 2020;8(4):e15780
URL: http://mhealth.jmir.org/2020/4/e15780/
doi:10.2196/15780
PMID:32352390

Original Paper

Quality, Functionality, and Features of Chinese Mobile Apps for Diabetes Self-Management: Systematic Search and Evaluation of Mobile Apps

Enying Gong[1,2], MSc; Zongmuyu Zhang[1,3], MRes; Xurui Jin[2], MBBS; Yishan Liu[4,5], MSc; Lumin Zhong[4,6], MSc; Yao Wu[1,7], MIT; Xuefeng Zhong[8], PhD; Lijing L. Yan[2], MPH, PhD; Brian Oldenburg[1], Mpsycho, PhD

[1]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia

[2]Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, China

[3]School of Medicine, Tsinghua University, Beijing, China

[4]School of Public Health, Shanghai Jiaotong University, Shanghai, China

[5]College of Science and Engineering, University of Edinburgh, Edinburgh, United Kingdom

[6]Department of Epidemiology and Public Health, University College London, London, United Kingdom

[7]Melbourne School of Engineering, University of Melbourne, Melbourne, Australia

[8]Department of Personal Health, Philips Research China, Shanghai, China

Corresponding Author:
Enying Gong, MSc
Melbourne School of Population and Global Health
University of Melbourne
333 Exhibition Street
Melbourne, Victoria
Australia
Phone: 61 0452389420
Email: egong@student.unimelb.edu.au

Abstract

Background: The emergence and advancement of mobile technologies offer a promising opportunity for people with diabetes to improve their self-management. Despite the proliferation of mobile apps, few studies have evaluated the apps that are available to the millions of people with diabetes in China.

Objective: This study aimed to conduct a systematic search of Chinese mobile apps for diabetes self-management and to evaluate their quality, functionality, and features by using validated rating scales.

Methods: A systematic search was conducted to identify Chinese apps for diabetes self-management in the four most popular Chinese language mobile app stores. Apps were included if they were designed for diabetes self-management and contained at least one of the following components: blood glucose management, dietary and physical activity management, medication taking, and prevention of diabetes-related comorbidities. Apps were excluded if they were unrelated to health, not in Chinese, or the targeted users are health care professionals. Apps meeting the identified inclusion criteria were downloaded and evaluated by a team of 5 raters. The quality, functionalities, and features of these apps were assessed by using the Mobile App Rating Scale (MARS), the IMS Institute for Healthcare Informatics Functionality score, and a checklist of self-management activities developed based on the Chinese diabetes self-management guideline, respectively.

Results: Among 2072 apps searched, 199 were eligible based on the inclusion criteria, and 67 apps were successfully downloaded for rating. These 67 apps had an average MARS score of 3.42 out of 5, and 76% (51/67) of the apps achieved an acceptable quality (MARS score >3.0). The scores for the four subdomains of MARS were 3.97 for functionality, 3.45 for aesthetics, 3.21 for information, and 3.07 for engagement. On average, reviewed apps applied five out of the 19 examined behavior change techniques, whereas the average score on the subjective quality for the potential impact on behavior change is 3 out of 5. In addition, the average score on IMS functionality was 6 out of 11. Functionalities in collecting, recording, and displaying data were mostly presented in the reviewed apps. Most of the apps were multifeatured with monitoring blood glucose and tracking lifestyle behaviors as common features, but some key self-management activities recommended by clinical guidelines, such as stress and emotional management, were rarely presented in these apps.

Conclusions: The general quality of the reviewed apps for diabetes self-management is suboptimal, although the potential for improvement is significant. More attention needs to be paid to the engagement and information quality of these apps through co-design with researchers, public health practitioners, and consumers. There is also a need to promote the awareness of the public on the benefit and potential risks of utilizing health apps for self-management.

(JMIR Mhealth Uhealth 2020;8(4):e14836) doi:10.2196/14836

KEYWORDS

diabetes mellitus; self-management; mobile apps; China

Introduction

Background

Diabetes is a complex chronic disease that affects more than 100 million individuals in China [1]. The prevalence of diabetes was estimated to be 10.9% in China in 2013 [2], whereas the overall awareness, treatment, and control rates were 38.6%, 35.6%, and 33.0%, respectively [3]. Many challenges exist regarding the provision of high-quality care for patients with diabetes and improving the control rate at the population level. Effective management and control of diabetes require intensive long-term efforts from patients and health care providers. Although primary health care services related to diabetes care are 1 of the 14 items of the Basic Public Health Service in China [4], access to high-quality health services is still quite limited. In addition, low health literacy and lack of awareness on appropriate self-management strategies have further restrained the effective control among people with diabetes [5]. Therefore, developing low-cost and effective strategies for improving the self-management of diabetes is essential.

The emergence and advancement of mobile health (mHealth), defined as the use of mobile technologies for improving health care processes and outcomes, offer a promising opportunity for people with diabetes to improve their self-management and health outcomes [6-8]. Mobile phone usage among Chinese adults is now almost ubiquitous (96.8 mobile cellular subscriptions per 100 people) [9], and the number of people who surf online through their phones reached 800 million in China in 2018 [10]. Smartphone penetration accounts for 48% of all phone users [11], and there are more than 4 million mobile apps already available in the market [10]. This widespread dissemination of mobile technologies creates a huge opportunity to transform health care delivery in China. Although the evidence remains inconclusive, empirical studies have shown that mHealth interventions have the potential to improve patients' access to low-cost care, facilitate patient-provider communication, and make an impact on patients' health outcomes and quality of life [12-15].

The number of publicly available mobile apps related to diabetes and diabetes self-management has grown exponentially over the last 10 years in the global markets [8,15]. Previous studies examined English apps targeted on diabetes self-management from various dimensions, including the quality, the functions, the usability, and the application of behavior change techniques, and found that the quality and performance of these apps in general were suboptimal [8,15-19]. A few studies reviewed mobile apps for chronic disease management in China in general [20-22], but these reviews were not able to provide a comprehensive assessment of the quality of the existing mHealth solutions for people with diabetes in China.

Objectives

To address this gap, the main aim of this study was to conduct a systematic review and evaluation of mobile apps for diabetes self-management in China. The specific objectives were (1) to provide an overview of the available Chinese mobile apps for diabetes self-management, (2) to evaluate the quality of these apps with validated rating scales, and (3) to describe the key functions and features of these apps in helping people with diabetes. The study was conducted by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses framework [23].

Methods

Systematic Searching and Screening

We conducted a systematic search of mobile apps related to diabetes self-management from four Chinese language mobile app stores from January 2018 to March 2018. On the basis of the popularity of the app stores, we identified four app stores as target platforms, including Apple iTunes Store for the iPhone operating system (iOS) as well as Tencent Myapp, 360 Mobile Assistant, and Baidu Mobile Assistant for the Android system. According to a preliminary estimate, these three Android stores accounted for more than half of the market share of the Android app stores in China [24]. Keywords for searching include diabetes, blood glucose, diabetes prevention, diabetes control, and diabetes treatment. Each keyword was searched through the general search bar of the four app stores listed above.

The eligibility screening was conducted by 2 reviewers (EG and ZZ). The duplicate apps yielded from multiple stores or multiple searching terms were removed from the pool. Then, the 2 reviewers checked the eligibility of the apps based on the titles, descriptions, and screenshots shown in the app stores. Apps were included if they were used for diabetes self-management and contained at least one of the following components: blood glucose management, dietary and physical activity management, medication taking, and prevention of diabetes-related comorbidities. Apps were excluded if they were unrelated to health, not in Chinese, or the target users are health care professionals. After two rounds of screening, a list of apps was generated for further download and evaluation.

Evaluation and Data Extraction

A team of 5 raters downloaded the screened apps and independently tested the quality, functionality, and features of the apps from August 2018 to October 2018. All raters attended

a workshop for this project and received formal training regarding the study protocol and evaluation instruments, read through the handbook, and passed a pilot test before formally rating the apps. For each app reviewed, raters downloaded the app on their phone and used all the functions of the app to familiarize themselves with the app before conducting the rating. Raters then went through all the questions in the data extraction form and performed the rating. Moreover, more than 10% (8/80) of the apps were randomly selected and double rated to check their inter-rater reliability. Differences were discussed to examine whether these differences existed between systems or in the evaluation.

Rating Instrument and Measurement

The data extraction form was developed based on the Qualtrics online platform (Qualtrics). The form included four parts: (1) general information of the app, (2) the quality of the app based on the Mobile App Rating Scale (MARS) [25], (3) the scope of functionality of the app based on the IMS Institute for Healthcare Informatics Functionality score [26], and (4) a checklist regarding the features and contents of the app.

General Information

The general information section primarily extracted data that could be found from the description of the app in the app stores, including app name, operating platform, developer, version, date of the recent update, and cost. The star rating score and the number of raters were extracted based on the iOS app market because of the largely unavailable star rating in the Android stores and the diverse scale range applied across the three Android stores. The number of downloads was only available in the Android app store, and the statistics from each store were recorded if available. In addition to these descriptive information shown in the app stores, technical aspects of the app (eg, allowing password protection and requiring log-in) and 19 behavior change techniques (eg, assessment, feedback, information or education, monitoring, advice, and goal setting) were also collected based on the checklist of MARS [25] and previous studies [27].

Quality

The quality of each app was evaluated by using MARS, which is a simple, reliable, objective rating tool to provide a multidimensional measure of the app quality [25]. The scale has been shown to have excellent internal consistency and inter-rater reliability in previous studies [28]. The scale contains 19 items grouped into four domains, including engagement (entertainment, interest, customization, interactivity, and target group), functionality (performance, ease of use, navigation, and gestural design), aesthetics (layout, graphics, and visual appeal), and information quality (accuracy of app description, goals, quality and quantity of information, visual information, credibility, and evidence base). Each item was measured on a 5-point Likert scale, with 1 indicating inadequate and 5 indicating excellent. A mean score for each domain and a mean score for overall 19 items were computed as the score for the quality of the app. In addition to the objective assessment, five items were used to assess the subjective quality in terms of the perceived impact of the app on users' knowledge, attitude, and intention to change and the likelihood of actual change in diabetes self-management. These subjective quality items were scored separately.

Functionality

The functionality of the app was measured by using the IMS functionality score [26]. Unlike the functional domain within the MARS that reflects whether the app functions well, the IMS functionality score focused on the scope of the functions. The score contains seven functionality categories (informing, instructing, recording, displaying, guiding, reminding, and communicating information) and four subcategories (collecting data, sharing data, evaluating data, and intervening). Apps allowing the function were coded as 1, otherwise coded as 0. A functionality score ranging from 0 to 11 was generated for each app.

Features and Contents

For each app, we also evaluated its features and contents in promoting diabetes self-management activities. A checklist, including health indicators and behaviors monitoring and reminding, health education, and communication with professionals and peers, was derived from previous studies [8,17] and the Chinese diabetes self-management guideline [29].

Quality Control

To ensure the quality of the study, fidelity, and consistency of ratings among the raters, a handbook for the raters was developed, reviewed, and refined by experts before implementing the research activities. A workshop was undertaken in China in March 2018 to review and finalize the study protocol and the handbook.

Statistical Analysis

All the information collected through the Qualtrics online platform was downloaded for further analysis. Descriptive analysis was conducted, and the mean and SD were reported. If the data distribution was skewed, the median and IQR were reported. The inter-rater reliability score was calculated between two records generated from the double rating [28]. All analyses were conducted using Stata statistical software, version 14 (StataCorp LP), and the visualized figures were drawn using Excel (Microsoft Excel for Office 365).

Results

Systematic Search and Screening

A total of 2072 apps were identified from the initial search in four Chinese language app stores. After excluding the duplicates, 936 apps were enrolled for eligibility screening, and 199 of these apps met the inclusion criteria for further download and evaluation. However, among these apps, 108 were failed in download or registration or did not work properly after download. Moreover, 18 apps were excluded because the raters found that these apps did not meet the inclusion criteria on diabetes self-management, and another six apps were excluded because they had to be linked with specific devices to operate. Finally, 67 apps met the inclusion criteria and were formally

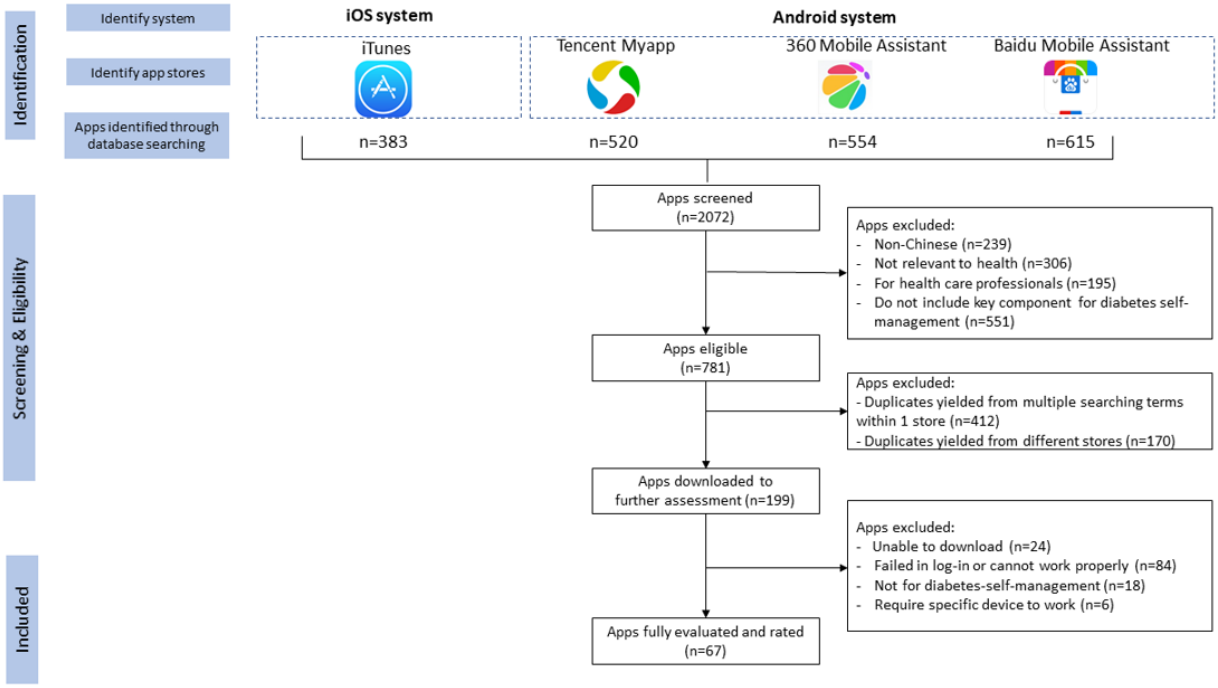
evaluated. Figure 1 provides an overview of the screening process.

Characteristics of the Apps

Multimedia Appendix 1 provides a list of the included apps and their characteristics. Moreover, 41 apps had both Android and iOS versions, 25 apps could run only on Android phones, and one app was only available in the iOS market. As for the developer of these apps, 24 apps were developed by informatic or internet technology companies, 38 by health management or biomedical companies, one by a pharmaceutical company, and three involved clinical institutions or science institutions as their codevelopers. About half (33/67 49%) of these apps released the latest version in 2018. All apps involved in the review were free to download.

On the basis of the statistics in the Android markets, the median number of downloads was 15,000 (IQR 1025-330,000), 11,000 (IQR 446-78,000), and 20,000 (IQR 1000-180,000) for Baidu, Tencent, and 360 app stores, respectively, and one app reached 7.7 million downloads as the largest number across the three stores, despite these number of downloads not necessarily implying the real users who suffered from diabetes. In addition, 73% (30/41) of apps in the iOS market had the star rating score available, with a median rating score of 4.7 (IQR 4-5), and the median number of raters was 54 (IQR 12-186). Of 67 apps, 48 (72%) required setting an account and logging in before using, and 40 (60%) allowed password protection.

Figure 1. Screening process based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram.



Presence of Behavior Change Techniques

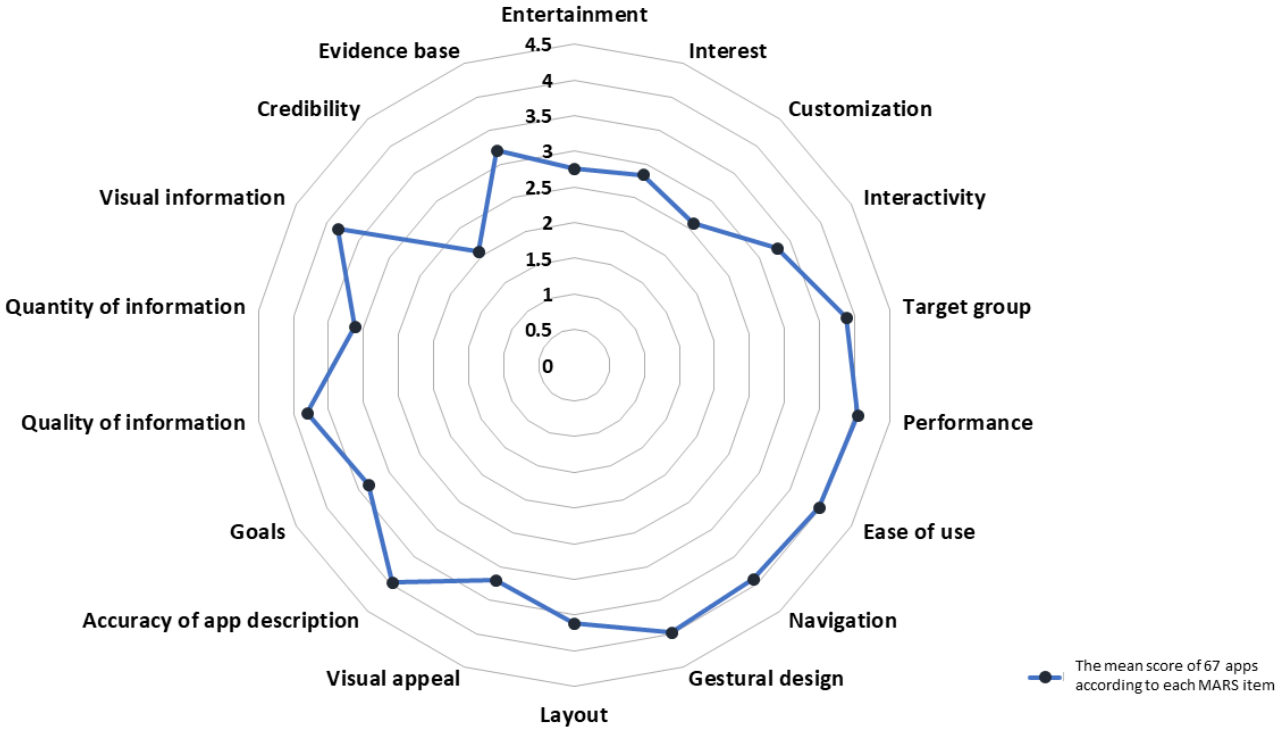
The average number of behavior change techniques used in the reviewed apps was 5 (SD 3.2) out of the 19 behavior change techniques assessed. The most frequently identified behavior change technique was self-monitoring or tracking contained in 87% (58/67) of apps, followed by information or education in 64% (43/67) of apps, assessment in 60% (40/67) of apps, feedback in 57% (38/67) of apps, and advice, tips or strategies in 54% (36/67) of apps.

Multimedia Appendix 2 presents the number of apps that applied each of the behavior change techniques.

App Quality

The average MARS score of all apps reviewed was 3.42 (SD 0.66). A total of 76% (51/67) of apps had a minimum score of 3.0, indicating that these apps reached an acceptable quality level. As for the four domains, functionality had the highest score with an average of 3.97 (SD 0.66), followed by aesthetics (mean 3.45, SD 0.88), information (mean 3.21, SD 0.73), and engagement (mean 3.07, SD 0.90). Figure 2 presents the mean score for each MARS item. Apps received a higher score in the items of performance, gestural design, ease of use, navigation, and accuracy of app description, whereas the score on credibility, customization, entertainment, and interest were relatively low. For item 19, only two apps were tested with quasi-experimental trials to evaluate their efficacy, despite the weakness of the study design in sample size and follow-up period [30,31]. Eight apps were double rated and reached an inter-rater reliability score of 0.56 (95% CI 0.41-0.69).

Figure 2. The average score of each Mobile App Rating Scale item.



The results of the subjective measurement of app quality are presented in Table 1. Raters evaluated that only three apps were likely to attract frequent users. Although all apps were free, seven apps were rated as "willing to pay" by raters. In terms of the perceived impact of these apps in promoting behavior change, apps reached an average score of 3.0 (SD 0.9) for five subjective items of MARS. Raters strongly agreed or agreed that half of these apps might improve users' awareness and knowledge, but only 30% (20/67) of the apps had the potential to change users' attitudes, intention to change, and lead to the actual change in diabetes self-management and related health outcomes.

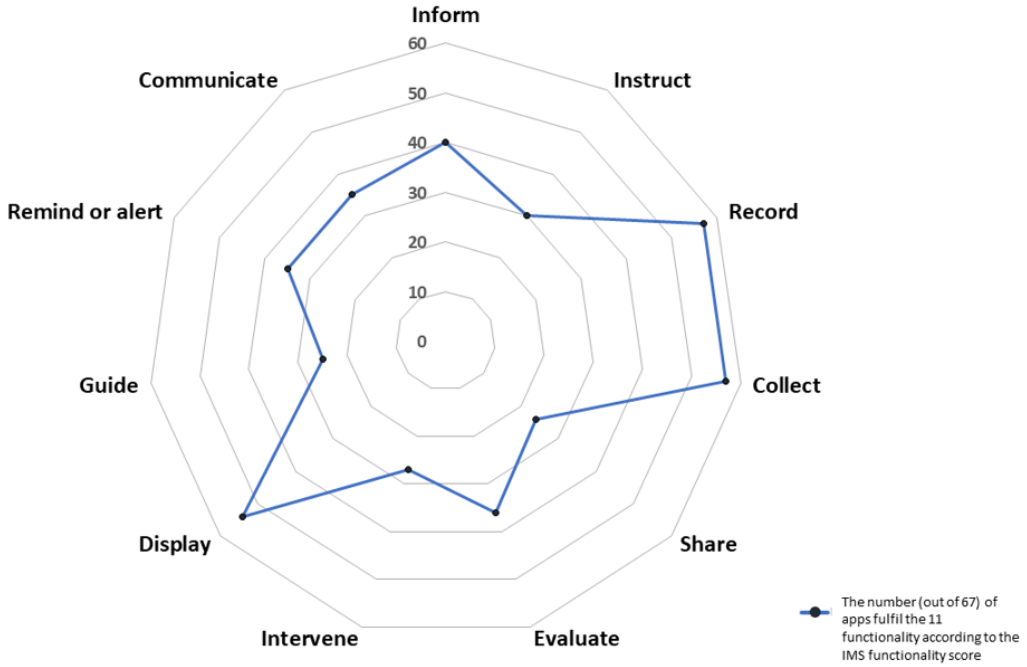
Table 1. Subjective measurement of quality of apps (N=67).

Subjective quality statement	Apps, n (%)
Recommend apps to people who might benefit from it	
Not at all	23 (34)
A few people	16 (24)
Several people	16 (24)
Many people	10 (15)
Everyone	2 (3)
Times of use of this app in the next 12 months if the app is relevant	
None	24 (36)
1-2	17 (25)
3-10	13 (19)
10-50	10 (15)
>50	3 (5)
Willing to pay for the app	
No	40 (60)
No or probably not	9 (13)
Maybe	11 (16)
Probably yes	7 (11)
Strongly agree or agree that the app will improve	
Awareness	35 (52)
Knowledge	36 (54)
Attitude	22 (33)
Intention to change	26 (39)
Behavior change	26 (39)

Functionality

Figure 3 presents the number of apps that meet the functionality category based on the scale. On average, the apps reviewed met 6.3 functionalities out of the 11 items of the scale (SD 2.8). Moreover, four apps had all 11 functionalities, and five apps had 10 functionalities. Among the 67 apps, 57 (85%) had the function to capture user-entered data and store the data on the user's phone, and 54 (80%) apps could graphically display user-entered data. However, only about half of the reviewed apps (36/67, 54%) were able to evaluate the entered data, 36% had the function to share and transmit the entered data, and 40% (27/67) were able to promote intervention based on the data collected. More than half of the apps had the function to provide information, instruction, and guidance to the users. Among the 67 apps reviewed, 35 apps provided support for communication with others.

Figure 3. Number of apps that meet the functionality categories.



Features and Contents

All the reviewed apps contained multiple features to fulfill users' activities in self-management (Table 2). Among all features, tracking and monitoring health conditions and lifestyle behaviors was the most popular one that presented in most of the apps. About 81% (54/67) of the reviewed apps fulfilled the users' needs in monitoring blood glucose, but less than half of these apps supported data transfer from wearable or medical devices. Setting individual reminders on self-management–related activities is another key feature in reviewed apps, and blood glucose testing (23/67 34%), medication taking (21/67 31%), and blood pressure testing (11/67 16%) were the most common purpose of reminders. In addition, apps also provided health education information in text, pictures, and videos and covered a broad range of topics, such as blood glucose control and diet (in 44/67, 66% apps), diabetes-related concepts (in 39/67, 58% apps), and physical activities (in 36/67, 54% apps). However, stress and emotional control and blood cholesterol control only presented in about one-fourth of the reviewed apps. In addition, less than 10% (7/67) of the apps were able to provide individualized health education information to meet the needs of users with various health literacy levels.

In addition to self-monitoring and health education, these apps also supported users to connect with health professionals and other patients online or offline. Among 67 apps, 26 apps (39%) supported online consultation with health care professionals, and four apps (6%) helped users make appointments with health care providers offline. Moreover, 11% (7/67) apps assisted users in navigating the closest pharmacies or clinics, and 22% (15/67) contained an online market for purchasing medicines, supplements, or devices. Besides, 33% (22/67) of apps contained social forums or blogs to facilitate communication between peers, and 27% (18/67) of apps supported data sharing with family members.

Table 2. Features to support diabetes self-management.

Features	Apps with manual data entry, n (%)	Apps with auto-transfer from wearable/medical devices, n (%)
Track and monitor health conditions		
Blood glucose level	54 (81)	29 (43)
Blood pressure level	34 (51)	10 (15)
Blood cholesterol level	10 (15)	2 (3)
Physical activity level	34 (51)	13 (19)
Diet pattern	26 (39)	0 (0)
Weight	33 (49)	7 (11)
Medication taking	24 (36)	0 (0)
Health education on		
Information about diabetes	39 (58)[a]	7 (11)[b]
Blood glucose control	44 (66)[a]	8 (12)[b]
Blood pressure control	22 (33)[a]	2 (3)[b]
Blood cholesterol control	15 (22)[a]	0 (0)[b]
Physical activity	36 (54)[a]	5 (8)[b]
Healthy diet	44 (66)[a]	5 (8)[b]
Weight control	26 (39)[a]	1 (2)[b]
Diabetes-related complications	35 (52)[a]	6 (9)[b]
Medication use and adherence	30 (45)[a]	7 (11)[b]
Stress and emotional control	17 (25)[a]	1 (2)[b]
Set individualized reminders on		
Blood glucose testing	23 (34)	N/A[c]
Blood pressure testing	11 (16)	N/A
Physical activity	10 (15)	N/A
Healthy diet	7 (11)	N/A
Weight test	7 (11)	N/A
Medication taking	21 (31)	N/A
Appointment with physicians	5 (8)	N/A
Make an appointment with physicians		
Online consultation	26 (39)	N/A
Face-to-face consultation	4 (6)	N/A
Share recorded data		
With health professionals	24 (36)	N/A
With family members/friends	18 (27)	N/A
Communicate through forums/blogs	22 (33)	N/A
Evaluate the risk of having complications	17 (25)	N/A
Purchase medicines/devices	15 (22)	N/A
Find out pharmacy stores/clinics	7 (11)	N/A

[a]General information.

[b]Individualized information.

[c]Not applicable.

Discussion

Principal Findings

In this study, we provide a snapshot of publicly available mobile apps for diabetes self-management in China. Through the searching and screening of more than 2000 apps, 67 apps were identified as being suitable for an in-depth evaluation. On the basis of a comprehensive review, our study found that the quality of these apps is suboptimal with considerable variability. Although most of the reviewed apps have multiple features, some common deficiencies were identified as poor engagement, low adherence to guidelines, and lacking evidence on health benefits.

We observed a consistent pattern with previous reviews showing that most of the apps performed better in the domains of functionality and aesthetics but poorly in information and engagement [18,27,32]. Although studies have shown that engagement is crucial for the users' uptake and the improvement in patients' health outcomes [33], most of the apps that we reviewed were not able to apply effective strategies to improve users' experience and engagement. For example, wireless sensors are now widely available, but less than half of the apps that we reviewed supported wireless automatic data acquisition for blood glucose monitoring. Manual data input exposes the users to erroneous workload, which may lead to poor engagement, low compliance, and abandoning [17].

In addition to the issue of engagement, the quality and evidence base of information provided by the apps could also be improved. Compared with previous studies that suggested insufficient health education features within apps [17], our review showed a higher proportion of apps designed with health education features. However, the source of information in some of the reviewed apps was not able to be identified and likely not to be evidence-based. The absence of such evidence-based information may expose end users to misleading or incorrect information, thus putting people with diabetes at risk of potential negative health outcomes. Another critical issue is that only a small proportion of the reviewed apps could set personalized reminders or provide personalized information to users. Although the advantages of using the mHealth tools in providing a large amount of tailored information in real time have been well acknowledged [34], most of the apps involved in our review were not able to achieve the expectation of providing patient-tailored support and health education.

It is also important to note that most of the reviewed apps only partially meet the requirements recommended by the clinical guidelines for diabetes self-management [29]. Supports in managing diabetes-related comorbidities and complications were unavailable in about half of the reviewed apps. For example, activities such as tracking blood pressure and blood cholesterol, stress management, emotional control, frequent foot check, and eye tests were absent in most of the apps. This result is consistent with previous studies that focused on the adherence to guidelines of diabetes apps [16,20]. As most of these apps were developed by informatics or health care technology companies, the involvement of health professionals with a deeper understanding of the best clinical guidelines was insufficient. Therefore, the public, researchers, and policymakers should be aware of the limitations and potential risks of the currently available apps. A multidisciplinary development and co-design process with extensive involvement from health professionals should be highly recommended for future app design.

Another emergent finding from this study was that about 40% (27/67) of the reviewed apps provided an online consultation to users and supported offline appointment making, and 5% (34/67) of the apps contained features to link the online consultation to the offline health care services. This feature had not been identified in previous reviews of English apps. As China faces significant challenges in addressing the increasing burden of chronic diseases, online consultation and telemedicine have been considered as an innovative approach with the potential to provide high-quality, accessible, and affordable care beyond geographical boundaries. Since April 2018, the Chinese government has promoted the "Internet plus healthcare" initiative as a strategy to alleviate the problem of inaccessible and expensive health care services and encourage online health care services, including consultation, appointment making, and test result inquiry [35]. Therefore, we observed an increasing number of apps containing features of online consultation or linking online services with offline care. However, only a few reviewed apps support the data sharing between patients and providers, let alone linking the data in the apps with existing health care records. This weakness in current apps indicates that further efforts are needed to integrate such services with the existing health care information system.

The actual impact of these apps in supporting users' behavior change and improving diabetes-related health outcomes is unclear. Although a variety of behavior change techniques were incorporated into many of these apps, most of these techniques were applied to support information exchange and records, such as monitoring, education, and assessment. Consistent with the findings from previous studies [36-38], behavior change techniques, as internal drivers, were applied only in a few apps, and the techniques that could provide just-in-time intervention for behavior changes were underutilized. In addition, only two apps in the review had been tested through quasi-experiment trials to evaluate its efficacy on health outcomes [30,31]. Similar to most of the studies that examined the effectiveness of apps for diabetes self-management [12,15], these two studies had a relatively small sample size and were unable to observe meaningful long-term effects. Therefore, the effectiveness and benefit of using these apps for diabetes self-management need to be further examined.

Comparison With Other Research

To the best of our knowledge, this is one of the few studies that utilized validated scales for evaluating Chinese language mobile apps. Although there are previous studies that reviewed Chinese language apps for either chronic disease in general [22] or diabetes [20,21], our study has a broader scope and utilized validated scales. The four most popular Chinese language mobile app markets were selected as the searching source, which ensured the representativeness and high coverage of apps in this review. This study also analyzed the quality, functionality, and

features of apps by using validated rating scales, which filled the research gap and makes the comparison across apps possible in the future.

Comparing these results with those obtained from previous reviews of comparable English apps, a similar level of the quality of apps was found [18,27,32,39]. The apps involved in this review also had similar deficiencies in terms of insufficient engagement strategies, lacking patient-tailored evidence-based information, inadequate functions for a variety of self-management activities recommended by the guideline, and the absence of integration with the existing health care system [8,17]. In addition, although there is an increasing number of trials investigating the effectiveness of English mobile apps for diabetes self-management [40,41], only a few such trials have been undertaken in China where there is the largest population of people with diabetes. Future studies are in great need to evaluate the effectiveness of these publicly available apps in improving the health outcomes among people with diabetes.

Limitations

This study has some limitations. First, the apps were evaluated by researchers based on short-term use. These results, therefore, cannot reflect the opinion of actual target users—people with diabetes in China—who are generally elderly with relatively poor health and digital literacy. Therefore, further research is needed to evaluate these apps by people with diabetes. Second, although we reached good coverage in selecting apps from the four most popular app stores, apps that were not publicly available were not included. For example, apps that were only available to download by private invitation and apps that can only be used with the support of specific medical devices were excluded. Third, some differences in results were observed through the double rating process. On the basis of the standard protocol, the differences between raters were discussed in detail, and it was found that the inconsistency in score could be explained by the differences in the operating systems or the phone model as well as the different information dispatched by the system based on the log-in information. In addition, although the MARS was rigorously developed, tested, and widely applied in evaluating health apps for various conditions worldwide [27,32], there has been no study evaluating the reliability and validity of the scale in evaluating Chinese apps. To minimize the potential issues in adopting the English version of the scale for the Chinese app review, we provided handbooks with detailed explanations and examples for all raters, organized training sessions with pilot testing, and discussed potential issues thoroughly before the rating. Finally, we were not able to evaluate apps in all conceivable dimensions. For example, privacy and information security were other key domains to evaluate health apps, but our study only assessed it through whether the app needed a log-in and allowed password protection. We were also not able to analyze professionally whether the advice provided for users were accurate and evidence-based. Future studies should focus on these aspects of the evaluation.

Conclusions

With the proliferation of available technologies, mobile apps are promising tools to support diabetes self-management. In general, the publicly available Chinese apps for diabetes self-management involved in the review contained multiple functions and features, but the quality is suboptimal with enormous potential for improvement. More work is needed to improve these apps by applying strategies to engage users, providing more comprehensive and evidence-based information, and to support a broader range of activities recommended by the clinical guidelines. Rigorous scientific evaluations of the effectiveness and value of these apps on behavior change and health outcomes are also greatly needed.

This study also provides important public health implications. The proliferation of smartphones enables the public to access mobile apps as a potential tool for health promotion and disease self-management. However, the public may not be able to fully realize the potential risks of using mobile apps if the apps have not been designed with full adherence to the clinical guidelines or have not been evaluated on its effectiveness. To overcome this challenge and increase the impact of mobile apps on health improvement at the population level, health researchers and professionals should be more engaged in the development and evaluation of the mHealth technologies to increase the quality of available apps and deliver evidence on the effectiveness. In addition, the government should pay attention to the quality and safety of publicly available apps, set up platforms to help health care providers and patients identify evidence-based health apps, and promote health literacy and awareness among the public on the potential benefit and risks of using health apps.

Acknowledgments

The authors would like to acknowledge the contribution of the Shanghai Center for Disease Control and Prevention for stimulating the original idea for this review as well as Shaira Baptista from the University of Melbourne, who was involved with the early conceptualization of the project. The authors would also like to thank their colleagues from the Shanghai Center for Disease Control and Prevention (Dong Yuan, Minna Cheng, Qinping Yang, and Fei Wu), the Shanghai Jiaotong University (Yong Cai, Xueqin Jiang, and Yiran Li), the Duke Kunshan University (Yunfan Ge, Ji Xiao, Duan Zhao, and Erik Yan), and the University of Melbourne (Ameera Katar) for providing feedback and support throughout the study. The authors also want to thank Stoyan R Stoyanov, who provided helpful suggestions related to the use of MARS and the strategies for ensuring the quality of the evaluation. EG was supported by the Graduate Research Scholarship provided by the University of Melbourne. ZZ was supported by the Melbourne Research Scholarship and the Chinese Scholarship Council. The study workshop was organized and supported by the Global Health Research Center, Duke Kunshan University.

Authors' Contributions

EG drafted the manuscript. EG and ZZ conducted the initial searching and screening. EG, ZZ, XJ, YL, and LZ performed app ratings and data extraction. EG and XJ performed the data analysis. BO, EG, ZZ, LY, and XZ contributed to the study design and instrument design. All coauthors contributed to the revision of the manuscript and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of evaluated apps.
[DOCX File , 672 KB - mhealth_v8i4e14836_app1.docx]

Multimedia Appendix 2
Number of apps that applied behavior change techniques.
[PNG File , 27 KB - mhealth_v8i4e14836_app2.png]

References

1. Xu Y, Wang L, He J, Bi Y, Li M, Wang T, 2010 China Noncommunicable Disease Surveillance Group. Prevalence and control of diabetes in Chinese adults. J Am Med Assoc 2013 Sep 4;310(9):948-959. [doi: 10.1001/jama.2013.168118] [Medline: 24002281]
2. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. J Am Med Assoc 2017 Jun 27;317(24):2515-2523 [FREE Full text] [doi: 10.1001/jama.2017.7596] [Medline: 28655017]
3. Chinese Center For Disease Control And Prevention. 2016. Report on Chronic Disease Risk Factor Surveillance in China URL: http://ncncd.chinacdc.cn/ [accessed 2020-01-23]
4. Li X, Lu J, Hu S, Cheng KK, de Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 9;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
5. Guo X, Yuan L, Lou Q, Shen L, Sun Z, Zhao F, Chinese Diabetes Education Status Survey Study Group. A nationwide survey of diabetes education, self-management and glycemic control in patients with type 2 diabetes in China. Chin Med J (Engl) 2012 Dec;125(23):4175-4180. [Medline: 23217382]
6. Kay M, Santos J, Takane M. World Health Organization. 2011. mHealth: New Horizons for Health Through Mobile Technologies URL: https://www.who.int/ehealth/mhealth_summit.pdf [accessed 2020-01-23]
7. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
8. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
9. The World Bank. 2016. Mobile Cellular Subscriptions (per 100 People) - China URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=CN [accessed 2019-02-08]
10. Office of the Cyber Security and Information Committee of the CPC Central Committee. 2018 Aug 20. The Statistic Report on China Internet Network Development URL: http://www.cac.gov.cn/2018-08/20/c_1123296882.htm [accessed 2020-01-23]
11. Statista. 2016. Smartphone Penetration as Share of Population in China 2015-2023 URL: https://www.statista.com/statistics/321482/smartphone-user-penetration-in-china [accessed 2019-03-08]
12. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
13. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
14. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Dec;15(12):109 [FREE Full text] [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
15. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
16. Breland JY, Yeh VM, Yu J. Adherence to evidence-based guidelines among diabetes self-management apps. Transl Behav Med 2013 Sep;3(3):277-286 [FREE Full text] [doi: 10.1007/s13142-013-0205-4] [Medline: 24073179]
17. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]

18. Chavez S, Fedele D, Guo Y, Bernier A, Smith M, Warnick J, et al. Mobile apps for the management of diabetes. Diabetes Care 2017 Oct;40(10):e145-e146 [FREE Full text] [doi: 10.2337/dc17-0853] [Medline: 28774944]
19. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. J Telemed Telecare 2012 Jun;18(4):235-238. [doi: 10.1258/jtt.2012.111002] [Medline: 22604278]
20. Gao C, Zhou L, Liu Z, Wang H, Bowers B. Mobile application for diabetes self-management in China: do they fit for older adults? Int J Med Inform 2017 May;101:68-74 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.02.005] [Medline: 28347449]
21. Nie L, Xie B, Yang Y, Shan YM. Characteristics of Chinese m-Health applications for diabetes self-management. Telemed J E Health 2016 Jul;22(7):614-619 [FREE Full text] [doi: 10.1089/tmj.2015.0184] [Medline: 27171016]
22. 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]
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Talking Data. Active Ranking URL: http://mi.talkingdata.com/app-rank.html?type=12 [accessed 2020-01-23]
25. 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 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Aitken M, Gauntlett C. ScienceOpen. New Jersey, USA: IMS Institute for Healthcare Informatics; 2013. Patient Apps for Improved Healthcare From Novelty to Mainstream URL: https://www.scienceopen.com/document?vid=696d6314-87cd-43c3-9a5a-421874cbb752 [accessed 2020-01-23]
27. 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 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
28. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
29. Chinese Diabetes Society. Guidelines for the prevention and control of type 2 diabetes in China (2017 Edition). Chin J Pract Intern Med 2018;38(4):292-344 [FREE Full text] [doi: 10.19538/j.nk2018040108]
30. Yang Y, Shen C. The efficacy of using 'Zhang Shang Tang Yi' mobile App for managing blood glucose among people with diabetes (in Chinese). Today Nurse 2018;25(9):38-39 [FREE Full text]
31. Liang C, Lai X, Zhong S. Effect of mobile applications of diabetes management on the compliance of diety treatment of type 2 diabetes mellitus in community. Nurse J Chin PLA 2013;30(13):10-13 [FREE Full text]
32. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
33. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
34. Oldenburg B, Taylor C, O'Neil A, Cocker F, Cameron L. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
35. China government website_central government portal. 2018. General Office of the State Council Opinions on Promoting the Development of 'Internet + Medical Health' URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2019-02-02]
36. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
37. Fitzgerald M, McClelland T. What makes a mobile app successful in supporting health behaviour change? Health Educ J 2017;76(3):373-381 [FREE Full text] [doi: 10.1177/0017896916681179]
38. Cheng F, Xu J, Su C, Fu X, Bricker J. Content analysis of smartphone apps for smoking cessation in China: empirical study. JMIR Mhealth Uhealth 2017 Jul 11;5(7):e93 [FREE Full text] [doi: 10.2196/mhealth.7462] [Medline: 28698170]
39. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
40. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
41. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
iOS: iPhone operating system

Edited by G Eysenbach; submitted 28.05.19; peer-reviewed by L de Witte, R Sun, M Gaman; comments to author 07.11.19; revised version received 18.11.19; accepted 15.12.19; published 07.04.20.

Please cite as:
Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, Zhong X, Yan LL, Oldenburg B
Quality, Functionality, and Features of Chinese Mobile Apps for Diabetes Self-Management: Systematic Search and Evaluation of Mobile Apps
JMIR Mhealth Uhealth 2020;8(4):e14836
URL: https://mhealth.jmir.org/2020/4/e14836
doi:10.2196/14836
PMID:32255432

Original Paper

Mobile Health Apps for Improvement of Tuberculosis Treatment: Descriptive Review

Lina Keutzer[1*], RPh; Sebastian G Wicha[2*], PhD; Ulrika SH Simonsson[1*], MSc, PhD

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

[2]Department of Clinical Pharmacy, Institute of Pharmacy, University of Hamburg, Hamburg, Germany

[*]all authors contributed equally

Corresponding Author:
Ulrika SH Simonsson, MSc, PhD
Department of Pharmaceutical Biosciences
Uppsala University
Box 591
Uppsala, 751 24
Sweden
Phone: 46 18 471 400
Email: Ulrika.Simonsson@farmbio.uu.se

Abstract

Background: Mobile health (mHealth) is a rapidly emerging market, which has been implemented in a variety of different disease areas. Tuberculosis remains one of the most common causes of death from an infectious disease worldwide, and mHealth apps offer an important contribution to the improvement of tuberculosis treatment. In particular, apps facilitating dose individualization, adherence monitoring, or provision of information and education about the disease can be powerful tools to prevent the development of drug-resistant tuberculosis or disease relapse.

Objective: The aim of this review was to identify, describe, and categorize mobile and Web-based apps related to tuberculosis that are currently available.

Methods: PubMed, Google Play Store, Apple Store, Amazon, and Google were searched between February and July 2019 using a combination of 20 keywords. Apps were included in the analysis if they focused on tuberculosis, and were excluded if they were related to other disease areas or if they were games unrelated to tuberculosis. All apps matching the inclusion criteria were classified into the following five categories: adherence monitoring, individualized dosing, eLearning/information, diagnosis, and others. The included apps were then summarized and described based on publicly available information using 12 characteristics.

Results: Fifty-five mHealth apps met the inclusion criteria and were included in this analysis. Of the 55 apps, 8 (15%) were intended to monitor patients' adherence, 6 (11%) were designed for dosage adjustment, 29 (53%) were designed for eLearning/information, 3 (6%) were focused on tuberculosis diagnosis, and 9 (16%) were related to other purposes.

Conclusions: The number of mHealth apps related to tuberculosis has increased during the past 3 years. Although some of the discovered apps seem promising, many were found to contain errors or provided harmful or wrong information. Moreover, the majority of mHealth apps currently on the market are focused on making information about tuberculosis available (29/55, 53%). Thus, this review highlights a need for new, high-quality mHealth apps supporting tuberculosis treatment, especially those supporting individualized optimized treatment through model-informed precision dosing and video observed treatment.

(JMIR Mhealth Uhealth 2020;8(4):e17246) doi:10.2196/17246

KEYWORDS

mHealth; video observed treatment; eHealth; model-informed precision dosing; MIPD; tuberculosis; mobile apps; therapeutic drug monitoring; smartphone

Introduction

Tuberculosis is an infectious disease caused by *Mycobacterium tuberculosis*, which usually affects the lungs. In 2018, the World Health Organization (WHO) reported a global tuberculosis incidence of 10 million [1]. With approximately 1.5 million fatalities each year [1], tuberculosis remains one of the most common causes of death from an infectious disease worldwide.

The treatment success rate for new and recurrent tuberculosis cases was estimated at 85% globally in 2017 [1]. This means that a substantial number of patients still fail to respond to treatment, have a relapse of disease, or develop drug-resistant tuberculosis [1]. There are multiple reasons for an unsuccessful treatment outcome, such as suboptimal plasma concentrations of tuberculosis drugs [2], lack of patient adherence [3], difficulties in diagnosis, or lack of education [4]. Mobile health (mHealth) apps could be valuable tools to overcome these challenges in tuberculosis treatment. Despite accumulating evidence that improving patient adherence through mobile technologies has a positive impact on treatment outcome [5-9], a clinical benefit remains to be proven for mobile interventions with respect to individualized dosing, patient education, and diagnosis.

This aim of this review was to discover, describe, and categorize Web-based and mHealth apps related to tuberculosis on the market. In 2016, Iribarren et al [10] published a review and evaluation on this matter; however, given the rapid development and further expansion of mHealth [11], an update on the topic is needed.

Methods

Search Strategy

The PubMed database, Google Play Store, Apple Store, Amazon, and Google were searched extensively in Sweden between February and July 2019 using the keywords “TB,” “Tuberculosis,” “Tuberkulos,” “Tuberkulose,” “Tuberculose,” “TDM,” “Therapeutic Drug Monitoring,” “Model-informed precision dosing,” “Decision-support software,” “Clinical pharmacokinetics,” “Dosing,” “Individualized dosing,” “Personalized medicine,” “Dose calculator,” “VOT,” “VDOT,” “videoDOT,” “eDOT,” “video observed treatment,” and “virtually observed treatment”. The keywords were selected by searching the literature for reviews dealing with mobile interventions for tuberculosis treatment and video observed treatment (VOT). Furthermore, studies and reviews from references in previously discovered sources were included. The search was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12].

Inclusion and Exclusion Criteria

mHealth apps in all languages were included if they focused on active or latent tuberculosis and were excluded if they were dedicated to other infectious diseases, or if they were not created for health improvement (eg, games).

Description of Discovered Mobile Health Apps

Discovered mHealth apps are summarized in tables in Multimedia Appendix 1, which are classified based on the following characteristics: description, intended end user, cost, available languages, downloads, store, country developed, and qualification as a medical device. For evaluation of apps intended for therapeutic drug monitoring (TDM) and model-informed precision dosing (MIPD) for available tuberculosis drugs, the provided output from the program, electronic health record (EHR) integration, and required training were additionally assessed. All identified mHealth apps were assessed based on the information provided in the app stores, the product’s webpage, or accompanying publications.

Results

Apps Identified and Classification

Our search identified a total of 376 mHealth apps according to the selected keywords. After removal of duplicates and irrelevant apps, 69 apps were screened and assessed in detail, 11 of which were excluded because they were focused on other disease areas, and 3 were excluded because they were games unrelated to tuberculosis. Finally, 55 mHealth apps were included in this review (Figure 1).

The 55 mHealth apps meeting the inclusion criteria were categorized (Table 1) and a descriptive, qualitative analysis was conducted. A detailed summary of all apps reviewed is provided in Multimedia Appendix 1 categorized according to intended use: monitor patients’ adherence, dosage adjustment, eLearning/information on tuberculosis, tuberculosis diagnosis, and other purposes. Each category is described in detail below.

Figure 1. Flow chart of search strategy following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

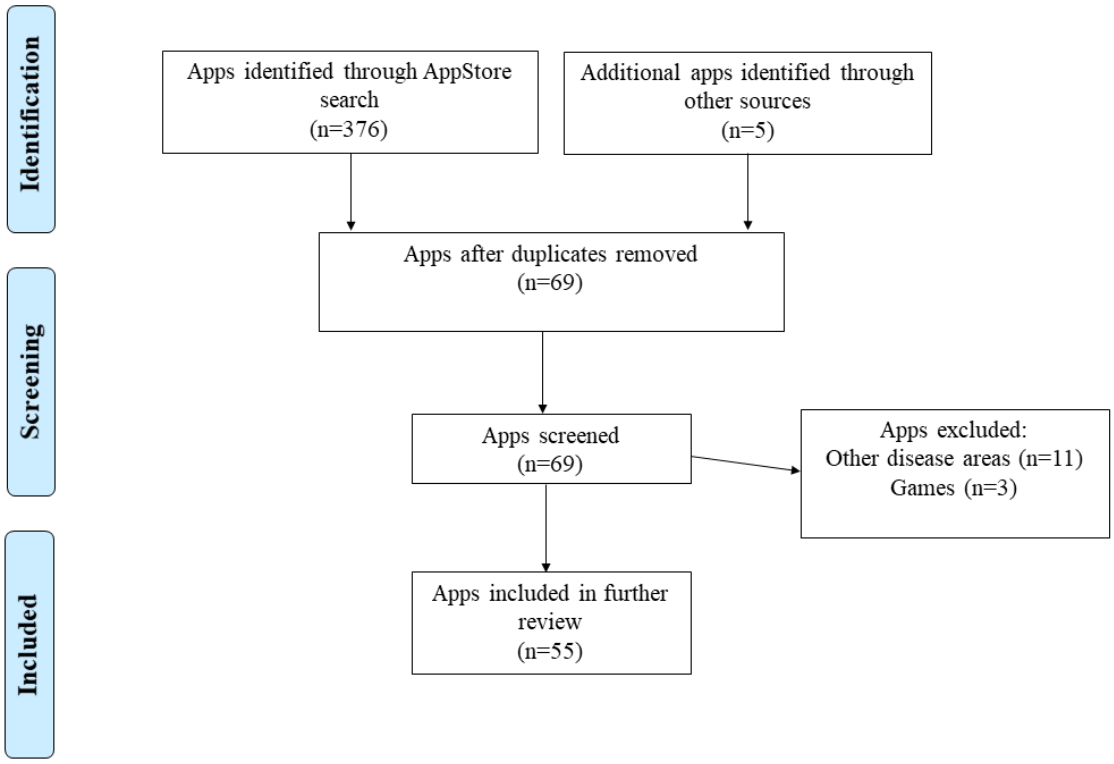


Table 1. Categorization of the included mHealth apps (N=55).

Category	Description	n (%)	Examples
Adherence monitoring	Assistance for patients to keep track of their daily drug intake; video observed treatment facilitates directly observed treatment	8 (15)	miDOT – EMOCHA; AiCure; SureAdhere; TBmCure; Wisepill evriMED
Individualized dosing	Assistance for doctors with dosage adjustments based on certain patient characteristics; dosage calculators based on guidelines	6 (11)	DoseMeRx; MwPharm; InsightRx; TDMx
eLearning/information	Depiction of official guidelines; information about the disease itself and its therapy; assistance for clinicians and students to learn about tuberculosis diagnosis and disease management; display of news in the field of tuberculosis	29 (53)	Tuberculosis TB treatment and Plan; ExplainTB; Tuberculosis TB Symptoms, Causes & Diet Help; TB mobile
(Self-) Diagnosis	Diagnostics based on data input (eg, questionnaire answers, cough sound); apps for clinicians and students to train in tuberculosis diagnosis	3 (6)	TuberSpot; TimBre for Tuberculosis (TB); Diagnosa Tuberkulosis (TB)
Others	Evaluation of treatment costs; tracing of people having contact with an infected person; monitoring and tracking of patients with tuberculosis; assistance with data gathering; creation of laboratory reports	9 (16)	CAD4TB; eDetection; TB eHealth; EdetectTB; LTBI Care

Adherence Monitoring

VOT (synonyms: video directly observed treatment, electronically directly observed treatment, video directly observed treatment, mobile directly observed treatment) describes the daily remote observation of drug intake using a smartphone app. There are two strategies to perform VOT: synchronous and asynchronous VOT. Synchronous VOT involves a live video call between patients and health care personnel. In asynchronous VOT, patients record a video of themselves, which is then stored and forwarded for later viewing by the provider. Advantages of asynchronous VOT compared to synchronous VOT are that it can be used outside of business hours and no time with the provider has to be scheduled [6,7].

Out of the 8 mHealth apps for adherence control, 3 are intended for VOT (miDOT-EMOCHA, SureAdhere, and AiCure).

MiDOT is a medication adherence app developed by emocha Mobile Health Inc (Baltimore, MD, USA). The app uses asynchronous VOT technology (store-and-forward), and care teams then review the videos and engage daily. Patients can report side effects, and the app has an additional function to filter patients struggling with adherence or experiencing side effects. The app can be purchased for Android and Apple [13].

SureAdhere is an mHealth app for asynchronous VOT. The app development was initiated by Richard Garfein from the University of California San Diego, USA. The app is compatible with Apple and Android and must be purchased. Features include patient text message or email reminders, notification to providers after a missing dose, side effect reporting, and report generation [14].

AiCure is an mHealth app using artificial intelligence to confirm medication ingestion. The software captures video, audio, and behavioral data. This app has been used in clinical trials and for population health to ensure patient adherence. The built-in algorithms have been validated against plasma blood levels. The software gathers data that can then be reviewed by health care staff [15]. In contrast to the other two main VOT tools on the market (miDOT, SureAdhere), the software of AiCure itself evaluates correct drug intake, which means that no human review is needed. Therefore, the technology is designated as automated directly observed treatment rather than VOT. The app identifies facial ID, the medication, and its ingestion. Improper administration or missed doses trigger alerts to health care workers [16]. Features of the app include reminders, interactive assistance, real-time communication with clinicians, EHR integration, and display of treatment progress [15].

Wisepill evriMED is a VOT-like app, which is connected to a smart pill dispenser that registers opening of the pillbox and subsequently sends a signal to the app. This adherence monitoring app is classified as an electronically directly observed treatment solution [17].

Other apps developed to facilitate patients' daily drug intake include Adhere2Tx-TB, Stop TB, TBmCure, and Sembuh TB, which alert and remind patients to take their medication on a regular basis (see Multimedia Appendix 1).

Individualized Dosing

The 6 mHealth apps providing health care professionals with assistance in dose optimization are TB Doctor, Medical Management of MDR-TB, DoseMeRx, MwPharm, InsightRx, and TDMx.

TB Doctor and Medical Management of MDR-TB are mHealth apps that enable clinicians to calculate individual doses depending on a patient's body weight with the help of dosing tables based on current guidelines both for drug-susceptible and resistant TB.

The remaining four tools (DoseMeRx, MwPharm, InsightRx, TDMx) have been developed for dose individualization of TB medication at the bedside based on more information than body weight (Table 2).

DoseMeRx is a decision-support software for precision dosing using MIPD. The commercial cloud-based Web app, which is also available as a mobile app for Android and Apple, uses several published clinically validated population pharmacokinetic models for the calculation of individualized doses to reach the therapeutic target. All patient data are stored in a patient file and EHR integration is also possible. The software is registered as a class l medical device in Europe and Australia [18]. DoseMeRx is currently used in public and private hospitals, as well as in many teaching institutions worldwide [19], and is considered to be very user-friendly [20].

Table 2. Overview of mHealth apps for model-informed precision dosing of tuberculosis drugs.

Feature	DoseMeRx	MwPharm	InsightRx	TDMx
Available tuberculosis drugs	linezolid, bedaquiline, isoniazid, rifampicin, pyrazinamide, ethambutol, para-aminosalicylic acid, moxifloxacin, levofloxacin	isoniazid, rifampicin, ethambutol, streptomycin	ciprofloxacin, linezolid, rifampicin, meropenem, amikacin	meropenem, amikacin, rifampicin
Compatibility	Mac, Windows, Linux, Android, iOS	Cloud-based platform	Cloud-based platform	Cloud-based platform
Output from program	Doses and pharmacokinetic parameter estimates	Doses and pharmacokinetic parameter estimates	Doses and pharmacokinetic parameter estimates	Doses and pharmacokinetic parameter estimates
Electronic health record integration	Yes: EPIC App Orchard, Cerner Millennium, Allscripts	Yes	Yes: EPIC App Orchard, Cerner Millennium, Meditech, Centricity	No
Availability	Web-based	Web-based	Web-based	Web-based
Required training	Minimal	Minimal	Minimal	Minimal
Further information (reference)	[18]	[22]	[24]	[29]
Cost	Available at [18]	1250 Euro per license	Not publicly available	Free
Medical device	Yes	Yes	No	No

MwPharm (Mediware a.s.) was developed in 1982 at the University of Groningen, the Netherlands [21]. The Web-based version can be used from any device, including smartphones. MwPharm is a decision-support software and is therefore classified as a class l medical device [22]. The clinical value of MwPharm has been proven at the majority of Dutch hospitals and has been declared a Dutch standard by the Dutch Association of Hospital Pharmacists. Furthermore, a review from 2013 by Fuchs et al [23] comparing TDM software ranked MwPharm in the leading position. Pharmacokinetic data are analyzed using either a Bayesian approach or nonlinear regression. It is possible to modify the implemented models or to add a new model [23]. Features of the software include storage of patient records, EHR integration, report generation, genotype/phenotype analysis, and availability of models for different patient groups [22].

InsightRx (San Francisco, CA, USA) is a cloud-based Web app for precision dosing using a Bayesian approach; all implemented models are clinically validated [24]. InsightRx is considered to be very comprehensive, easy to use, and visually rich [20]. The main features include patient file storage, tracking of dosing practices across institutions, dosage history, dosing reference tables, printable reports, and constant model improvement using data from institutions [24,25].

TDMx is a cloud-based platform for precision dosing using a Bayesian approach (lead developer: Sebastian Wicha, University of Hamburg, Germany) [26]. The software is Web-based and can be freely accessed from any device with an internet connection. No data are stored in the cloud platform. Dosing can be optimized in several ways. Using the "Probabilistic Dosing" module, a likely effective first dose can be calculated using Monte Carlo simulations. When TDM data are available, the "Bayesian Dosing" module can be used to derive precision dosing based on Bayesian forecasting using TDM data. The "Optimal Sampling" module can be used as a guide to optimal sampling time points. The population pharmacokinetic models implemented in TDMx are successively validated against clinical data (eg, [27,28]).

eLearning and Information

A total of 29 mHealth apps were identified that focus on providing patients and health care professionals with information on tuberculosis. Most of these apps provide information related to causes, risk factors, symptoms, diagnostics, treatment, or diet. The apps mainly depict treatment guidelines, provide doctors and students the opportunity to improve their skills in tuberculosis treatment, or explain the disease and its therapy to patients (see Multimedia Appendix 1). Problems were detected in many of these apps, including spelling or grammar mistakes (eg, Tuberculosis Disease [popularp, Nigeria]), wrong information (eg, in Tuberculosis [TB] [Rikki], under "description of Tuberculosis," diabetes mellitus is described), did not include up to date information based on current guidelines or presented potentially harmful information (eg, in Tuberculosis TB Home Remedies [StatesApps, USA], it is stated that "custard apple can help to cure tuberculosis to a large extent […] and rejuvenate the drugs that are delivered for curing tuberculosis").

Diagnostics

Three mHealth apps (TuberSpot, TimBre for Tuberculosis [TB], Diagnosa Tuberkulosis [TB]) supporting health care professionals and patients with a TB diagnosis were included in this review.

TuberSpot (SpotLab, Spain) is a game to identify tuberculosis bacilli in samples. This app teaches the user about shape, color, clusters, and how to differentiate tuberculosis bacilli from artifacts [10]. TimBre for Tuberculosis (TB) (Rahul Pathri, India) is a screening tool in which patients are asked to cough into their smartphone microphone and are then referred to a physician if necessary. The Indonesian app Diagnosa Tuberkulosis (TB) (Informatika Unsada, Indonesia) includes features to determine the likelihood of infection.

Other Apps

Nine mHealth apps could not be categorized within the above-mentioned categories. Their functionalities ranged from simulation of treatment costs (CAD4TB [Interactive Health Solutions, Pakistan]) [10], tracing of people that have been in contact with infected patients (eDetection [Operation Asha, India]) [10], tracking and monitoring of patients with tuberculosis (eCompliance [Operation Asha, India]), enabling researchers to identify chemical structures with activity against *M. tuberculosis* (TB mobile [Collaborative Drug, USA]) [30], facilitating data gathering by health care personnel (EdetectTB [CTMobi srls, Italy], Smart TB Puskesmas Andalas Padang [Puskesmas Andalas Padang, Indonesia], LTBI Care [WHO]), and the creation of laboratory reports (TB eHealth [iMoSyS, Malawi]).

Discussion

Principal Findings

Fifty-five mHealth apps related to tuberculosis were included in this review. In comparison to previous work on the topic [10], the amount of mHealth apps in the field of tuberculosis has more than doubled since 2016, which demonstrates how fast mHealth is advancing. Clearly, these findings are affected by differing inclusion criteria, but since they were comparable to the review by Iribarren et al [10], it is still reasonable to hypothesize that the amount of apps on the market has increased substantially over the past 3 years.

Adherence Monitoring

Adherence is a major challenge in tuberculosis therapy since the treatment length ranges from 6 months for drug-susceptible tuberculosis [4] to 18 months or more in cases of drug-resistant tuberculosis [31]. Additionally, side effects can severely limit quality of life [32]. However, adherence is particularly crucial for tuberculosis treatment since interruptions in therapy can lead to suboptimal plasma drug concentrations and consequently to the development of drug-resistant bacteria and finally treatment failure [3]. According to the WHO, cases of drug-resistant tuberculosis are increasing [1], which is partly caused by a lack of patient adherence [3]. There are several approaches to improve tuberculosis patients' adherence. One method is directly observed treatment to control drug

administration. This technique has been developed further by the introduction of VOT [4].

Studies comparing directly observed treatment and VOT show that VOT is convenient for patients and providers, time and cost-effective, widely accepted among patients and health care personnel, and flexible [6,8,32]. The WHO has thus recommended the use of VOT since 2017 [4].

Potential drawbacks of VOT could be less frequent interaction between patients and providers, and therefore a lack of side effect detection, that some patients might not have access to internet or smartphones, and that it requires a secure data and video transfer [33]. Three smartphone apps are currently available for VOT (miDOT-EMOCHA, SureAdhere, AiCure).

Individualized Dosing

A substantial number of patients still fail to respond to treatment, have a relapse of disease, or develop drug-resistant tuberculosis [1]. One of the multiple reasons for an unsuccessful treatment outcome are suboptimal plasma concentrations of tuberculosis drugs [2]. One strategy to avoid insufficient plasma concentrations, and therefore enhance cure rates, is MIPD, which is an approach to obtain individual pharmacokinetic parameters using both a priori data from population pharmacokinetic models and individually measured drug concentrations to adjust future treatment. To calculate an initial dose, patient covariates (eg, age, weight, genotype) are utilized [34-36]. Several mHealth apps have been developed as an aid for health care professionals with dosage adjustments of tuberculosis drugs. Out of the six tools currently on the market, four (DoseMeRx, MwPharm, InsightRX, TDMx) are designed for MIPD of tuberculosis medication. mHealth apps performing MIPD of tuberculosis drugs enable dosage calculation at bedside, which is very convenient for health care personnel, representing an important contribution to the improvement of tuberculosis treatment [20].

eLearning and Information About Tuberculosis

The majority of the included apps focused on education for patients and health care professionals. Although patient education is certainly of great importance [4], it should be noted that many of these apps included errors such as grammar or spelling mistakes, or communicated false or potentially harmful information. mHealth apps without a designation as a medical device must therefore be used with caution.

Based on our analysis, only 2 of the 55 apps (4%) are currently designated as medical devices (DoseMeRx, MwPharm), which is important to ensure the quality of mHealth apps in the European Union. In the future, once the new Medical Device Regulation is enforced, more apps claiming to have a medical purpose will have to be marketed as medical devices. This will likely lead to an increase in the number of high-quality apps for tuberculosis treatment.

Development of the Market

Since the last review on this matter in 2016, 31 new apps have been introduced to the market, representing a 129% increase in the total number of apps available. In 2016, there were no apps for improvement of adherence [10], whereas there are currently 8 tools featuring adherence monitoring for patients with tuberculosis, which is a substantial improvement. However, there is still room for further development. As discussed above, VOT tools are very valuable and have been shown to improve treatment outcomes in patients with tuberculosis [5-9]. However, VOT apps are labor- and cost-extensive, whereas approaches involving artificial intelligence for facial recognition and verification of drug intake involve less health care personnel and costs. If validated thoroughly, these technologies could be very valuable for improvement of patient adherence monitoring. Moreover, since 2016, 6 apps designed for dose individualization have entered the market. Tools for dose individualization should, whenever possible, include both individual covariate information as well as measured drug concentrations in order to predict the next dose with high accuracy and precision [34-36]. These MIPD approaches are superior to techniques based on merely covariate information such as body weight [34-36], and thus more tools enabling clinicians to use MIPD at bedside are needed. At present, 4 tools for MIPD are available; however, there are still tuberculosis drugs such as pretomanid that have not been covered by any MIPD app. In addition, none of the dose individualization apps takes within-subject variability into account, which has been shown to improve accuracy and precision in dose predictions [37,38] (personal communication, L Keutzer). Furthermore, the usage of “big data” for dose individualization is increasingly coming into the spotlight. As more data on patient covariates and drug concentration become available, machine-learning and artificial intelligence algorithms can be used to constantly update and improve the implemented dosing algorithms [24,25]. InsightRx includes such a feature, but more apps moving into the area of artificial intelligence are needed.

The number of apps providing patients and health care personnel with information on tuberculosis has increased by 164% (from 11 to 29 apps) since 2016 [10]. This shows that there has been a huge increase in apps providing information with which the market is saturated.

The number of apps for diagnostics or other purposes has remained rather constant in the last few years. In the review from 2016 [10], 8 apps for the screening and surveillance of patients with tuberculosis were discovered, and 8 such tools are currently available. Both currently and in 2016 [10], 3 apps aiding with tuberculosis diagnosis were on the market.

One major gap that was identified during this work was the availability of languages. Most apps are only available in English (39/55, 71%). This is problematic since the highest prevalence of tuberculosis cases includes nonEnglish-speaking countries such as China or Indonesia [1]. This points to a clear need to either add additional language availability to existing apps or develop new apps in other languages.

Limitations

There are several limitations to this work. One drawback is that apps that are not free of charge or requiring a license were not purchased and consequently not scrutinized with respect to their functionalities. However, such apps were evaluated in this work and were described based on information retrieved from the app store or accompanying publications. Although it was not the aim of this review, the apps were not rated and ranked regarding

their functionality and quality, as suggested in previous work [39], but were rather merely described in order to provide an overview of existing mHealth apps currently on the market. Furthermore, we might have missed including some available apps because the databases were only searched in English. In addition, the documentation of the number of downloads (see Multimedia Appendix 1) could be a misleading indicator of an app's success, especially if the app cannot be used without payment of a fee. The number of downloads also does not account for how long the app has been available.

Conclusions

Although the importance of the various mHealth apps currently on the market, such as tools for VOT, MIPD, or simplification of data management, for the improvement of tuberculosis treatment cannot be discounted, the majority (53%) of apps identified in this work were merely focused on providing information about tuberculosis, and many of these exhibited issues regarding spelling, grammar, or correctness of the information provided. Thus, this review demonstrates a need for more mHealth apps of high quality supporting tuberculosis treatment.

Authors' Contributions

LK, SW, and US conducted the app review, wrote the manuscript, and revised the manuscript.

Conflicts of Interest

SW is the lead developer of TDMx. LK and US declare no conflicts of interest.

Multimedia Appendix 1
Tables of mHealth applications included in the review.
[DOCX File , 45 KB - mhealth_v8i4e17246_app1.docx]

References

1. World Health Organization. Global tuberculosis report 2019. Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO; 2019.
2. Pasipanodya JG, McIlleron H, Burger A, Wash PA, Smith P, Gumbo T. Serum drug concentrations predictive of pulmonary tuberculosis outcomes. J Infect Dis 2013 Nov 01;208(9):1464-1473 [FREE Full text] [doi: 10.1093/infdis/jit352] [Medline: 23901086]
3. Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clin Infect Dis 2016 Dec 01;63(7):e147-e195 [FREE Full text] [doi: 10.1093/cid/ciw376] [Medline: 27516382]
4. World Health Organization. Guidelines For Treatment Of Drug-susceptible Tuberculosis And Patient Care: 2017 Update. Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO; 2017.
5. Salehitali S, Noorian K, Hafizi M, Dehkordi AH. Quality of life and its effective factors in tuberculosis patients receiving directly observed treatment short-course (DOTS). J Clin Tuberc Other Mycobact Dis 2019 May;15:100093 [FREE Full text] [doi: 10.1016/j.jctube.2019.100093] [Medline: 31720420]
6. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
7. Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010 Jul;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
8. Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016 May;20(5):588-593. [doi: 10.5588/ijtld.15.0738] [Medline: 27084810]
9. Garfein RS, Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. Int J Tuberc Lung Dis 2015 Sep;19(9):1057-1064 [FREE Full text] [doi: 10.5588/ijtld.14.0923] [Medline: 26260824]
10. Iribarren SJ, Schnall R, Stone PW, Carballo-Diéguez A. Smartphone Applications to Support Tuberculosis Prevention and Treatment: Review and Evaluation. JMIR Mhealth Uhealth 2016 May 13;4(2):e25 [FREE Full text] [doi: 10.2196/mhealth.5022] [Medline: 27177591]
11. World Health Organization. Mhealth: New Horizons For Health Through Mobile Technologies: Second Global Survey On Ehealth (global Observatory For Ehealth). Geneva: World Health Organization; 2011.
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

13. Emocha Mobile Health Inc. Emocha Mobile Health. 2019. URL: https://www.emocha.com/home [accessed 2019-02-25]
14. SureAdhere Mobile Technology, Inc. SureAdhere Mobile Technology. 2016. URL: http://www.sureadhere.com [accessed 2019-03-18]
15. AiCure. AiCure. 2019. URL: https://aicure.com/ [accessed 2019-03-18]
16. Hanina A. Automated DOT® - Facial Recognition Software To Conirm Medication Adherence On Mobile Devices. 2014. URL: https://tinyurl.com/rx7fnac [accessed 2019-03-18]
17. Wisepill Technologies. evriMED. URL: https://www.wisepill.com/evrimed [accessed 2019-04-14]
18. DoseMe Pty Ltd. DoseMeRx - Precision Dosing: A smarter way to dose. 2019. URL: https://doseme-rx.com/ [accessed 2019-04-14]
19. Bourne DAW. Pharmacokinetic Software. 2017. URL: https://www.pharmpk.com/soft.html [accessed 2019-04-14]
20. Kumar AA, Burgard M, Stacey S, Sandaradura I, Lai T, Coorey C, et al. An evaluation of the user-friendliness of Bayesian forecasting programs in a clinical setting. Br J Clin Pharmacol 2019 Oct;85(10):2436-2441. [doi: 10.1111/bcp.14066] [Medline: 31313335]
21. Stojanova S, Luque S. Available Dose Adaptation Tools for Clinical Use. In: Udy AA, Roberts JA, Lipman J, editors. Antibiotic Pharmacokinetic/pharmacodynamic Considerations In The Critically Ill. Singapore: Springer Singapore; 2018:176-177.
22. Mediware a.s. MwPharm++. 2014. URL: http://www.mediware.cz/en/mwpharm-plus-plus/ [accessed 2019-02-22]
23. Fuchs A, Csajka C, Thoma Y, Buclin T, Widmer N. Benchmarking therapeutic drug monitoring software: a review of available computer tools. Clin Pharmacokinet 2013 Jan;52(1):9-22. [doi: 10.1007/s40262-012-0020-y] [Medline: 23196713]
24. InsightRX, Inc. InsightRX - Precision Medicine Made Simple. 2019. URL: https://insight-rx.com/ [accessed 2019-02-22]
25. Keizer RJ, Mangat R, Goswami S. Experiences in applied clinical pharmacometrics: challenges, recommendations, and research opportunities. 2018 Presented at: Population Approach Group Europe (PAGE) 27; June 2018; Montreux URL: https://www.page-meeting.org/?abstract=8758
26. Wicha SG, Kees MG, Solms A, Minichmayr IK, Kratzer A, Kloft C. TDMx: a novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine. Int J Antimicrob Agents 2015 Apr;45(4):442-444. [doi: 10.1016/j.ijantimicag.2014.12.010] [Medline: 25631677]
27. Broeker A, Nardecchia M, Klinker KP, Derendorf H, Day RO, Marriott DJ, et al. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. Clin Microbiol Infect 2019 Oct;25(10):1286.e1-1286.e7. [doi: 10.1016/j.cmi.2019.02.029] [Medline: 30872102]
28. Barras MA, Serisier D, Hennig S, Jess K, Norris RLG. Bayesian Estimation of Tobramycin Exposure in Patients with Cystic Fibrosis. Antimicrob Agents Chemother 2016 Nov;60(11):6698-6702 [FREE Full text] [doi: 10.1128/AAC.01131-16] [Medline: 27572411]
29. TDMx-model-supported Therapeutic Drug Monitoring for Precision Dosing. URL: http://www.tdmx.eu/ [accessed 2020-03-20]
30. Ekins S, Clark AM, Sarker M. TB Mobile: a mobile app for anti-tuberculosis molecules with known targets. J Cheminform 2013 Mar 06;5(1):13. [doi: 10.1186/1758-2946-5-13] [Medline: 23497706]
31. World Health Organization. WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: World Health Organization; License: CC BY-NC-SA 3.0 IGO; 2019.
32. Nsengiyumva NP, Mappin-Kasirer B, Oxlade O, Bastos M, Trajman A, Falzon D, et al. Evaluating the potential costs and impact of digital health technologies for tuberculosis treatment support. Eur Respir J 2018 Nov;52(5):1801363. [doi: 10.1183/13993003.01363-2018] [Medline: 30166325]
33. Mehregany M, Saldivar E. Chapter 2: Opportunities and obstacles in the adoption of mHealth. In: Krohn R, Metcalf D, editors. mHealth from smartphones to smart systems. Chicago: Health Information and Management Systems Society (HIMSS); 2012:7-24.
34. Sheiner LB, Beal S, Rosenberg B, Marathe VV. Forecasting individual pharmacokinetics. Clin Pharmacol Ther 1979 Sep;26(3):294-305. [Medline: 466923]
35. Keizer RJ, Ter Heine R, Frymoyer A, Lesko LJ, Mangat R, Goswami S. Model-Informed Precision Dosing at the Bedside: Scientific Challenges and Opportunities. CPT Pharmacometrics Syst Pharmacol 2018 Dec;7(12):785-787. [doi: 10.1002/psp4.12353] [Medline: 30255663]
36. Darwich AS, Ogungbenro K, Vinks AA, Powell JR, Reny J, Marsousi N, et al. Why has model-informed precision dosing not yet become common clinical reality? lessons from the past and a roadmap for the future. Clin Pharmacol Ther 2017 May;101(5):646-656. [doi: 10.1002/cpt.659] [Medline: 28182269]
37. Abrantes JA, Jönsson S, Karlsson MO, Nielsen EI. Handling interoccasion variability in model-based dose individualization using therapeutic drug monitoring data. Br J Clin Pharmacol 2019 Jun;85(6):1326-1336. [doi: 10.1111/bcp.13901] [Medline: 30767254]
38. Wicha SG, Hennig S. Handling inter-occasion variability in model implementation for Bayesian forecasting: A comparison of methods and metrics. 2018 Presented at: Population Approach Group Europe (PAGE) 27; June 2018; Montreux URL: https://www.page-meeting.org/?abstract=8580

39. 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]

Abbreviations

EHR: electronic health record
mHealth: mobile health
MIPD: model-informed precision dosing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TDM: therapeutic drug monitoring
VOT: video observed treatment
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.11.19; peer-reviewed by A Meyer, R Garfein; comments to author 16.12.19; revised version received 10.02.20; accepted 22.02.20; published 21.04.20.

Please cite as:
Keutzer L, Wicha SG, Simonsson USH
Mobile Health Apps for Improvement of Tuberculosis Treatment: Descriptive Review
JMIR Mhealth Uhealth 2020;8(4):e17246
URL: http://mhealth.jmir.org/2020/4/e17246/
doi:10.2196/17246
PMID:32314977

Original Paper

Adoption and Attitudes of eHealth Among People Living With HIV and Their Physicians: Online Multicenter Questionnaire Study

Christine Jacomet[1], MD; Roxana Ologeanu-Taddei[2], PhD; Justine Prouteau[1], MSci; Céline Lambert[3], MSci; Françoise Linard[4], MD; Pascale Bastiani[5], AAS; Pierre Dellamonica[6], MD, PhD

[1]Infectious Diseases Department, Clermont-Ferrand University Hospital, Clermont-Ferrand, France

[2]Systèmes d'Information – Recherche en Management & Polytech, Montpellier University, Montpellier, France

[3]Delegation Recherche Clinique et Innovation, Clermont-Ferrand University Hospital, Clermont-Ferrand, France

[4]Infectious Disease Department, Assistance Publique Hôpitaux de Paris – Sorbonne Université, Paris, France

[5]Agissons Contre le Sida, Nice, France

[6]Cote d'Azur University, Nice, France

Corresponding Author:
Christine Jacomet, MD
Infectious Diseases Department
Clermont-Ferrand University Hospital
Hôpital Gabriel Montpied
58 rue Montalembert
Clermont-Ferrand
France
Phone: 33 473754931
Email: cjacomet@chu-clermontferrand.fr

Abstract

Background: The development of electronic health (eHealth) has offered the opportunity for remote care provision. eHealth addresses issues for patients and professionals favoring autonomy and compliance, respectively, while fostering closer links both between patients and health care professionals and among health care professionals themselves.

Objective: The aim of this study was to analyze the patterns of use, benefits, and perceived obstacles in eHealth among people living with HIV (PLHIV) and their caring physicians at hospitals.

Methods: An online multicenter observational survey was conducted October 15-19, 2018 in 51 medical units across France by means of self-administered questionnaires to collect sociodemographic and medical data, and perceptions of eHealth. Multiple correspondence analysis followed by mixed unsupervised classification were performed to analyze data of the respondents.

Results: A total of 279 PLHIV and 219 physicians responded to all parts of the questionnaire. Three groups of PLHIV were identified based on multivariate analysis. Group 1 comprised “eHealth believers” (121/279, 43.4%), who were more frequently above 60 years old and more likely to be receiving treatments other than antiretrovirals. Group 2, the “technology skeptics” (86/279, 30.8%), comprised more women with at least one child. Group 3, the “internet adopters” (72/279, 25.8%), were more frequently under 49 years of age, men who have sex with men, and more likely to use mobile apps for obtaining wellness/health information and related subjects. Three groups of physicians also emerged. Group 1 comprised those “strongly confident in eHealth” (95/219, 43.4%), who more frequently used mobile apps for wellness/health information and were more likely to accept prescription assistance software. Group 2 comprised physicians “strongly opposed to eHealth” (80/219, 36.5%), frequently asserting that eHealth challenges confidentiality. Group 3 were “open to eHealth” (44/219, 20.1%), comprising a higher proportion of infectious disease specialists, and were more likely to believe that medical apps are useful for patient education and information. No link was found between the groups of PLHIV and physicians.

Conclusions: The literature on eHealth mainly classifies people as enthusiasts and skeptics; however, we identified a third profile among both PLHIV and physicians, albeit without a direct link between them. For PLHIV, this third group is attentive to eHealth for improving their health condition, and for physicians, this group considers eHealth to offer benefits to patients and their own practice.

(JMIR Mhealth Uhealth 2020;8(4):e16140) doi:10.2196/16140

KEYWORDS

survey; HIV; eHealth; internet for information retrieval; health applications; connected objects; telemedicine; collection of digitized personal information

Introduction

The World Health Organization defines eHealth (electronic health) as the application of information and communication technologies to all activities connected to health [1]. As eHealth offers remote care provision, such technologies can favor patient autonomy and compliance, while fostering closer links between patients and health care professionals, as well as among health care professionals themselves [2]. Data processing via computers, connected devices (eg, mobile phones, smartphones, tablets) [3-6], and connected objects has witnessed exponential development in recent years involving a broad diversity of actors [7].

The emergence and spread of HIV infection paralleled the development of eHealth within a short time period. In the last three decades, there has been a rise in freely accessible online journals, eLearning, and massive open online courses, which now supersede the printed page as the primary source of scientific knowledge for health care professionals [8]. Moreover, the development of new channels of communication between patients and their health care professionals poses a threat to medical confidentiality [9]. Improvements in the traceability and storage of medical data are raising ethical questions [10], while the economic value of such information has increased. In response, the General Data Protection Regulation was approved on April 14, 2016 by the European Parliament, and came into effect in France on May 25, 2016, which restricts the use of such data [11].

At the same time, France devoted 11% of its gross domestic product (GDP; 198.5 billion Euros [214 billion USD]) to health [12]. Health costs are increasing faster than the GDP (+1.6%), calling for an economic and ethical review of public health spending. The French health care system must meet many challenges, including the increased prevalence in chronic illness (estimated to affect 37% of the French population in 2014 [13]), an aging society, medically deprived areas, shortage of caregivers, and increased costs of care. One approach to maintaining quality health care is through reorganization. The use of eHealth tools to follow patients remotely and coordinate their care teams should be a step in this direction.

Analysis of the literature in the field of eHealth, and in particular of connected objects [14], considering current trends, utility, and limits shows that eHealth technologies are perceived in two distinct, but specific, ways. The first is clear enthusiasm and especially high hopes for new health technologies, and the other is criticism of the development of these technologies, in some cases questioning their usefulness and highlighting the potential danger to confidentiality, while destroying the physician-patient relationship through the dehumanization of care.

In the present study, we analyzed the behaviors, benefits, and barriers perceived by people living with HIV (PLHIV) and their physicians via online self-administered questionnaires (for details on questionnaire development, see personal communication with the first author CJ) to determine whether any additional profiles of eHealth perception exist besides these two sharply opposing viewpoints, and whether there is a link between the perception profiles of PLHIV and physicians.

Methods

Recruitment and Questionnaire Design

We conducted a multicenter observational single-week survey involving all of the patients infected with HIV who were consulted during the week of October 15-19, 2018 in French hospitals via regional coordination structures. The patients were involved in the design, implementation, reporting, and dissemination of this survey.

The inclusion criterion was aged over 18 years. Exclusion criteria were inability to reply to the questionnaire, inability to speak French, and refusal to participate. This observational survey was compliant with the MR003 reference specification [15]. The protocol was filed with the French data protection agency (CNIL, No. M18009).

Data Collection

Study data were collected and managed using Research Electronic Data Capture (REDCap) electronic data capture tools hosted at the teaching hospital Centre Hospitalier Universitaire (CHU) de Clermont-Ferrand (Clermont-Ferrand, France) [16,17]. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing an intuitive interface for validated data capture, audit trails for tracking data manipulation and export procedures, automated export procedures for seamless data downloads to common statistical packages, and procedures for data integration and interoperability with external sources. Both patients and health care professionals generated their own anonymous data using the REDCap app created for this survey. Access to the survey was provided through a QR code or internet link, followed by entry of an access code. The data were accessible only to the staff of the Clinical Research Department of CHU Clermont-Ferrand, which takes responsibility for their security and confidentiality.

Data Analysis

All statistical analyses were performed using Stata software (version 13, StataCorp, College Station, TX, USA) and R 3.3.3 software (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-sided with a type I error set at .05. Baseline characteristics (physicians and patients) are presented as frequencies and associated percentages for categorical parameters, and as mean (SD) or median (interquartile range) for continuous variables, according to the data distribution.

Multiple correspondence analysis followed by mixed unsupervised classification (k-means clustering applied to the partition obtained from an ascending hierarchical classification using Ward distance) were implemented to (i) examine the

relations between the modalities of the variables and (ii) determine PLHIV profiles (clusters of individuals sharing closely similar characteristics). For these analyses, the variables were chosen according to univariate results, clinical relevance, and the data distribution (parameters always present or always absent were not considered), which included sociodemographic and medical characteristics of PLHIV, patterns of behavior and opinions regarding information retrieval via the internet and social media, representations of using apps and connected objects, and opinions on distance consultation and on collection of personal health data. The association between the groups obtained and their characteristics were compared by the Chi-squared test.

The same analyses were performed for the physicians with the same variables. The questionnaire for physicians also addressed perceptions of their patients' use of eHealth provisions.

Finally, the clinically identified groups of PLHIV were compared with those of physicians by a Chi-squared test to examine similarities between profiles.

Results

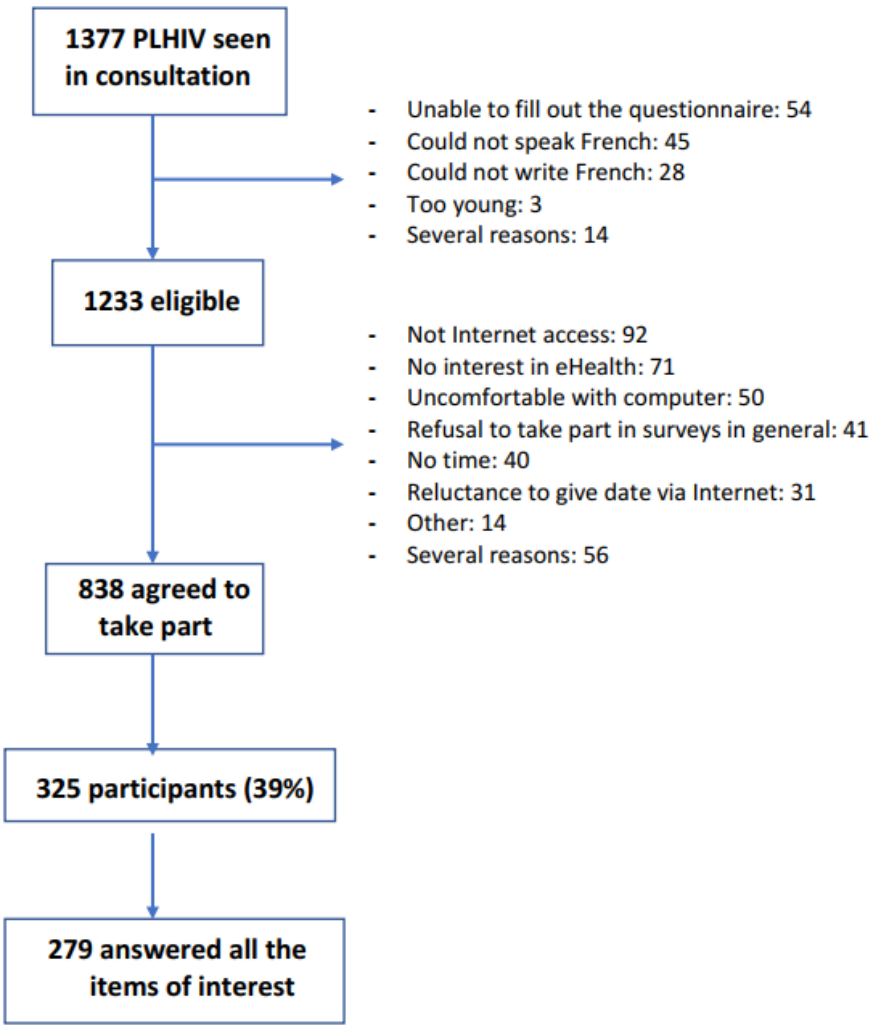
Overall Response

The survey was conducted in 51 medical units throughout France after consent was obtained from 255 physicians who consulted with 1377 patients during the study period. Among these patients, 144 were not eligible, 395 refused to take part in the study, and 838 agreed to participate and were given access to the online survey. Overall, 325/838 (38.8%) patients followed in 46 centers, including 191 (58.8%) at teaching hospitals, filled out the online questionnaire; 117 (36.0%) of the patients were residents of the Ile-de-France region. Of the patients, 279 answered all of the questions of interest, and their questionnaires were ultimately included in the analysis. These patients were comparable to the 46 excluded patients with regard to the variables used for analyzing patterns of use and behavior for information retrieval via the internet and social media, use of apps and connected objects, distance consultations, and collection of personal health data.

Nearly all the physicians who replied to the questionnaire (219/227, 96.5%) were included in the factorial analysis.

The reasons for noninclusion and refusal are given in Figure 1.

Figure 1. Flowchart of the 279 people living with HIV (PLHIV) who answered all items of interest.



Sociodemographic and Clinical Characteristics of Patients and Physicians

Table 1 provides a summary of the sociodemographic characteristics, and Table 2 summarizes the key clinical and care path characteristics of the patients included in the analysis.

The 279/325 (85.8%) PLHIV who responded to all of the items included in the analysis (no missing data) were mostly middle-aged males born in France. Half of them lived with a partner. The majority had high school or higher education, and were approximately evenly split between those with a stable job and those in a precarious employment situation. They had been living with HIV for a mean of 17 years and had been receiving antiretroviral treatment for a mean of 14 years with an undetectable viral charge (<50 copies/ml) in most cases and an average CD4 count of 620/mm^3; less than half of the patients were also receiving associated treatments. Most of the patients consulted their general practitioner 1-3 times per year and their HIV specialist twice a year, and less than a third did not consult any other specialist physicians. At the time the questionnaire was filled out, the average fitness of the patients was self-evaluated to be relatively high on a visual analog scale.

Of the 255 physicians who agreed to take part in the survey, 227 (89.0%) answered the medical questionnaire, and 219 (85.9%) answered all of the questions of interest for the analysis. The majority worked full-time at the hospital, mainly in an infectious disease department; the other physician characteristics are summarized in Table 3.

Table 1. Sociodemographic characteristics of people living with HIV (N=279).

Characteristic	Value
Age (years), mean (SD)	53 (12)
Gender, n (%)	
Male	199 (71.3)
Female	80 (28.7)
Live with a partner, n (%)	142 (50.9)
Sexual orientation, n (%)	
Heterosexual	127 (45.5)
Homosexual	120 (43.0)
Other	11 (3.9)
Decline to reply	21 (7.5)
At least one child, n (%)	119 (42.7)
Country of birth=France, n (%)	218 (78.1)
Region of birth=Paris/Ile-de-France, n (%)	55 (19.7)
Region of residence=Paris/Ile-de-France, n (%)	89 (31.9)
High school or above education level, n (%)	184 (65.9)
Occupation, n (%)	
Stable job	130 (46.6)
Retired	61 (21.9)
Invalid	35 (12.5)
Job-seeker	29 (10.4)
Other	24 (8.6)
Precarity	
EPICES[a] precarity score, median (IQR)	25 (15-46)
Precarious, n (%)	127 (45.5)
Meeting places, n (%)	
Bars, clubs without sex	81 (29.0)
Sex clubs	40 (14.3)
Through geolocating dating sites	57 (20.4)

[a]EPICES: Evaluation of Precarity and Inequalities in Health Examination Centers.

Table 2. Clinical and care path characteristics of people living with HIV (N=279).

Characteristic	Value
Medical characteristics	
Last viral load undetectable, n (%)	255 (91.4)
Last CD4 count (mm^3), median (IQR)	600 (400-842)
Years since HIV infection detected, mean (SD)	17 (10)
Years of antiviral treatment, mean (SD)	14 (8)
Tobacco use, n (%)	
Active	76 (27.2)
Nonsmoker or exsmoker	203 (72.8)
Alcohol consumption more than once a week, n (%)	136 (48.7)
Consumption of recreational drugs, n (%)	
Active or former user	55 (19.7)
Non-user	224 (80.3)
Lipodystrophy, n (%)	56 (20.1)
Associated treatments, n (%)	
Presence	123 (44.1)
Anti-HBP[a]	57 (20.4)
Psychiatric	43 (15.4)
Cardiovascular	27 (9.7)
Anti-diabetes	24 (8.6)
Hyperlipidemia	15 (5.4)
Bone and joints	15 (5.4)
Neurological	13 (4.7)
Hepatitis B or C	9 (3.2)
Renal	8 (2.9)
Cancer	5 (1.8)
Care path, n (%)	
Follow up in teaching hospital	239 (85.7)
No consultation in general medical practice	39 (14.0)
One to three consultations in general medical practice	163 (58.4)
Four or more consultations in general medical practice	77 (27.6)
One or more HIV-specific consultations per year	159 (57.0)
Three or more HIV-specific consultations per year	120 (43.0)
No other specialized consultations	81 (29.0)
One to three other specialized consultations	150 (53.8)
Four or more other specialized consultations	48 (17.2)
Fitness VAS[b], mean (SD)	78 (19)

[a]HBP: high blood pressure.

[b]VAS: visual analog scale (0-100).

Table 3. Sociodemographic characteristics of physicians (N=219).

Sociodemographic characteristic	Value
Age (years), mean (SD)	48 (10)
Gender, n (%)	
Male	94 (42.9)
Female	125 (57.1)
Medical specialty, n (%)	
Infectious diseases	158 (72.1)
General practitioner	37 (16.9)
Internal medicine	13 (5.9)
Dermatology	4 (1.8)
Hematology	2 (0.9)
Gastroenterology	1 (0.5)
Geriatrics	1 (0.5)
Immunology	1 (0.5)
Psychiatry	1 (0.5)
Public health	1 (0.5)
Hospital practice, n (%)	
Full time	159 (72.6)
Part time	60 (27.4)
Department of practice in Paris/Ile-de-France, n (%)	80 (36.5)

Patient Profiles on eHealth Perceptions

Overall Response

The analysis was focused on the 279/325 PLHIV for whom there were no missing data among the variables selected for the analysis. Multiple correspondence analysis identified 3 groups (G) of patients: G1 (121/279, 43.4%), termed “eHealth believers;” G2 (86/279, 30.8%), termed “technology skeptics;” and G3 (72/279, 25.8%), termed “internet adopters” (Figure 2).

Their main epidemiological, medical, and eHealth perception characteristics are summarized by group in Table 4 and Table 5 (see Multimedia Appendix 1 for detailed results). Only statistically different variables are presented.

Figure 2. Factorial analysis of people living with HIV by sociodemographic and medical characteristics, and their answers to 113 questions concerning eHealth: searching for information on the internet and social media, collection of digitized personal information, and mHealth apps and connected objects for health/wellness.

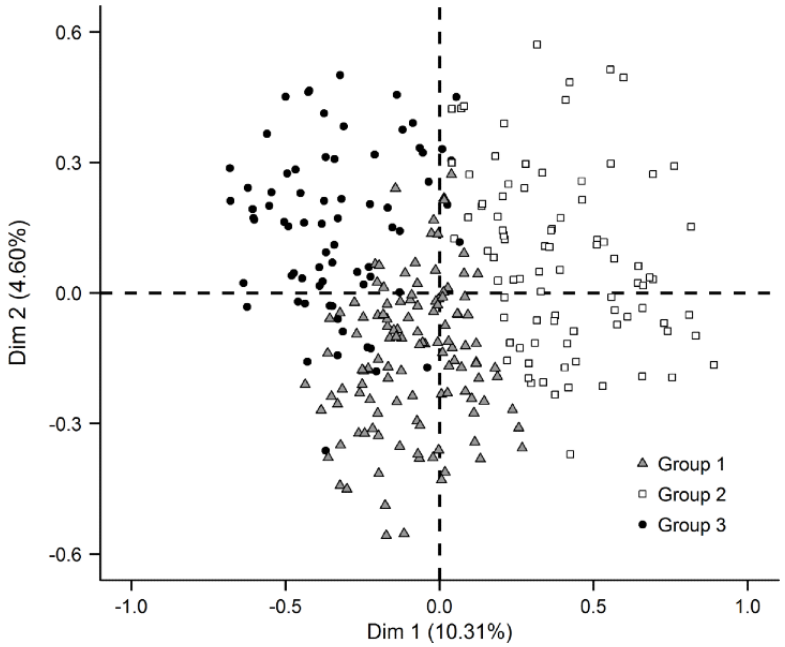


Table 4. Sociodemographic and medical characteristics, internet and social media, app use, and connected objects for three groups of people living with HIV obtained by mixed unsupervised classification.

Characteristic	Total (N=279)	Group 1 (n=121)	Group 2 (n=86)	Group 3 (n=72)
Sociodemographic characteristics, n (%)				
Aged <49 years	106 (38.0)	33 (27.3)	26 (30.2)	47 (65.3)
Aged >60 years	79 (28.3)	48 (39.7)	25 (29.1)	6 (8.3)
Male	199 (71.3)	90 (74.4)	51 (59.3)	58 (80.6)
MSM[a]	120 (43.0)	53 (43.8)	21 (24.4)	46 (63.9)
At least one child	119 (42.7)	50 (41.3)	48 (55.8)	21 (29.2)
Higher education	184 (65.9)	73 (60.3)	51 (59.3)	60 (83.3)
Frequent use of geolocating dating sites	57 (20.4)	14 (11.6)	10 (11.6)	33 (45.8)
Medical characteristics, n (%)				
Duration of treatment<9 years	93 (33.3)	32 (26.4)	22 (25.6)	39 (54.2)
Receiving treatments other than antiretroviral	123 (44.1)	65 (53.7)	40 (46.5)	18 (25.0)
History of illegal drug use	55 (19.7)	18 (14.9)	9 (10.5)	28 (38.9)
Internet and social media, n (%)				
Used the internet in the last 12 months to look for information or advice on health or wellness	154 (55.2)	64 (52.9)	32 (37.2)	58 (80.6)
Changed the way they attend to their health/wellness after these searches	74 (26.5)	31 (25.6)	5 (5.8)	38 (52.8)
Possess a social media account (eg, Facebook, Twitter)	174 (62.4)	55 (45.5)	47 (54.7)	72 (100.0)
No longer as trusting after confidentiality problems	66 (23.7)	21 (17.4)	17 (19.8)	28 (38.9)
App use, n (%)				
Currently use mobile apps for monitoring physical activity	45 (16.1)	8 (6.6)	8 (9.3)	29 (40.3)
Would be willing to use an app if it was recommended by a physician	199 (71.3)	94 (77.7)	41 (47.7)	64 (88.9)
Would be willing to use an app if it was recommended by an associate	51 (18.3)	13 (10.7)	6 (7.0)	32 (44.4)
Would be willing to use an app if they could manage on their own	102 (36.6)	35 (28.9)	50 (58.1)	17 (23.6)
Think an ideal app should help follow adverse effects of medical drugs	208 (74.6)	110 (90.9)	41 (47.7)	57 (79.2)
Think an ideal app should help follow vaccinations	212 (76.0)	107 (88.4)	45 (52.3)	60 (83.3)
Think an ideal app should help get in touch with other patients	88 (31.5)	46 (38.0)	16 (18.6)	26 (36.1)
Trust an app more than a health care professional	26 (9.3)	19 (15.7)	2 (2.3)	5 (6.9)
Connected objects, n (%)				
Possess a connected object	61 (21.9)	16 (13.2)	11 (12.8)	34 (47.2)
Could be persuaded to have connected objects and use them if medical insurance schemes reduced their contributions as an incentive	88 (31.5)	40 (33.1)	10 (11.6)	38 (52.8)

[a]MSM: men who have sex with men.

Table 5. Comparison of the three groups of people living with HIV obtained by mixed unsupervised classification.

Characteristic	Total (N=279)	Group 1 (n=121)	Group 2 (n=86)	Group 3 (n=72)
Telemedicine, n (%)				
In favor of consultations by video conference	166 (59.5)	84 (69.4)	16 (18.6)	66 (91.7)
Would prefer to use distance consultation to get a new prescription for treatment	207 (74.2)	95 (78.5)	45 (52.3)	67 (93.1)
Would prefer to use distance consultation to consult for health problems that seem minor (eg, sore throat, cold)	123 (44.1)	63 (52.1)	15 (17.4)	45 (62.5)
Would prefer to use distance consultation to monitor evolution of their HIV infection	92 (33.0)	49 (40.4)	8 (9.3)	35 (48.6)
Think having a free internet terminal in the medical unit where they can enter data directly into their medical files before consultation would be a good thing	124 (44.4)	73 (60.3)	11 (12.8)	40 (55.6)
Collection of personal data, n (%)				
Think their personal data might be misused	167 (59.9)	65 (53.7)	63 (73.3)	39 (54.1)
Think the law adequately oversees the collection and use of personal data	99 (35.5)	54 (44.6)	11 (12.8)	34 (47.2)
Think artificial intelligence will speed progress towards more individualized diagnosis and treatment	159 (57.0)	88 (72.7)	19 (22.1)	52 (72.2)
Would like to have health digital safe space on a dedicated site hosted by a health data organization	94 (33.7)	41 (33.9)	16 (18.6)	37 (51.4)
eHealth, n (%)				
Think the development of eHealth is a good thing	197 (70.6)	109 (90.1)	21 (24.4)	67 (93.1)
Think the development of eHealth would be efficient for improving coordination among different health care practitioners	226 (81.0)	104 (86.0)	52 (60.4)	70 (97.2)
Think the development of eHealth would be efficient for reducing travel	135 (48.4)	69 (57.0)	14 (16.3)	52 (72.2)
Think the development of eHealth would be efficient for servicing medically deprived areas	157 (56.3)	74 (61.2)	34 (39.5)	49 (68.1)
Think the development of eHealth would be efficient for reducing the social Security burden	125 (44.8)	61 (50.4)	19 (22.1)	45 (62.5)

Group 1: eHealth Believers

The 121 PLHIV grouped as “eHealth believers” were most often aged over 60 years, and were more likely to be receiving treatments other than antiretroviral drugs. Compared to the patients in the other two groups, fewer of the patients in Group 1 had Facebook, Twitter, Instagram, LinkedIn, or other social media accounts, and fewer had downloaded mobile apps for wellness, health, or physical activity monitoring. More of these patients agreed that an ideal app should first show their vaccinations and adverse effects to their drugs, and then provide help with overall psychological wellness, monitor their physical state, and show their biological HIV status, history of HIV-related biology reports, and history of antiretroviral treatments. More of the patients in this group were also in contact with other individuals with the same interests, and trusted an app more than a health care professional. Many of them agreed that the collection of personal health data would increase in the coming years, that this trend would help to improve the quality of patient care and follow up, and that it was an acceptable price to be paid to gain benefits from the apps. Many feared that their personal health data might be misused, but more of them thought that artificial intelligence would progress toward more individualized diagnosis and treatments. In general, the great majority agreed that eHealth was a good thing, that it was efficient for coordination among health care professionals, and for monitoring the evolution of their HIV infection more regularly and more rapidly, allowing them greater autonomy, improved medical care and treatment, and reducing the societal burden.

Group 2: Technology Skeptics

The 86 PLHIV classified as “technology skeptics” were more often women, and more frequently had at least one child compared to patients in the other two groups. Fewer had smartphones, but more of these patients used mobile apps they had chosen themselves and not on the advice of a friend, physician, pharmacist, associate, or another patient. Fewer were comfortable with technology or trusted it. More had no time to use apps, were skeptical about their scientific value, and found them stressful. Fewer would be persuaded to use connected objects by their medical insurance scheme offering lower contributions. More of these patients were unwilling to share their data, including with their physician, and were less often in favor of distance consultations for serious health problems, to address an intimate subject, renew a prescription, ask for a

medical certificate or for advice, monitor their HIV evolution, or get an emergency consultation. More of them would not want to use an internet terminal at their care facility to enter data directly into their hospital files. More of them were worried about the collection of personal health information and felt that the law did not oversee such collection to a sufficient degree. Concerning the digital safety of health information, more of these patients preferred hosting of apps by a health insurance organization, and considered the development of eHealth as a bad trend overall.

Group 3: Internet Adopters

The 72 PLHIV classified as "internet adopters" were most often aged under 49 years, male, and men who have sex with men, most of whom had higher education and stable jobs. Patients in this group were more likely to frequent bars and sex clubs and to use dating sites with geolocating apps, and to have consumed illegal drugs. They had been infected with HIV for less than 12 years, and had been receiving treatment for less than 9 years.

The main characteristic of this group was that they are regular internet users, including following an association connected with HIV via social media, and using mobile apps for wellness, health, or physical activity monitoring. More of them agreed that an ideal app should help them monitor their physical activity and keep track of their appointments. Fewer thought the apps intruded too much into their lives, and more used connected objects. Compared to the other two groups, more of these patients were willing to communicate their results to their general practitioner and their specialist physician, and more of them had already used email for such contact. Concerning the digital safety of health information, more of them would choose a health data host. In general, more of these patients considered that eHealth was useful overall, especially for monitoring their health indicators and reducing travel.

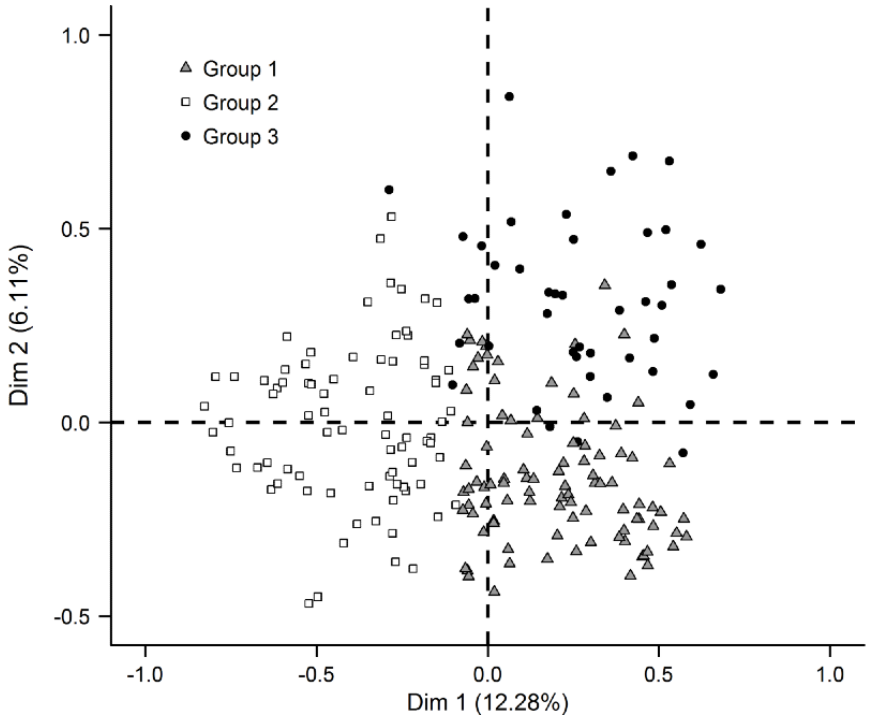
Physician Profiles on eHealth Perceptions

Overall Response

The analysis included 219 physicians for whom no data were missing among the variables selected for analysis. Multiple correspondence analysis identified three groups (G): G1 (95/219, 43.4%), termed "strongly confident in eHealth;" G2 (80/219, 36.5%), termed "strongly opposed to eHealth;", and G3 (44/219, 20.1%), termed "open to eHealth" (Figure 3).

The eHealth perception characteristics of the physicians are summarized in Multimedia Appendix 2.

Figure 3. Factorial analysis of physicians by sociodemographic and medical characteristics, and their answers to 53 questions concerning eHealth: searching for information on the internet and social media, collection of digitized personal information and telemedecine, and mHealth apps and connected objects for health/wellness.



Group 1: Strongly Confident in eHealth

The 95 physicians "strongly confident in eHealth" were more often themselves users of wellness/health apps and more often reported having patients that actively use mobile apps. Compared to the other two groups, more of the physicians in Group 1 stated that they would like to see connected objects for health become generalized, and think that connected objects should be financed by health insurance schemes and that their use would reduce health costs. In addition, since the advent of new computerized medical information systems in their hospitals, these physicians largely agreed that they have better tools to work with, that diagnostic aid is a reality, and medical decision making is facilitated. More of them use computer-assisted prescribing, consider patient safety is improved, and that the eHealth provision favors a transfer of skills among professionals. More of them think that eHealth

facilitates communication among physicians and are keen to conduct consultations by video conference.

Group 2: Strongly Opposed to eHealth

The 80 physicians "strongly opposed to eHealth" were more likely not to know their patients' habits concerning connected objects. More of them agreed that eHealth challenges the confidentiality of medical information, consider a risk for data security and quality, and that eHealth destroys the care system. More of them disagreed that eHealth will reduce the number of consultations, will improve the quality and efficacy of patient care, or embody technical progress. More of the physicians in this group were against the collection of personal health information, and disagreed that it will improve patient care and follow up. More of these physicians were also mistrustful and uneasy about the potential uses of personal information, and did not consider that the law appropriately oversees the collection of health data.

Group 3: Open to eHealth

The 44 physicians "open to eHealth" were more often infectious disease specialists compared to the specialties of physicians in the other two groups. More of these physicians considered that mobile medical apps are useful for informing and educating their patients, assisting clinical decision making, facilitating online entering of data in medical files, and improving patient monitoring. More of them considered that connected objects enable better quality of data for care, and that eHealth facilitates communication with paramedics. More of them asserted that the development of eHealth is a good trend overall, because it will be efficient in improving coordination among health care professionals, following patients' health indicators more closely, improving the quality of care and treatment, servicing medically deprived areas, and making more data available for public health.

Association Between Physician and Patient Profiles

There was no link between the three physician and three patient groups (P=.37).

Discussion

Principal Findings

This is the first study to examine the patterns of use, benefits, as well as perceived obstacles and challenges for eHealth in PLHIV and their physicians using a detailed questionnaire on different aspects of eHealth, including retrieval of information on the internet and social media, mobile apps, and connected objects, along with the use of telemedicine and the collection of digitized personal health information in France and elsewhere.

Our findings revealed three distinct clusters of patients: (i) those for whom eHealth is part of a connected lifestyle; (ii) those who mistrust technology, although they are more averse to technology in general than to eHealth specifically; and (iii) those keen to adopt eHealth because they see it as a benefit for their health, and for whom eHealth does not represent any risk. Three clusters were also found for the physicians: (i) those strongly opposed to eHealth (resisters), (ii) those who believe in eHealth (enthusiasts), and (iii) those who are open to eHealth, and who rise to the challenge. This third group overlaps the second to some extent.

Strengths and Limitations

It should be emphasized that some PLHIV are logically exempt from hospital surveys, including those diagnosed but without follow up and those followed outside a hospital setting, whose number is increasing. In addition, recruitment was conducted on a voluntary basis and those unable to speak/read French or unable to complete a questionnaire were excluded. Among those eligible, PLHIV who did not have access to the internet or who were not comfortable with computers, and were thus likely technology skeptics, did not agree to participate in the study. Thus, if 838/1233 (67.96%) PLHIV agreed to participate, only 325/838 (38.8%) of them answered all of the questions, which could impact the representativeness of the sample. However, the design of this survey, conducted over a single week in 51 care units in France and its overseas departments, and the fact that the characteristics of those who answered all of the questions were not widely different from those reported for the general PLHIV population in France [18] entails the soundness of results, besides the rate of 86% of physicians included. In addition, these PLHIV were only 2 years older on average than the whole French PLHIV population, and this difference is not generational. Moreover, 30% is the usual rate of responses to online surveys [19].

The lifelong drug therapy requiring regular follow-up visits with laboratory tests is well documented from this survey. The high prevalence of multiple clinical comorbidities with concomitant polypharmacy necessitates effective interdisciplinary care, but also access to web-based information or connection of apps enabling PLHIV to monitor their health [20].

Implications

Overall, our study found that 55% of the PLHIV had already searched for information on the internet. This finding adds to a study published in 2013 that found 49% of people affected by a chronic illness or serious condition consulted the internet for health-related questions [21]. The fact that online searches could enable self-diagnosis or self-medication has already been discussed elsewhere [22]. In our survey, 74 of 154 (48.1%) PLHIV respondents changed the way they cared for their health after relevant internet searches. Participating in an association connected with HIV was less frequent (9%), probably because of the discrimination that PLHIV might be subject to, which is in contrast to patients with other chronic health disorders, 30% of whom discuss or exchange information about their illnesses [23].

The data presented herein suggest that an ideal app for PLHIV in France would be one that enables follow up of vaccinations and monitoring of physical health conditions linked to HIV. The purpose of such an app would be to "communicate better" with the patient's physician, extending the first findings of the Emerge project [24]. The questions then arise as to how this communication will be managed for physicians, how it will be integrated into the patient's care path, and who will pay for its implementation. The responses of physicians to these issues did

not differ across groups: 42% of the physicians saw no use for mobile apps, whereas 41% thought they might be useful for their patients to keep track of their appointments. An Odoxa survey conducted in 2015 found that 29% of patients regularly used common connected objects, and only 5% had been advised by their physician to use connected technology [25].

Interestingly, 22% of PLHIV and 11% of physicians who stated that some of their patients reported using connected objects emphasized that digital technologies in medical practice were still in their infancy. One reason for this lack of eager promotion may be related to controversial reports [26] on the positive consequences of such technologies for morbidity and mortality.

In January 2019, 52% of French respondents were in favor of distance consultations [27]. The percentage was higher among PLHIV (60%) in the present study. However, serious misgivings about data collection remain: more than one third of respondents, both PLHIV and physicians, considered that the law did not oversee such data collection to a sufficient degree. The security and privacy concerns about both the devices that collect data and the systems in which these data are stored are prominent, and thus require more specific regulations [28].

In our study, the PLHIV and physicians most often using eHealth were those who also generally used the internet frequently. Those who used eHealth least were those who were averse to technology in general. Some researchers in this field [29] have indicated a continuum from technophiles to technophobes representing a continuum of idealization to skepticism toward eHealth. Rio del Carral and colleagues [30] recently published the results of a survey conducted among a population attending a public fair on health in Lausanne. In 2018, most of the individuals surveyed (55%) expressed a reluctance to possess connected objects and physical health apps because they feared misuse of shared data. The study further showed that these tools reinforced the existing habits and practices of some people (eg, physical activity and diet), rather than being "useful" for those who did not yet have a "healthy" lifestyle, indicating a health technology divide among the population studied. It is well-established that health-related and literacy-related disparities should be minimized for eHealth to be widely accepted and adopted [20]. Certainly, we observed a divide of this sort in our study, but it is noteworthy that a third profile appeared within both PLHIV and physicians, including PLHIV attentive to eHealth for improving their condition and physicians who perceive a benefit for their patients or their practice. Only 31% of patients and 36% of physicians remained respectively mistrustful of or hesitant about using eHealth. With a fine-scale analysis of the sociodemographic and medical characteristics of respondents, we can better describe the characteristics and representations of people that fall somewhere between the extremes of "in favor" and "against," between enhancement by technology and enfeeblement and enslavement by technology [31]. This intermediate group comprised PLHIV "eHealth believers" who were older, with more comorbidities, and who viewed eHealth as a response to their multiple health disorders and the more complex care path they were offered. Their responses emphasize that they were not especially technophiles or attentive to their health, and while perceiving the benefits of eHealth, they did not easily fit along a technophile-technophobe continuum. Compared to the "internet adopters," fewer of the "eHealth believers" wanted to have consultations by video conference, especially for an intimate matter. It is noteworthy that such a group of older "eHealth believers" has not yet been identified in similar studies focused on other health conditions, in particular cardiac illnesses [32].

Méadel and Akrich [33] reported that studies documenting how physicians interact with "informed patients" show polarized reactions. Some studies found that patients who consulted the internet were poorly informed (overanxious and overdemanding), which prolonged consultations because the information brought by the patient had to be discussed and explained. Others, often more familiar with this tool, considered such consultation to be positive and likely to help patients become more active in their own health care. Here, we describe an additional group of physicians, "open to eHealth," who were most often infectious disease specialists, and were neither idealists nor skeptics, but saw the possibilities of eHealth for developing greater autonomy among their patients even though not all information can be entrusted to machines, similar to the argument put forth by Besnier [34]. In other words, some patients and physicians seem to pay greater attention to eHealth projects than other online initiatives; they consider eHealth as a contemporary "civilizational" tool designed to help users and improve their health condition or mode of practice [35]. In that respect, eHealth could play the part of a "mediator" [36].

It is interesting that the debate on eHealth extends beyond the scope of individual medical practice. Although PLHIV who were "adopters" and "believers" with regard to eHealth both recognized the benefit of eHealth in terms of coordination among different health care professionals, only the physicians "open to eHealth" favored overcoming the opposition between medicine and public health. The absence of any statistical link between the patient and physician groups suggests that little cogent discussion on the subject was taking place during consultations.

The coordinated adoption of eHealth by PLHIV and physicians may be the next step following the dissemination of our results within the French society in the fight against HIV/AIDS. Once physicians and patients share the same perceptions of eHealth, it is possible that the common use of tools accepted by both parties will lead to improved care.

Conclusion

Given the successful scale up of antiretroviral therapy globally, eHealth apps have the potential to transform HIV care beyond viral suppression in terms of comorbidity management and patient-reported outcome [37]. Our study shows that 26% of PLHIV and 43% of physicians are eHealth enthusiasts, while 31% and 37% are skeptics. For the latter group, there is a need to scale down literacy-related disparities; enforce regulations that will reduce concerns about security, privacy, and sharing personal data; as well as to ensure that physicians assisted by better hospital management are involved in the promotion of apps and connected objects in telemedicine. However, a third profile overcoming these challenges appeared in both PLHIV (43%), who were often older and attentive to eHealth for improving their health condition, and physicians (20%), who

find benefits of eHealth for patients or their own practice, although without a direct link between the two groups. It seems necessary for both PLHIV and physicians to address eHealth apps during consultations, which could lead to improvement in care and to a reduction of the cost of care pathways.

Acknowledgments

The study was supported by funding from CHU Clermont-Ferrand. The authors acknowledge all of the collaborators of this study at the participating hospitals across France for offering the online survey to PLHIV and physicians and participating in the study themselves. C Faudon and T Allegre (Aix-en-Provence); M Lombard and JL Schmit (Amiens); S Rehaiem and V Rabier (Angers), A Valran and E Piet (Annecy); J Gerbe and F Caby (Argenteuil), E Chevalier and A Foltzer (Besancon); L Traore and O Bouchaud (Bobigny); JC Duthe, S Ansart, and L de Saint Martin (Brest); P Goubin and R Verdon (Caen); J Prouteau and C Jacomet (Clermont-Ferrand); C Chesnel, L Richier, S Gallien, and B Elharrar (Créteil); I Calmont and A Cabié (Fort-de-France); M Marcou and P de Truchis (Garches); C Sautron and MP Moiton (La Réunion Saint-Denis); L Lainé, and S Leautez Nainville (La Roche-sur-Yon); P Camps and M Roncato-Saberan (La Rochelle); D Bornarel and A Greder-Belan (Le Chesnay); AS Keita and A Cheret (Le Kremlin-Bicêtre); F Dangeul-Potier and H Hitoto (Le Mans); B Montoya and JJ Laurichesse (Mantes-la-Jolie); S Tassi and E Frogel (Marne-la-Vallée); A Suzel-Ritleng and O Zaegel-Faucher (Marseille Sainte-Marguerite); C Fernandez and J Reynes (Montpellier); M Benomar and G Beck-Wirth (Mulhouse); A Soria and F Raffi (Nantes); I Touitou and P Pugliese (Nice); P Gougeon and S Sunder (Niort); MJ Dulucq and L Slama (Paris Hôtel-Dieu); C Louisin, C Duvivier, and PH Consigny (Paris Necker Institut Pasteur); M Pauchard, D Beniken, C Katlama, A Simon (Paris Pitié-Salpêtrière); C Tran and D Bollens (Paris Saint-Antoine); A Adda Lievin, G Thibaut, and P Pialoux (Paris Tenon); L Pinheiro and H Aumaitre (Perpignan), L Gerard and H Masson (Poissy); D Plainchamp and G Lemoal (Poitiers); JC Duthe and P Perfezou (Quimper), I Kmiec and F Bani-Sadr (Reims); C Guennoun and C Arvieux (Rennes); G Unal and A Lesourd (Rouen); V Ronat and A Fresard (Saint-Etienne); B Montoya and Y Welker (Saint-Germain-en-Laye); C Grand Courault and C Michau (Saint-Nazaire); P Fisher and D Rey (Strasbourg); V Lambry and A Lafeuillade (Toulon); M Digumber and O Robineau (Tourcoing); E Racamier and H Champagne (Valence); L Richier and S Dellion (Villeneuve-Saint-Georges).

Authors' Contributions

CJ, RO, and PD conceived the study. CJ, supported by RO and CL, designed the methodology. JP was responsible for project administration. CL performed the statistical analysis. CJ wrote the original draft. CJ, RO, JP, CL, FL, and PD revised and edited successive drafts of the paper. CJ, RO, FL, and PD were responsible for the validation. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of the three groups of people living with HIV obtained by mixed unsupervised classification.
[DOCX File , 22 KB - mhealth_v8i4e16140_app1.docx]

Multimedia Appendix 2
Comparison of the three physician groups obtained by mixed unsupervised classification.
[DOCX File , 20 KB - mhealth_v8i4e16140_app2.docx]

References

1. World Health Organization. Geneva; 2019. WHO Guideline: recommendations on digital interventions for health system strengthening URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-03-08]
2. Mitchell J. Increasing the cost-effectiveness of telemedicine by embracing e-health. J Telemed Telecare 2000;6 Suppl 1:S16-S19. [doi: 10.1258/1357633001934500] [Medline: 10793960]
3. Swan M. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. Big Data 2013 Jun;1(2):85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
4. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
5. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]

6. Schüll N. Data for life: Wearable technology and the design of self-care. BioSocieties 2016 Oct 13;11(3):317-333 [FREE Full text] [doi: 10.1057/biosoc.2015.47]
7. del Río Carral M, Roux P, Bruchez C, Santiago-Delefosse M. Santé digitale: promesses, défis et craintes. Une revue de la littérature. Prat Psychol 2017 Mar;23(1):61-77. [doi: 10.1016/j.prps.2016.06.004]
8. Salamati F, Pasek ZJ. Personal Wellness: Complex and Elusive Product and Distributed Self-services. Procedia CIRP 2014;16:283-288 [FREE Full text] [doi: 10.1016/j.procir.2014.02.016]
9. Brouard B. [Chapter 2. Internet of Things help to collect Big Data]. J Int Bioethique Ethique Sci 2017 Oct 27;28(3):27-30. [doi: 10.3917/jib.283.0027] [Medline: 29561096]
10. Appelboom G, LoPresti M, Reginster J, Sander Connolly E, Dumont EP. The quantified patient: a patient participatory culture. Curr Med Res Opin 2014 Dec 28;30(12):2585-2587. [doi: 10.1185/03007995.2014.954032] [Medline: 25118077]
11. Le Parlement Européen et le Conseil de l'Union Européenne. Réglement (UE) 2016/679 du parlement européen et du conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règement général sur la protection des données). Journal Officiel de l'Union Europeenne 2016;119:1-88 [FREE Full text]
12. DREES. Les dépenses de santé en 2016. 2017. Vues d'ensembles URL: https://drees.solidarites-sante.gouv.fr/IMG/pdf/ve-3.pdf [accessed 2020-03-08]
13. DREES. Ministere des Solidarites et de la Sante. 2015. Rapport « État de santé de la population en France » : Les Français en bonne santé, mais des disparités territoriales, sociales et de genre perdurent URL: https://tinyurl.com/sr7wgzu [accessed 2020-03-08]
14. Williams V, Rutter H, Christy T, Tarassenko L, Farmer A. Exploring patients' perspectives of an mHealth application: a qualitative study as part of EDGE COPD. Int J Integr Care 2013 Nov 20;13(7). [doi: 10.5334/ijic.1447]
15. Legifrance. Délibération n° 2016-263 du 21 juillet 2016 portant homologation d'une méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches dans le domaine de la santé ne nécessitant pas le recueil du consentement exprès ou écrit de la personne concernée (MR-003) URL: https://beta.legifrance.gouv.fr/jorf/id/JORFTEXT000033028290 [accessed 2020-03-08]
16. 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]
17. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
18. Morlat P. Épidémiologiede l'infection à VIH en France (juillet 2017). France: Conseil National du SIDA et des hépatites virales; 2017. Prise en charge médicale des personnes vivant avec le VIH URL: https://cns.sante.fr/wp-content/uploads/2017/10/experts-vih_epidemio.pdf [accessed 2020-03-08]
19. Survey Anyplace. 2019. What's the average survey response rate? [2019 benchmark] URL: https://surveyanyplace.com/average-survey-response-rate/ [accessed 2020-03-08]
20. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
21. La Francais et l'internet sante: Maladie chronique et numerique attitude.: Lauma; 2013. A la recherche du ePatient URL: http://www.patientsandweb.com/wp-content/uploads/2013/04/A-la-recherche-du-ePatient-externe.pdf [accessed 2020-03-08]
22. Delisle J, Servy H. Danger des sites patients? Lett Infectiolog 2018 Sep;5:23.
23. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. BMJ 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: 10.1136/bmj.38753.524201.7C] [Medline: 16513686]
24. Marent B, Henwood F, Darking M, EmERGE Consortium. Development of an mHealth platform for HIV Care: Gathering User Perspectives Through Co-Design Workshops and Interviews. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e184 [FREE Full text] [doi: 10.2196/mhealth.9856] [Medline: 30339132]
25. ODOXA. 2015 Jan 19. Baromètre-santé 360- La santé connectée URL: http://www.odoxa.fr/sondage/barometre-sante-360-sante-connectee/ [accessed 2020-03-08]
26. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012 Jun 21;344:e3874-e3892 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
27. Levy J, Potéreau J, Hauser M. Harris interactive study for LIVI. Paris: Harris Interactive; 2019 Jan. The French and telehealth URL: http://harris-interactive.fr/opinion_polls/les-francais-et-la-teleconsultation [accessed 2020-03-08]
28. Fraser AG, Butchart EG, Szymański P, Caiani EG, Crosby S, Kearney P, et al. The need for transparency of clinical evidence for medical devices in Europe. Lancet 2018 Aug;392(10146):521-530. [doi: 10.1016/s0140-6736(18)31270-4]
29. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Health Soc Rev 2016 Aug 05;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]

30. del Río Carral M, Schweizer A, Papon A, Santiago-Delefosse M. Les objets connectés et applications de santé : étude exploratoire des perceptions, usages (ou non) et contextes d'usage. Prat Psychol 2019 Mar;25(1):1-16. [doi: 10.1016/j.prps.2018.05.001]
31. Sharon T. Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. Philos Technol 2016 Apr 18;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]
32. Gallagher R, Roach K, Sadler L, Glinatsis H, Belshaw J, Kirkness A, et al. Mobile Technology Use Across Age Groups in Patients Eligible for Cardiac Rehabilitation: Survey Study. JMIR Mhealth Uhealth 2017 Oct 24;5(10):e161 [FREE Full text] [doi: 10.2196/mhealth.8352] [Medline: 29066425]
33. Méadel C, Akrich M. Internet, third party nebulous patient-physician relationship. Health Forums 2010;29(4):41-48. [doi: 10.3917/seve.029.0041]
34. Besnier JM. L'homme simplifié. Le syndrome de la touche étoile. France: Fayard; Oct 10, 2012.
35. Vygotskij L, Piaget J. Pensée et langage. France: La Dispute; 2013:1-546.
36. Chuah S, Rauschnabel P, Krey N, Nguyen B, Ramayah T, Lade S. Wearable technologies: The role of usefulness and visibility in smartwatch adoption. Comput Hum Behav 2016 Dec;65:276-284. [doi: 10.1016/j.chb.2016.07.047]
37. van den Berk G, Leoni M, Behrens G, Taljaard J, Arends JE. Improving HIV-related care through eHealth. Lancet HIV 2020 Jan;7(1):e8-e10. [doi: 10.1016/s2352-3018(19)30348-0]

Abbreviations

CHU: Centre Hospitalier Universitaire
eHealth: electronic health
EPICES: Evaluation of Precarity and Inequalities in Health Examination Centers
GDP: gross domestic product
MSM: men who have sex with men
PLHIV: people living with HIV
REDCap: Research Electronic Data Capture
VAS: visual analog scale

Edited by G Eysenbach; submitted 05.09.19; peer-reviewed by L Apers, H Mehdizadeh, E Da Silva, H Sandhu; comments to author 22.01.20; revised version received 30.01.20; accepted 26.02.20; published 15.04.20.

Please cite as:
Jacomet C, Ologeanu-Taddei R, Prouteau J, Lambert C, Linard F, Bastiani P, Dellamonica P
Adoption and Attitudes of eHealth Among People Living With HIV and Their Physicians: Online Multicenter Questionnaire Study
JMIR Mhealth Uhealth 2020;8(4):e16140
URL: http://mhealth.jmir.org/2020/4/e16140/
doi:10.2196/16140
PMID:32293581

Review

Exploring the Use of Telemonitoring for Patients at High Risk for Hypertensive Disorders of Pregnancy in the Antepartum and Postpartum Periods: Scoping Review

Maria Aquino[1,2], RN, MHI; Sarah Munce[3], PhD; Janessa Griffith[4,5], BA, MSc; Maureen Pakosh[6], BA, MISt; Mikayla Munnery[1], BComm; Emily Seto[1,2], PEng, PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[3]Rumsey Centre, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[4]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

[5]Women's College Hospital Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[6]Library & Information Services, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

Corresponding Author:
Maria Aquino, RN, MHI
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College Street, 4th Floor
Toronto, ON, M5T3M6
Canada
Phone: 1 4168039963
Email: maria.aquino@mail.utoronto.ca

Abstract

Background: High blood pressure complicates 2% to 8% of pregnancies, and its complications are present in the antepartum and postpartum periods. Blood pressure during and after pregnancy is routinely monitored during clinic visits. Some guidelines recommend using home blood pressure measurements for the management and treatment of hypertension, with increased frequency of monitoring for high-risk pregnancies. Blood pressure self-monitoring may have a role in identifying those in this high-risk group. Therefore, this high-risk pregnancy group may be well suited for telemonitoring interventions.

Objective: The aim of this study was to explore the use of telemonitoring in patients at high risk for hypertensive disorders of pregnancy (HDP) during the antepartum and postpartum periods. This paper aims to answer the following question: What is the current knowledge base related to the use of telemonitoring interventions for the management of patients at high risk for HDP?

Methods: A literature review following the methodological framework described by Arksey et al and Levac et al was conducted to analyze studies describing the telemonitoring of patients at high risk for HDP. A qualitative study, observational studies, and randomized controlled trials were included in this scoping review.

Results: Of the 3904 articles initially identified, 20 met the inclusion criteria. Most of the studies (13/20, 65%) were published between 2017 and 2018. In total, there were 16 unique interventions described in the 20 articles, all of which provide clinical decision support and 12 of which are also used to facilitate the self-management of HDP. Each intervention's design and process of implementation varied. Overall, telemonitoring interventions for the management of HDP were found to be feasible and convenient, and they were used to facilitate access to health services. Two unique studies reported significant findings for the telemonitoring group, namely, spontaneous deliveries were more likely, and one study, reported in two papers, described inductions as being less likely to occur compared with the control group. However, the small study sample sizes, nonrandomized groups, and short study durations limit the findings from the included articles.

Conclusions: Although current evidence suggests that telemonitoring could provide benefits for managing patients at high risk for HDP, more research is needed to prove its safety and effectiveness. This review proposes four recommendations for future research: (1) the implementation of large prospective studies to establish the safety and effectiveness of telemonitoring interventions; (2) additional research to determine the context-specific requirements and patient suitability to enhance accessibility to healthcare

services for remote regions and underserved populations; (3) the inclusion of privacy and security considerations for telemonitoring interventions to better comply with healthcare information regulations and guidelines; and (4) the implementation of studies to better understand the effective components of telemonitoring interventions.

(JMIR Mhealth Uhealth 2020;8(4):e15095) doi:10.2196/15095

KEYWORDS

high-risk pregnancy; blood pressure; preeclampsia; telemetry; telemedicine; mHealth; eHealth; smartphone; mobile phone

Introduction

Background

High blood pressure is one of the most common complications during pregnancy, affecting 2% to 8% of pregnancies [1,2]. The spectrum of hypertensive disorders of pregnancy (HDP) includes chronic hypertension, gestational hypertension, preeclampsia, eclampsia, and preeclampsia superimposed on chronic hypertension [1]. HDP pose short- and long-term risks for maternal and fetal health, including preterm delivery; hemolysis, elevated liver enzymes, and low platelet (HELLP) syndrome; disseminated intravascular coagulation; chronic renal failure; coronary artery disease; and premature death [3-5]. In addition, eclampsia and HELLP syndrome can occur for some patients postpartum [6-9]. Eclampsia, which is the new onset of grand mal seizures occurring in the absence of another identifiable cause [10], can develop at any point antepartum (38%-53%), intrapartum (18%-36%), and postpartum (11%-44%) [7]. Signs of preeclampsia usually diminish by 6 to 12 weeks postpartum [10].

During pregnancy, blood pressure measurements are routinely monitored, and high blood pressure is identified using blood pressure values obtained at the hospital or clinic visit [11]. International guidelines, such as the American College of Obstetricians and Gynecologists Hypertension in Pregnancy Guidelines for the treatment of hypertension, define it as clinic blood pressure >140/90 mm Hg [11]. Compared with clinic blood pressure measurements, ambulatory blood pressure measurements (ie, blood pressure measurements collected at various points, usually in a 24-hour period, while patients conduct routine activities) and home blood pressure measurements (ie, blood pressure values collected under a fixed schedule over a prolonged timeframe) [12] have a stronger association with end-organ damage and long-term health effects in the general population outside of pregnancy [13-15]. In fact, some international guidelines, such as the American Heart Association, American Society of Hypertension, Preventive Cardiovascular Nurses Association joint statement [16], and the European Society of Hypertension guidelines, emphasize the importance of blood pressure self-monitoring [17].

The benefits of home blood pressure monitoring have been demonstrated for clinical decision making. For example, home blood pressure between clinic visits can help to identify blood pressure changes in pregnant patients at home [17] and, combined with demographic risk factors such as chronic kidney disease, diabetes mellitus, and autoimmune disease such as systemic lupus erythematosus [18] may allow clinicians to estimate a woman's risk for HDP. Current UK guidelines recommend increased blood pressure monitoring for those at higher risk of preeclampsia [18]. Self-monitored blood pressure readings may have a role in identifying those in this high-risk group as well as those with white-coat hypertension (ie, high clinic blood pressure measurements but normal blood pressure otherwise) [19] and true chronic or gestational hypertension [20]. In addition, the 2013 guidelines from the American College of Obstetricians and Gynecologists recommend home monitoring for pregnant patients with chronic, poorly controlled, and gestational hypertension [21], and the National Institute for Health and Care Excellence recommends the self-monitoring of preeclampsia symptoms [22]. In general, the World Health Organization (WHO) recommends pregnant women to maintain their own case notes or home-based records related to their pregnancy to improve continuity and quality of care [23].

Telemonitoring may be well suited for patients at high risk for HDP for several reasons. First, more pregnant patients are tracking home blood pressure measurements through their own volition or by instruction from their health care provider (HCP) [24]. Second, pregnant patients who require additional monitoring have indicated a preference for the self-measurement of blood pressure at home over the more frequent visits to the prenatal clinic or ambulatory blood pressure monitoring [25,26]. Furthermore, telemonitoring has been shown to be effective for the management of chronic conditions, including cardiopulmonary disease, asthma, and heart failure, which have contributed to reduced patient travel, absenteeism, hospital length of stay, readmissions [27], and overall costs [28].

Objectives

There are three published reviews related to the telemonitoring of pregnant patients using various technologies. In 2017, Lanssens et al [29] published a scoping review on the telemonitoring of patients during the prenatal period. However, their review primarily focused on the telemonitoring of patients at high risk for preterm labor and gestational diabetes [29]. In addition, Lanssens et al [29] employed a narrow time range, from 1988 to 2010, for their inclusion criteria, potentially leading to missed relevant publications before 1988 and after 2010. Few studies have explored the existing telemonitoring technologies during the prenatal period, and in fact, the review by Lanssens et al [29] identified only 14 papers on the effectiveness of telemonitoring in obstetrics. Rivera-Romero et al [30] published a scoping review in 2018, which identified 11 articles exploring mobile health (mHealth) solutions for HDP. The authors found that only four studies collected physiological data and only two studies collected blood pressure measurements. Van den Heuvel et al [31] described 71 studies that reported on electronic health (eHealth) use during prenatal, perinatal, and postnatal care. In their review, the authors found 12 studies describing telemonitoring and teleconsulting

interventions and stated that telemonitoring of pregnancy may be a mechanism through which the potential of eHealth technologies is realized [31]. This scoping review aims to describe the available studies on telemonitoring interventions for the detection and management of HDP. This paper aims to answer the following research question: What is known about telemonitoring interventions for the management of patients at high risk for HDP? Specifically, this study sought to understand the types of telemonitoring interventions that have been used, the study designs employed, and the results of the studies.

Methods

Literature Review

A literature review consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidelines and following the methodological framework described by Arksey et al [32] and Levac et al [33] was conducted to analyze studies describing the telemonitoring of adult patients at high risk for HDP.

Inclusion Criteria

Peer-reviewed studies were included if they met the following criteria: (1) published in English, (2) included antepartum and postpartum adults aged 18 years and older at high risk for HDP (ie, history of high blood pressure, advanced maternal age, elevated blood lipids, high body mass index, and history of diabetes mellitus) [34], and (3) described a telemonitoring intervention that included feedback from the HCPs to the patients. For this scoping review, telemonitoring interventions were defined in the same way as described by Lanssens et al [29], involving periodic measurements of physiological metrics (eg, blood pressure, weight, and physical activity) and using an information and communication technology to relay these metrics from the patient's home to a health care facility. These telemonitoring interventions included measurements taken by patients themselves as well as through home visits by HCPs or community workers to ensure a comprehensive understanding of the available models of telemonitoring. All study designs were included in this review, such as randomized controlled trials (RCTs), prospective and retrospective studies, feasibility studies, economic evaluations, and case studies. No limitations with respect to year of publication were imposed. Abstracts, books and book chapters, literature reviews, and research in progress were excluded because the intent was to review completed research and peer-reviewed publications of telemonitoring interventions for patients at high risk for HDP.

Search Strategy

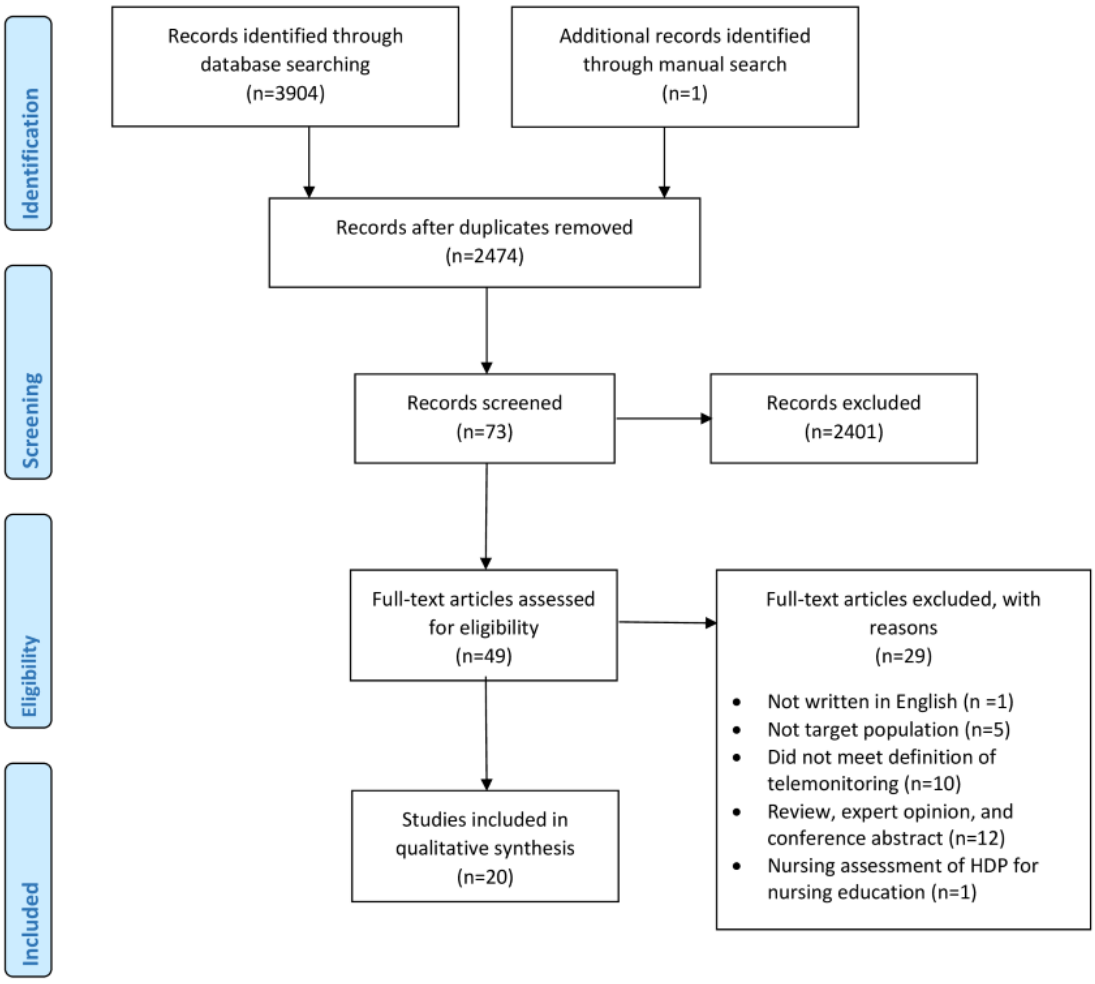
A comprehensive literature search was conducted in July 2018 in the following databases: Medical Literature Analysis and Retrieval System Online (MEDLINE), PubMed, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cumulative Index Nursing and Allied Health Literature, PsycINFO, Excerpta Medica Database, and EMCare.

Search strategies were developed with the aid of an experienced information specialist (MP) who reviewed and refined them through discussion with the research team. Keywords for this review included the target population, health condition, and telemonitoring system (ie, telemonitoring of adult patients at high risk for HDP). An article from a manual review of journals was identified and added to the search results for screening. The final search strategy for MEDLINE can be found in Multimedia Appendix 1.

Selection Procedure

Selection criteria forms were developed by the primary reviewer (MA) and pilot tested with the research team using five randomly selected articles. This was done to facilitate consistency among reviewers in the selection process. The initial database search resulted in 3904 articles. The selection procedure is illustrated in the flow diagram in Figure 1. After duplicate articles were removed, 2474 articles remained. The titles and abstracts of 2474 publications were independently evaluated and assessed for eligibility by two sets of reviewers (complete review by MA and joint review by SM and JG) using the systematic reviews web app Rayyan Qatar Computing Research Institute (Hamad Bin Khalifa University, Doha, Qatar) [35]. Of these, 49 publications were found to be eligible by 2 reviewers. The full text of the 49 publications were reviewed independently by MA and JG. A total of 29 publications were excluded because of the following reasons: (1) the article was not written in English; (2) the study did not include our target population; (3) the study did not meet our definition of telemonitoring; (4) the article was a review, conference abstract, or featured expert opinion; or (5) the study focused on the nursing assessment of HDP for nursing education. Disagreements among the reviewers were resolved by discussion until consensus was reached. A total of 20 articles were included for full data extraction and analysis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. HDP: hypertensive disorders for pregnancy.



Data Extraction

A data-charting table was developed collectively with the research team to determine relevant variables to include, and it was pilot tested to facilitate consistency during data extraction. A total of two sets of reviewers (complete review by MA and joint review by MM and JG) independently completed data extraction from each of the 20 articles. Any disagreements among the reviewers were resolved through discussion. Relevant data from the included studies were recorded on the data-charting table, including information on study characteristics, telemonitoring intervention description, outcome measures, and results.

Results

Study Characteristics

The included papers were published from 1987 to 2018, with more than half (n=13) published in 2017 (n=5) and 2018 (n=8). A total of eight [36-43] papers were published in the United Kingdom, followed by the United States (n=4) [44-47], Belgium (n=4) [48-51], Dominican Republic (n=1) [52], Netherlands (n=1) [53], and Guatemala (n=1) [54]. One study [55] described two different versions of a telemonitoring intervention being implemented in Nigeria, Mozambique, Pakistan, and India. A total of 15 papers reported on antepartum women, whereas five papers reported on postpartum women (refer to Multimedia Appendices 2 and 3 for further details).

A total of three study designs were identified in the included articles: a qualitative study (n=1), observational studies (n=16), and RCTs (n=3). A total of seven articles reported on the distinct aspects of the same three studies: for the blood pressure self-monitoring in pregnancy study, an observational study [37] and a qualitative study [38] were conducted; for the pregnancy remote monitoring study, observational studies at 1- [48] and 2-year [49] intervals as well as a cost-benefit analysis [50] were conducted; and for the home blood pressure monitoring study, two observational studies [36,39] were conducted. Thus, there were a total of 16 unique interventions from the 20 included studies. To avoid duplication, papers reporting on similar study results are grouped together.

The total number of study participants for the included studies was 2709, with one study [55] predicting the telemonitoring of over 30,000 women when its interventions are implemented. Study participants from two studies [39,48] were disregarded in the final tally because study participants in these studies were already accounted for in another study. There were no standardized criteria for the suitability of patients undergoing blood pressure telemonitoring.

Critical Appraisal of Included Studies

In keeping with the PRISMA-ScR guidelines, an assessment of the methodological approaches of the included articles was performed to identify study limitations. Included studies were assessed by study design (ie, qualitative studies, observational studies, and RCTs), and emerging themes were extracted from each grouping of studies.

The key themes that emerged from the qualitative and observational studies were as follows: (1) small study sample sizes; (2) nonrandomized study groups; (3) demographic and characteristic differences among groups; (4) short study durations; (5) limitations of study findings to the clinical and regulatory practices of a particular country; and (6) costs analyses that did not account for other costs to the patient, such as transportation, travel costs, and lost income for time spent in the hospital or clinic visit. For the three RCTs, investigators were unable to blind the study personnel and participants to the intervention, and there was a lack of allocation sequence concealment.

Intervention Characteristics

The study duration for interventions during pregnancy ranged from a gestational age of 4 weeks to delivery or admission to the hospital, whereas interventions in the postpartum period ranged from after delivery to 6 months. A total of three studies [43,51,55] did not specify the duration of their interventions. The data collection period of the studies ranged from 7 months to 35 months for interventions during pregnancy, whereas the data collection period for studies in the postpartum period ranged from 4 to 12 months.

All interventions collected maternal blood pressure, with some interventions collecting additional metrics such as heart rate; weight; activities; urinalysis for glucose and protein; symptoms of preeclampsia such as headache, epigastric pain, and visual symptoms; temperature; peripheral capillary oxygen saturation; and psychosocial signs and symptoms. In addition, four articles [51-54] collected data from the fetus, including fetal heart rate and kick count. A total of six unique studies [37,38,41-43,45,46] reported interventions in which only maternal blood pressure was collected. Blood pressure thresholds varied for each intervention, with alarm thresholds ranging from >140/90 mm Hg (n=5) to >160/110 mm Hg (n=4), or were unspecified (n=7). Criteria for interventions with blood pressure thresholds were either applied using guidelines, such as the International Society for the study of Hypertension in Pregnancy [36,48-50], American College of Obstetricians and Gynecologists [45,46], and local clinic guidelines [37,38,42], or were unspecified [41,43].

The schedule for collecting maternal blood pressure measurements varied across the included studies and ranged from four times a day (n=1) [47], two times a day (n=8) [37,38,44-46,48-50], daily (n=2) [40,53], weekly (n=1) [43], to did not specify a schedule (n=4) [42,51,54,55]. A total of three unique studies [36,39,41,52] proposed a variable schedule depending on patient condition.

Intervention Design and Implementation

All 16 interventions provided clinical decision support. A total of four of these interventions were developed specifically to enhance the assessment capabilities of HCPs, including Indigenous Mayan traditional birth attendants [54], community-based HCPs [55], community health workers [52], and physicians and midwives [53]. The remaining 12 interventions also facilitated the self-management of pregnant patients' signs and symptoms related to HDP.

Instructions in response to abnormal signs and symptoms were provided to patients for self-management interventions manually (n=6) [44,47-51], automatically (n=4) [40,43,45,46], or both (n=4) [36,38,39,56], or they were unspecified (n=2) [42,57]. Manual instructions for patients and HCPs included written or verbal instructions, and automatically provided instructions were given through a digital platform, such as a mobile or web app or a web-based dashboard.

Physiological data from interventions developed for enhancing HCP assessment capabilities (n=4) were inputted by HCPs when they visited patients at home, whereas interventions with the additional self-management component (n=12) were inputted by the patient. Physiological data of pregnant patients were entered either manually or automatically. Manual input included writing results of the patients in a journal [36-39,42], texting [45,46], or calling [47] to convey these results to HCPs. Automatic input involved a real-time transfer of blood pressure readings from an automated blood pressure machine to a digital platform. Automatic input included the use of a mobile phone or smartphone [36,38-40,43,52,54,55], whereas other technological components involved a web-based dashboard [48-50], web app [46,51], modem [41,53], or cloud-based portal [44].

The clinical team of HCPs, such as nurses, physicians, obstetricians, midwives, traditional birth attendants, community health workers, and community HCPs, who managed and supported pregnant patients at home varied depending on the intervention. A total of three studies [44,45,47] described interventions with 1 designated HCP who reviewed and monitored pregnant patients, whereas most studies (n=17) defined a more team-based approach in which a nurse or midwife triaged patients and consulted with an obstetrician on recommendations for the patient.

A total of 11 studies [36-39,41,42,44-47,51] described a training session for patients on how to use the intervention and provided information for the normal and abnormal values of physiological metrics, as well as actionable steps for critical results. A total of six studies [40,43,48-50,53] did not specify a training plan for patients or HCPs, and three studies [52,54,55] described training for HCPs. Training for HCPs to use interventions to enhance their assessment capabilities tended to be more intensive than training for patients to facilitate the self-management of HDP. For example, training for traditional birth attendants in Guatemala involved a 4-day training session by nurses on key maternal and neonatal assessment concepts, a minimum passing score of 90% for three standardized patient encounters, and retraining and reevaluation sessions for those who failed the initial evaluation in addition to training on

intervention use and knowledge of normal and abnormal values [54]. Similarly, training for community health workers in San Juan Province, Dominican Republic, entailed a 2-day workshop, an overview of maternal anatomy and obstetric assessment, 8 days of individual training, ongoing evaluations through discussion, observation, and a return demonstration [52].

The use of theoretical frameworks to guide the intervention was uncommon, with only one study describing the use of a theoretical framework in its implementation of an intervention. Bonnell et al [52] employed a community-based participatory research approach in examining how mHealth technology can be used in a community health worker model in San Juan Province, Dominican Republic.

Intervention Technology and Home Monitoring Platform

As all interventions collected maternal blood pressure readings, all interventions included the use of blood pressure monitors. A total of 11 interventions, described in 15 studies [36-41,43-51], provided blood pressure monitors to patients, whereas the other four interventions [52-55] required HCPs to take the patient's blood pressure readings during home visits. One intervention [42] did not specify whether the blood pressure monitor was provided to the patient for home use or whether HCPs brought along the monitors during home visits. Interestingly, not all interventions using blood pressure monitors were validated for pregnancy and preeclampsia. In fact, only three interventions, described in five studies [36-40], used the Microlife WatchBP monitor, which is validated for use in pregnancy and preeclampsia [58]. One intervention [54] used the Omron M7 monitor, which is validated for pregnancy but not for severe preeclampsia [59]. A total of two unique interventions [45,48-50] used blood pressure monitors that were not validated for use in pregnancy or preeclampsia (ie, Withings Blood Pressure Monitor) [60,61] or required consultation with a physician before use in pregnant patients (ie, Omron 3 series Blood Pressure Monitor) [62]. A total of eight interventions [41-43,46,51-53,55] did not specify the blood pressure monitor or model used, and validation information on 2 blood pressure monitors (eg, Ideal Life and Vasoplex) [44,47] could not be found. Additional technological components for each intervention depended on supplementary physiological metrics that were collected, such as a pulse oximeter, urine dipsticks, fetal heart monitor, activity tracker (ie, Withings Smart Body Analyzer), thermometer, and scale.

Study Results and Outcomes

Included studies reported on four main outcomes: (1) maternal and fetal outcomes, (2) health system utilization, (3) user (ie, patient or HCP) experience, and (4) intervention feasibility (see Multimedia Appendix 4 for further details).

A total of eight studies [36,39,40,44,48,49,53,54] reported on maternal and fetal health outcomes. Maternal outcomes included medication adherence, gestational outcomes, mode of delivery (eg, cesarean or vaginal), and adverse outcomes (including but not limited to acute renal failure, acute myocardial ischemia, intravenous medication for blood pressure control, hypertensive encephalopathy, and death). Fetal outcomes included neonatal and adverse outcomes such as preterm delivery, small for gestational age, fetal growth restriction, and death. A total of two unique studies [48,49,53] reported significant findings for the telemonitoring group, namely, spontaneous deliveries were more likely to take place. One study, described in two papers, reported that inductions were less likely to take place for patients in the telemonitoring group compared with the control group (P<.01) [48,49]. A total of six studies [36,39,40,44,53,54], which included two RCTs [40,53], reported nonsignificant findings regarding medication adherence and maternal or fetal health outcomes for patients in the telemonitoring group.

Health system utilization outcomes included admission and readmission or referral to the hospital, admission to the neonatal intensive care unit, and costs associated with admission to the hospital. A total of 11 papers reported on health system utilization. Moreover, two studies described more return visits to a health care facility: one study noted that mHealth users returned to a medical facility compared with none of the nonusers (P=.004) [44]. Another study reported that women using a mobile app to monitor home blood pressure returned to the hypertension clinic more times than the nonapp-based home blood pressure monitoring and control groups (P<.001) [39]. However, the authors did not explain whether an escalation of care was required and whether the cause of the return visit was related to hypertension. A total of four studies [36,39,48,49] reported fewer prenatal hospitalizations and clinic visits for patients in the telemonitoring group compared with the conventional monitoring group. Another study [53] reported no difference between the groups with regard to maternal and neonatal hospital admission rates. Furthermore, three studies [39,50,51] described cost savings as a result of fewer hospitalizations and clinic visits. One study [45] reported no hospital readmission in a postpartum population undergoing telemonitoring. Martinez et al [54] described an increase in referral rates to facility-level care when traditional birth attendants had access to the mHealth intervention.

Overall, five studies [38,43,45,47,52] reported on user (eg, patient or clinician) experience, which included the intervention's ease of use, and users' preference for and perceptions of blood pressure self-monitoring at home. A total of four of the five studies [38,43,45,47] discussed patients' preference for and perceptions of telemonitoring interventions compared with hospital or clinic visits. In addition, patients with a previous history of preeclampsia perceived that having the telemonitoring intervention empowered and reassured them [38]. Bonnell et al [52] reported that patients felt that home visits by community health workers using an mHealth app for the monitoring and assessment of HDP were of comparable clinical value to prenatal visits to local health care facilities.

Intervention feasibility included perceived benefits and barriers, data accuracy, recruitment, compliance, and effectiveness of detecting HDP. In one study, patients who elected to be in the telemonitoring group perceived telemonitoring to be beneficial [44]. Other studies found that home blood pressure readings were accurately collected or transferred to a digital platform [40,42,43] and that HCPs were able to reliably use home blood pressure measurements for clinical decision making in a similar way as clinic blood pressure measurements [37,47,54]. In

addition, three interventions described a high rate of compliance related to blood pressure reading collection and submission. One study found that 84% of the participants texted at least one blood pressure measurement [45], whereas another study found that 92.2% of the participants from the telemonitoring group sent one blood pressure measurement via text message compared with the 43.7% of the participants from the control group who sent one blood pressure measurement via text message [46]. The median compliance was 85% for submitting daily blood pressure readings for another study [40]. However, the lack of consistent internet connection in rural areas proved to be a technological barrier to telemonitoring [52].

Discussion

Principal Findings

This review presented a summary of existing telemonitoring interventions for patients at high risk for HDP both antepartum and postpartum. Of the 20 included studies, 13 were published between 2017 and 2018, suggesting that telemonitoring interventions for patients at high risk for HDP are a novel and burgeoning area of research. This study, which only identified 16 unique interventions, reflects the limited use of telemonitoring for antepartum [29] and postpartum [44] patients. Interventions for postpartum patients were included because current recommendations call for the postpartum monitoring of hypertension, as symptoms can develop regardless of the history of hypertension or preeclampsia in the antepartum period [6]. However, given the exploratory nature of a portion of the included studies, more research is needed before any recommendations can be made for the telemonitoring of this high-risk group.

Surprisingly, not all interventions used blood pressure measurement monitors that are validated for use in pregnancy and preeclampsia. Only five studies reported using a blood pressure instrument validated for pregnancy and preeclampsia, such as the Microlife WatchBP monitor. Accurate blood pressure measurements in pregnancy are essential for the appropriate management and treatment of patients [63]. HCPs should be cognizant of the impact of inaccurate blood pressure measurements and should consider recommending patients to use blood pressure measurement monitors that have been appropriately validated in this population [63]. Finally, implementing the practice of blood pressure self-monitoring in the antepartum and postpartum period would require a standardization of home blood pressure thresholds [20].

The feasibility [40,42,43] and accuracy [37,47,54] of blood pressure readings from the included interventions support the WHO recommendation of pregnant women maintaining home-based records throughout their pregnancy [23]. There was also a high compliance rate noted by three interventions, but only one of these interventions discussed security as an intervention consideration [45]. In patients with gestational diabetes mellitus, Homko et al [64] showed a significant correlation between income and blood sugar result transmissions, with women with higher income sending results more frequently. Therefore, consideration of socioeconomic status may be beneficial when developing and implementing self-management interventions. Given the ubiquitous nature of mobile phone use [65], there is potential for integrating mobile phones and smartphones into the management and treatment of this patient population. In addition, automatic data entry of blood pressure measurements may reduce the possibility of incorrectly entering blood pressure measurements and the burden associated with manual entry.

This review showed only two unique studies reporting a significant difference for maternal and fetal health outcomes, with the telemonitoring group experiencing a higher likelihood for spontaneous deliveries [48,49,53], and one study reported a lower likelihood for inductions [48,49]. Similarly, previous studies reporting on the health outcomes of women with gestational diabetes using mHealth technologies showed limited impact on maternal health outcomes [64,66,67]. Homko et al [64] concluded that the benefits of health care technologies may lie in their ability to streamline health care processes (eg, reducing the need for clinic follow-ups and unnecessary hospital admissions). Costs analyses included in this review highlight the ability for telemonitoring to reduce costs by monitoring stable high-risk pregnant women at home [39,50,51] instead of being admitted to a hospital, which is the standard management for patients diagnosed with preeclampsia [68,69]. As outpatient management of HDP occurs in the earlier stages of pregnancy [24] and after the baby has been delivered [68], telemonitoring may provide enhanced surveillance of disease progression. However, telemonitoring interventions should not encourage obstetricians to defer hospital admission [24] or replace HCP visits or contact points. The WHO recommends a minimum of eight contact points with an HCP to decrease antenatal mortality and improve women's experience during pregnancy [23]. Contraindications to the self-monitoring of home blood pressure, such as atrial fibrillation or other abnormal heart rhythms [16], should also be considered.

Identified barriers in the implementation of telemonitoring interventions for this high-risk group include mHealth infrastructure [52,55] and costs (eg, equipment or lack of HCP compensation for the provision of telemedicine) [44]. Similarly, a systematic review on the use of mHealth technologies for the self-management of diabetes noted the following barriers for its adoption and use: a lack of financial resources, constrained human and technical resources, integration challenges with existing health information systems, and limited incentives or reimbursement [70].

Despite the limitations of telemonitoring, the included studies generally showed telemonitoring to be positive for patients and HCPs. Patients stated their preference for telemonitoring over hospital or clinic visits for monitoring or follow-up because of its convenience [38,43,45,47]. Enhanced clinical decision making and assessment capabilities for HCPs in low-to-middle income countries were described in three telemonitoring interventions [52,54,55]. For these interventions, HCPs such as traditional birth attendants in Guatemala [54]; community-based HCPs in Nigeria, Mozambique, Pakistan, and India [55]; and community health workers in Dominican Republic [52], who have less formal education and training than traditional medical professions, successfully assessed and referred pregnant patients in remote communities to facility-level care as required. These

findings are supported by the WHO guidelines for antenatal care, which provide a context-specific recommendation for antenatal home visits to enhance the use of antenatal care services and perinatal health outcomes, especially for those in remote locations with limited access to health services [23]. Evaluations of antenatal health services in Nigeria, Mozambique, India, and Dominican Republic described the need to eliminate possible barriers to access, such as the following: (1) transportation and cost for patients to seek antenatal care [71], (2) reducing delays in referrals from primary health care to more comprehensive antenatal health services [72] by improving the capacity of primary health care facilities to do so [73], and (3) planning for accessible and equitable antenatal health programs [74]. An exploratory study in Pakistan and the WHO guidelines propose the shifting of some antenatal care tasks to other HCPs, such as trained lay health workers (eg, Lady Health Workers in Pakistan) [75], auxiliary nurses, and midwives [23]. Designing more culturally appropriate antenatal programs and government policies may reduce the health risks associated among indigenous pregnant women in countries such as Guatemala, where a significant proportion of the indigenous population resides [76].

Research Gaps and Suggestions for Future Research

Future studies on the effectiveness of telemonitoring for patients at high risk for HDP need to include more rigorous study designs with larger sample sizes. In addition, longer intervention duration periods may allow for more robust study results and reflect implementation processes more consistent with real-world practice. Telemonitoring interventions for this high-risk group have the potential to enhance antenatal care for women in high-, middle-, and low-income countries. However, addressing implementation and adoption barriers, such as mHealth infrastructure and costs, is key. In addition, which patient group or groups within the high-risk HDP population would benefit most from telemonitoring needs to be determined. More qualitative studies may help to identify this group and guide the researchers in their development and implementation of a telemonitoring intervention.

This review revealed the need for research to establish more robust evidence for the safety and effectiveness of these interventions for this high-risk group. This review proposes 4 recommendations for future research: (1) the implementation of large prospective studies to establish the safety and effectiveness of telemonitoring interventions; (2) additional research to determine the context-specific requirements and patient suitability to enhance patient accessibility to health care services for remote regions and underserved populations; (3) the inclusion of key considerations such as privacy and security in the development and implementation of telemonitoring interventions to better comply with health care information regulations and guidelines; and (4) the implementation of evaluation studies to better understand the effective components of telemonitoring interventions.

Study Limitations and Strengths

This scoping review has some limitations. It evaluated peer-reviewed journal articles written in English from eight relevant databases, therefore, telemonitoring interventions described in other languages may have been missed. Furthermore, additional search results may have been found in other databases and sources (eg, grey literature, conference proceedings, and books) not included in this review.

Strengths of this review include the use of an experienced information specialist for the development of the search strategy, duplication for each phase of the review (article screening and selection, data extraction, and full-text analysis), and the absence of limitations with respect to year of publication.

Conclusions

The short- and long-term impacts of HDP on maternal and fetal health are significant and can include multiorgan disease and mortality. Although there are increasingly more studies being published on the telemonitoring of patients at high risk for HDP, this review found only 16 unique interventions on the subject. The current knowledge base of telemonitoring interventions shows some promise for the use of telemonitoring in detecting and managing HDP in patients during and after pregnancy. Specifically, studies indicate that telemonitoring interventions can be feasible, convenient, and cost-effective. However, there is currently very limited evidence on the benefits of telemonitoring on health outcomes. This lack of evidence, combined with mHealth infrastructure and financial barriers, may impede the adoption of these potentially beneficial technologies for patients.

Acknowledgments

This study was supported by the Trainee Award in Women's Heart Health and was funded by the Canadian Institutes of Health Research' s (CIHR) Institute for Gender and Health, CIHR's Institute for Circulatory and Respiratory Health, and the Heart and Stroke Foundation of Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medical Literature Analysis and Retrieval System Online search strategy.
[DOCX File , 24 KB - mhealth_v8i4e15095_app1.docx]

Multimedia Appendix 2
Overview of included studies on the telemonitoring of women at high risk for hypertensive disorders of pregnancy.
[DOCX File , 36 KB - mhealth_v8i4e15095_app2.docx]

Multimedia Appendix 3
Characteristics and design of the telemonitoring interventions for patients at high risk for hypertensive disorders of pregnancy.
[DOCX File , 38 KB - mhealth_v8i4e15095_app3.docx]

Multimedia Appendix 4
Outcomes of the telemonitoring interventions for patients at high risk for hypertensive disorders of pregnancy.
[DOCX File , 39 KB - mhealth_v8i4e15095_app4.docx]

References

1. Mammaro A, Carrara S, Cavaliere A, Ermito S, Dinatale A, Pappalardo EM, et al. Hypertensive disorders of pregnancy. J Prenat Med 2009 Jan;3(1):1-5 [FREE Full text] [Medline: 22439030]
2. Nathan HL, Duhig K, Hezelgrave NL, Chappell LC, Shennan AH. Blood pressure measurement in pregnancy. Obstet Gynecol 2015;17(2):91-98. [doi: 10.1111/tog.12173]
3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol 2009 Jun;33(3):130-137. [doi: 10.1053/j.semperi.2009.02.010] [Medline: 19464502]
4. Butalia S, Audibert F, Côté AM, Firoz T, Logan AG, Magee LA, Hypertension Canada. Hypertension Canada's 2018 Guidelines for the Management of Hypertension in Pregnancy. Can J Cardiol 2018 May;34(5):526-531. [doi: 10.1016/j.cjca.2018.02.021] [Medline: 29731014]
5. Ghulmiyyah L, Sibai B. Maternal mortality from preeclampsia/eclampsia. Semin Perinatol 2012 Feb;36(1):56-59. [doi: 10.1053/j.semperi.2011.09.011] [Medline: 22280867]
6. Sibai BM. Etiology and management of postpartum hypertension-preeclampsia. Am J Obstet Gynecol 2012 Jun;206(6):470-475. [doi: 10.1016/j.ajog.2011.09.002] [Medline: 21963308]
7. Sibai BM. Diagnosis, prevention, and management of eclampsia. Obstet Gynecol 2005 Feb;105(2):402-410. [doi: 10.1097/01.AOG.0000152351.13671.99] [Medline: 15684172]
8. Chames MC, Livingston JC, Ivester TS, Barton JR, Sibai BM. Late postpartum eclampsia: a preventable disease? Am J Obstet Gynecol 2002 Jun;186(6):1174-1177. [doi: 10.1067/mob.2002.123824] [Medline: 12066093]
9. Sibai BM, Stella CL. Diagnosis and management of atypical preeclampsia-eclampsia. Am J Obstet Gynecol 2009 May;200(5):481.e1-481.e7. [doi: 10.1016/j.ajog.2008.07.048] [Medline: 19019323]
10. Turner J. Diagnosis and management of pre-eclampsia: an update. Int J Womens Health 2010 Sep 30;2:327-337 [FREE Full text] [doi: 10.2147/IJWH.S8550] [Medline: 21151680]
11. Mikami Y, Takai Y, Era S, Ono Y, Saitoh M, Baba K, et al. Provisional criteria for the diagnosis of hypertension in pregnancy using home blood pressure measurements. Hypertens Res 2017 Jul;40(7):679-684 [FREE Full text] [doi: 10.1038/hr.2017.6] [Medline: 28179623]
12. Imai Y, Obara T, Asamaya K, Ohkubo T. The reason why home blood pressure measurements are preferred over clinic or ambulatory blood pressure in Japan. Hypertens Res 2013 Aug;36(8):661-672. [doi: 10.1038/hr.2013.38] [Medline: 23595050]
13. Karpettas N, Destounis A, Kollias A, Nasothimiou E, Moyssakis I, Stergiou GS. Prediction of treatment-induced changes in target-organ damage using changes in clinic, home and ambulatory blood pressure. Hypertens Res 2014 Jun;37(6):543-547. [doi: 10.1038/hr.2014.24] [Medline: 24572912]
14. Ohkubo T, Imai Y, Tsuji I, Nagai K, Kato J, Kikuchi N, et al. Home blood pressure measurement has a stronger predictive power for mortality than does screening blood pressure measurement: a population-based observation in Ohasama, Japan. J Hypertens 1998 Jul;16(7):971-975. [doi: 10.1097/00004872-199816070-00010] [Medline: 9794737]
15. Her AY, Kim YH, Rim SJ, Kim JY, Choi EY, Min PK, et al. Home blood pressure is the predictor of subclinical target organ damage like ambulatory blood pressure monitoring in untreated hypertensive patients. Anadolu Kardiyol Derg 2014 Dec;14(8):711-718 [FREE Full text] [doi: 10.5152/akd.2014.5119] [Medline: 25188760]
16. Pickering TG, Miller NH, Ogedegbe G, Krakoff LR, Artinian NT, Goff D, American Heart Association, American Society of Hypertension, Preventive Cardiovascular Nurses Association. Call to action on use and reimbursement for home blood pressure monitoring: executive summary: a joint scientific statement from the American Heart Association, American Society Of Hypertension, and Preventive Cardiovascular Nurses Association. Hypertension 2008 Jul;52(1):1-9. [doi: 10.1161/HYPERTENSIONAHA.107.189011] [Medline: 18497371]
17. Hodgkinson JA, Tucker KL, Crawford C, Greenfield SM, Heneghan C, Hinton L, et al. Is self monitoring of blood pressure in pregnancy safe and effective? Br Med J 2014 Nov 18;349:g6616. [doi: 10.1136/bmj.g6616] [Medline: 25406132]
18. National Institute for Health and Care Excellence.: National Institute for Health and Care Excellence; 2008. NICE Clinical Guideline 62: Antenatal Care for Uncomplicated Pregnancies URL: https://www.nice.org.uk/guidance/cg62/chapter/1-Guidance#clinical-examination-of-pregnant-women [accessed 2019-06-12]

19. Pickering TG, James GD, Boddie C, Harshfield GA, Blank S, Laragh JH. How common is white coat hypertension? J Am Med Assoc 1988 Jan 8;259(2):225-228. [doi: 10.1001/jama.1988.03720020027031] [Medline: 3336140]
20. Tucker KL, Bankhead C, Hodgkinson J, Roberts N, Stevens R, Heneghan C, et al. How do home and clinic blood pressure readings compare in pregnancy? Hypertension 2018 Sep;72(3):686-694 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.118.10917] [Medline: 30354754]
21. American College of ObstetriciansGynecologists, Task Force on Hypertension in Pregnancy. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. Obstet Gynecol 2013 Nov;122(5):1122-1131. [doi: 10.1097/01.AOG.0000437382.03963.88] [Medline: 24150027]
22. National Institute for Health and Care Excellence. 2019 Jun. Hypertension in Pregnancy: Diagnosis and Management: NICE guideline [NG133] URL: https://www.nice.org.uk/guidance/ng133 [accessed 2019-06-30]
23. WHO. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. Geneva, Switzerland: World Health Organization; 2016.
24. Lavallee L, Tucker K, McManus R. Pregnant women are doing it for themselves: Self-monitoring of blood pressure in pregnancy. Br J Midwifery 2018;26(7):451-454. [doi: 10.12968/bjom.2018.26.7.451]
25. Cartwright W, Dalton KJ, Swindells H, Rushant S, Mooney P. Objective measurement of anxiety in hypertensive pregnant women managed in hospital and in the community. Br J Obstet Gynaecol 1992 Mar;99(3):182-185. [doi: 10.1111/j.1471-0528.1992.tb14495.x] [Medline: 1606113]
26. Denolle T, Weber J, Calvez C, Getin Y, Daniel J, Lurton O, et al. Diagnosis of white coat hypertension in pregnant women with teletransmitted home blood pressure. Hypertens Pregnancy 2008;27(3):305-313. [doi: 10.1080/10641950802000950] [Medline: 18696359]
27. Averwater NW, Burchfield DC. No place like home: telemonitoring can improve home care. Healthc Financ Manage 2005 Apr;59(4):46-8, 50, 52. [Medline: 15853034]
28. Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]
29. Lanssens D, Vandenberk T, Thijs IM, Grieten L, Gyselaers W. Effectiveness of telemonitoring in obstetrics: scoping review. J Med Internet Res 2017 Sep 27;19(9):e327 [FREE Full text] [doi: 10.2196/jmir.7266] [Medline: 28954715]
30. Rivera-Romero O, Olmo A, Muñoz R, Stiefel P, Miranda ML, Beltrán LM. Mobile health solutions for hypertensive disorders in pregnancy: scoping literature review. JMIR Mhealth Uhealth 2018 May 30;6(5):e130 [FREE Full text] [doi: 10.2196/mhealth.9671] [Medline: 29848473]
31. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 5;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
32. 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]
33. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
34. Quan LM, Xu QL, Zhang GQ, Wu LL, Xu H. An analysis of the risk factors of preeclampsia and prediction based on combined biochemical indexes. Kaohsiung J Med Sci 2018 Feb;34(2):109-112 [FREE Full text] [doi: 10.1016/j.kjms.2017.10.001] [Medline: 29413226]
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
36. Perry H, Sheehan E, Thilaganathan B, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population. Ultrasound Obstet Gynecol 2018 Apr;51(4):524-530 [FREE Full text] [doi: 10.1002/uog.19023] [Medline: 29468771]
37. Tucker KL, Taylor KS, Crawford C, Hodgkinson JA, Bankhead C, Carver T, et al. Blood pressure self-monitoring in pregnancy: examining feasibility in a prospective cohort study. BMC Pregnancy Childbirth 2017 Dec 28;17(1):442 [FREE Full text] [doi: 10.1186/s12884-017-1605-0] [Medline: 29284456]
38. Hinton L, Tucker KL, Greenfield SM, Hodgkinson JA, Mackillop L, McCourt C, et al. Blood pressure self-monitoring in pregnancy (BuMP) feasibility study; a qualitative analysis of women's experiences of self-monitoring. BMC Pregnancy Childbirth 2017 Dec 19;17(1):427 [FREE Full text] [doi: 10.1186/s12884-017-1592-1] [Medline: 29258469]
39. Xydopoulos G, Perry H, Sheehan E, Thilaganathan B, Fordham R, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population: cost-minimization study. Ultrasound Obstet Gynecol 2019 Apr;53(4):496-502. [doi: 10.1002/uog.19041] [Medline: 29516615]
40. Cairns AE, Tucker KL, Leeson P, Mackillop LH, Santos M, Velardo C, SNAP-HT Investigators. Self-Management of Postnatal Hypertension: The SNAP-HT Trial. Hypertension 2018 Aug;72(2):425-432. [doi: 10.1161/HYPERTENSIONAHA.118.10911] [Medline: 29967037]
41. Dalton KJ, Manning K, Robarts PJ, Dripps JH, Currie JR. Computerized home telemetry of maternal blood pressure in hypertensive pregnancy. Int J Biomed Comput 1987 Nov;21(3-4):175-187. [doi: 10.1016/0020-7101(87)90085-7] [Medline: 3679578]

42. Waugh J, Bosio P, Habiba M, Boyce T, Shennan A, Halligan A. Home monitoring of blood pressure in pregnancy at high risk of pre-eclampsia. Eur J Obstet Gynecol Reprod Biol 2001 Nov;99(1):109-111. [doi: 10.1016/s0301-2115(01)00353-0] [Medline: 11604197]
43. Ganapathy R, Grewal A, Castleman JS. Remote monitoring of blood pressure to reduce the risk of preeclampsia related complications with an innovative use of mobile technology. Pregnancy Hypertens 2016 Oct;6(4):263-265. [doi: 10.1016/j.preghy.2016.04.005] [Medline: 27939464]
44. Rhoads SJ, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring implementation of m-Health monitoring in postpartum women with hypertension. Telemed J E Health 2017 Oct;23(10):833-841 [FREE Full text] [doi: 10.1089/tmj.2016.0272] [Medline: 28475431]
45. Hirshberg A, Bittle M, VanDerTuyn M, Mahraj K, Asch D, Rosin R, et al. Rapid-cycle innovation testing of text-based monitoring for management of postpartum hypertension. J Clin Outcomes Manag 2017 Feb;24(2) [FREE Full text]
46. Hirshberg A, Downes K, Srinivas S. Comparing standard office-based follow-up with text-based remote monitoring in the management of postpartum hypertension: a randomised clinical trial. BMJ Qual Saf 2018 Nov;27(11):871-877. [doi: 10.1136/bmjqs-2018-007837] [Medline: 29703800]
47. Naef RW, Perry KG, Magann EF, McLaughlin BN, Chauhan SP, Morrison JC. Home blood pressure monitoring for pregnant patients with hypertension. J Perinatol 1998;18(3):226-229. [Medline: 9659655]
48. Lanssens D, Vandenberk T, Smeets CJ, de Cannière H, Molenberghs G, van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 9;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
49. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
50. Lanssens D, Vandenberk T, Smeets CJ, de Cannière H, Vonck S, Claessens J, et al. Prenatal remote monitoring of women with gestational hypertensive diseases: cost analysis. J Med Internet Res 2018 Mar 26;20(3):e102 [FREE Full text] [doi: 10.2196/jmir.9552] [Medline: 29581094]
51. Buysse H, de Moor G, van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. Int J Med Inform 2008 Jul;77(7):470-476. [doi: 10.1016/j.ijmedinf.2007.08.009] [Medline: 17923433]
52. Bonnell S, Griggs A, Avila G, Mack J, Bush RA, Vignato J, et al. Community health workers and use of mHealth: improving identification of pregnancy complications and access to care in the Dominican republic. Health Promot Pract 2018 May;19(3):331-340. [doi: 10.1177/1524839917708795] [Medline: 28578606]
53. Monincx WM, Zondervan HA, Birnie E, Ris M, Bossuyt PM. High risk pregnancy monitored antenatally at home. Eur J Obstet Gynecol Reprod Biol 1997 Dec;75(2):147-153. [doi: 10.1016/s0301-2115(97)00122-x] [Medline: 9447367]
54. Martinez B, Ixen EC, Hall-Clifford R, Juarez M, Miller AC, Francis A, et al. mHealth intervention to improve the continuum of maternal and perinatal care in rural Guatemala: a pragmatic, randomized controlled feasibility trial. Reprod Health 2018 Jul 4;15(1):120 [FREE Full text] [doi: 10.1186/s12978-018-0554-z] [Medline: 29973229]
55. Dunsmuir DT, Payne BA, Cloete G, Petersen CL, Görges M, Lim J, et al. Development of mHealth applications for pre-eclampsia triage. IEEE J Biomed Health Inform 2014 Nov;18(6):1857-1864. [doi: 10.1109/JBHI.2014.2301156] [Medline: 25375683]
56. Tucker KL, Bowen L, Crawford C, Mallon P, Hinton L, Lee M, et al. The feasibility and acceptability of self-testing for proteinuria during pregnancy: A mixed methods approach. Pregnancy Hypertens 2018 Apr;12:161-168 [FREE Full text] [doi: 10.1016/j.preghy.2017.11.009] [Medline: 29242046]
57. Swindells HE, Mooney P, Cartwright W, Dalton KJ, Perinatal Research Group. Blood pressure telemetry from home. Midwife Health Visit Community Nurse 1990 Mar;26(3):88-90 no doi, pmid, or electronic copy available.
58. Clark K, Snowball O, Nzelu D, Kay P, Kametas NA. Validation of the Microlife WatchBP Home blood pressure device in pregnancy for medium and large arm circumferences. Blood Press Monit 2018 Jun;23(3):171-174. [doi: 10.1097/MBP.0000000000000315] [Medline: 29596069]
59. Nouwen E, Snijder M, van Montfrans G, Wolf H. Validation of the Omron M7 and Microlife 3BTO-A blood pressure measuring devices in preeclampsia. Hypertens Pregnancy 2012;31(1):131-139. [doi: 10.3109/10641955.2010.544799] [Medline: 21332327]
60. Food and Drug Administration. 2014. 510(K) Summary K133125 Withings URL: https://www.accessdata.fda.gov/cdrh_docs/pdf13/k133125.pdf [accessed 2019-06-12]
61. Hay A, Ayis S, Nzelu D, James L, Kametas NA. Validation of the Withings BP-800 in pregnancy and impact of maternal characteristics on the accuracy of blood pressure measurement. Pregnancy Hypertens 2016 Oct;6(4):406-412. [doi: 10.1016/j.preghy.2016.09.005] [Medline: 27939491]
62. Omron Healthcare Wellness & Healthcare Products.: Omron Healthcare Inc; 2013. 3 Series Blood Pressure Monitor Model BP710N Instruction Manual URL: https://omronhealthcare.com/wp-content/uploads/BP710N-EN.pdf [accessed 2019-06-12]

63. Bello NA, Woolley JJ, Cleary KL, Falzon L, Alpert BS, Oparil S, et al. Accuracy of blood pressure measurement devices in pregnancy: a systematic review of validation studies. Hypertension 2018 Feb;71(2):326-335 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.117.10295] [Medline: 29229741]
64. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
65. Pew Research Center. Washington, D.C: Pew Research Center; 2019 Jun 12. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-03-03]
66. Kruger DF, White K, Galpern A, Mann K, Massirio A, McLellan M, et al. Effect of modem transmission of blood glucose data on telephone consultation time, clinic work flow, and patient satisfaction for patients with gestational diabetes mellitus. J Am Acad Nurse Pract 2003 Aug;15(8):371-375. [doi: 10.1111/j.1745-7599.2003.tb00410.x] [Medline: 14509102]
67. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, de la Cruz MJ, Martín P, et al. A Telemedicine system based on internet and short message service as a new approach in the follow-up of patients with gestational diabetes. Diabetes Res Clin Pract 2010 Feb;87(2):e15-e17. [doi: 10.1016/j.diabres.2009.12.002] [Medline: 20044162]
68. National Collaborating Centre for Women's and Children's Health (UK). NCBI - NIH. London, UK: RCOG Press; 2010 Aug. Hypertension in Pregnancy: The Management of Hypertensive Disorders During Pregnancy URL: https://www.ncbi.nlm.nih.gov/pubmed/22220321 [accessed 2020-02-14]
69. ObG Project.: The ObG Project Inc ACOG Preeclampsia Guidelines: Antenatal Management and Timing of Delivery URL: https://www.obgproject.com/2018/12/27/acog-preeclampsia-guidelines-antenatal-management-and-timing-of-delivery/ [accessed 2019-04-05]
70. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Dec;15(12):109 [FREE Full text] [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
71. Saidu R, August EM, Alio AP, Salihu HM, Saka MJ, Jimoh AA. An assessment of essential maternal health services in Kwara State, Nigeria. Afr J Reprod Health 2013 Mar;17(1):41-48. [Medline: 24069733]
72. Augusto O, Keyes EE, Madede T, Abacassamo F, de la Corte P, Chilundo B, et al. Progress in Mozambique: Changes in the availability, use, and quality of emergency obstetric and newborn care between 2007 and 2012. PLoS One 2018;13(7):e0199883 [FREE Full text] [doi: 10.1371/journal.pone.0199883] [Medline: 30020958]
73. Miller S, Cordero M, Coleman A, Figueroa J, Brito-Anderson S, Dabagh R, et al. Quality of care in institutionalized deliveries: the paradox of the Dominican Republic. Int J Gynaecol Obstet 2003;82(1):89-103. [doi: 10.1016/s0020-7292(03)00148-6]
74. Mony PK, Krishnamurthy J, Thomas A, Sankar K, Ramesh BM, Moses S, et al. Availability and distribution of emergency obstetric care services in Karnataka State, South India: access and equity considerations. PLoS One 2013;8(5):e64126 [FREE Full text] [doi: 10.1371/journal.pone.0064126] [Medline: 23717547]
75. Salam RA, Qureshi RN, Sheikh S, Khowaja AR, Sawchuck D, Vidler M, CLIP working group. Potential for task-sharing to Lady Health Workers for identification and emergency management of pre-eclampsia at community level in Pakistan. Reprod Health 2016 Sep 30;13(Suppl 2):107 [FREE Full text] [doi: 10.1186/s12978-016-0214-0] [Medline: 27719680]
76. Paulino NA, Vázquez MS, Bolúmar F. Indigenous language and inequitable maternal health care, Guatemala, Mexico, Peru and the Plurinational State of Bolivia. Bull World Health Organ 2019 Jan 1;97(1):59-67 [FREE Full text] [doi: 10.2471/BLT.18.216184] [Medline: 30618466]

Abbreviations

CIHR: Canadian Institutes of Health Research
eHealth: electronic health
HCP: health care provider
HDP: hypertensive disorders of pregnancy
HELLP: hemolysis, elevated liver enzymes, and low platelet
MEDLINE: Medical Literature Analysis and Retrieval System Online
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 20.06.19; peer-reviewed by J van den Heuvel, L Hinton, L Mackillop; comments to author 21.08.19; revised version received 15.10.19; accepted 24.01.20; published 17.04.20.

Please cite as:
Aquino M, Munce S, Griffith J, Pakosh M, Munnery M, Seto E
Exploring the Use of Telemonitoring for Patients at High Risk for Hypertensive Disorders of Pregnancy in the Antepartum and Postpartum Periods: Scoping Review
JMIR Mhealth Uhealth 2020;8(4):e15095
URL: http://mhealth.jmir.org/2020/4/e15095/
doi:10.2196/15095
PMID:32301744

Original Paper

Design, Development, and Testing of an App for Dual-Task Assessment and Training Regarding Cognitive-Motor Interference (CMI-APP) in People With Multiple Sclerosis: Multicenter Pilot Study

Andrea Tacchino[1], PhD; Renee Veldkamp[2], MSc; Karin Coninx[3], Prof Dr; Jens Brulmans[3], MSCS; Steven Palmaers[4], MSCS; Päivi Hämäläinen[5], PhD; Mieke D'hooge[2,6], MSc; Ellen Vanzeir[7], MSc; Alon Kalron[8,9], PT, PhD; Giampaolo Brichetto[1,10], MD, PhD; Peter Feys[2], PT, PhD; Ilse Baert[2], PhD

[1]Scientific Research Area, Italian Multiple Sclerosis Foundation, Genoa, Italy

[2]Rehabilitation Research Center (REVAL), Faculty of Rehabilitation Sciences, Hasselt University, Hasselt, Belgium

[3]Expertise Centre for Digital Media, Faculty of Sciences, Hasselt University, Hasselt, Belgium

[4]Smart ICT, PXL University College, Hasselt, Belgium

[5]Masku Neurological Rehabilitation Centre, Masku, Finland

[6]National MS Center Melsbroek, Melsbroek, Belgium

[7]Rehabilitation and MS Center Overpelt, Overpelt, Belgium

[8]Department of Physical Therapy, School of Health Professions, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

[9]Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

[10]AISM Rehabilitation Service of Genoa, Italian Multiple Sclerosis Society, Genoa, Italy

Corresponding Author:
Andrea Tacchino, PhD
Scientific Research Area
Italian Multiple Sclerosis Foundation
Via Operai 40
Genoa, 16149
Italy
Phone: 39 3491403917
Email: andrea.tacchino@aism.it

Abstract

Background: Dual tasking constitutes a large portion of most activities of daily living; in real-life situations, people need to not only maintain balance and mobility skills, but also perform other cognitive or motor tasks at the same time. Interest toward dual-task training (DTT) is increasing as traditional interventions may not prepare patients to adequately face the challenges of most activities of daily living. These usually involve simultaneous cognitive and motor tasks, and they often show a decline in performance. Cognitive-motor interference (CMI) has been investigated in different neurological populations, but limited evidence is present for people with multiple sclerosis (MS). The use of computerized tools is mandatory to allow the application of more standardized assessment and rehabilitation intervention protocols and easier implementation of multicenter and multilanguage studies.

Objective: To describe the design and development of CMI-APP, an adaptive and interactive technology tablet-based app, and to present the preliminary results of a multicenter pilot study involving people with MS performed in several European centers for evaluating the feasibility of and adherence to a rehabilitation program based on CMI-APP.

Methods: CMI-APP includes user-friendly interfaces for personal data input and management, assessment of CMI, and DTT. A dedicated team developed CMI-APP for Android tablets above API level 14 (version 4.0), using C# as the programming language and Unity and Visual Studio as development tools. Three cognitive assessment tests for working memory, information processing speed, and sustained attention and four motor assessment tests for walking at different difficulty levels were implemented. Dual cognitive-motor tasks were performed by combining single cognitive and motor tasks. CMI-APP implements exercises for DTT involving the following 12 cognitive functions: sustained attention, text comprehension, verbal fluency, auditory discrimination,

visual discrimination, working memory, information processing speed, auditory memory, visual memory, verbal analog reasoning, visual analog reasoning, and visual spatial planning, which can be performed during walking or stepping on the spot. Fifteen people with MS (mean age 52.6, SD 8.6 years; mean disease duration 9.4, SD 8.4 years; mean Expanded Disability Status Scale score 3.6, SD 1.1) underwent DTT (20 sessions). Adherence to the rehabilitation program was evaluated according to the percentage of performed sessions, perceived exertion during the training (Borg 15-point Ratings of Perceived Exertion [RPE] Scale), and subjective experience of the training (Intrinsic Motivation Inventory [IMI]).

Results: The adherence rate was 91%. DTT was perceived as "somewhat difficult" (mean RPE Scale score 12.6, SD 1.9). IMI revealed that participants enjoyed the training and felt that it was valuable and, to some extent, important, without feelings of pressure. They felt competent, although they did not always feel they could choose the exercises, probably because the therapist chose the exercises and many exercises had few difficulty levels.

Conclusions: CMI-APP is safe, highly usable, motivating, and well accepted for DTT by people with MS. The findings are fundamental for the preparation of future large-sample studies examining CMI and the effectiveness of DTT interventions with CMI-APP in people with MS.

(JMIR Mhealth Uhealth 2020;8(4):e15344) doi:10.2196/15344

KEYWORDS

tablet; mobile device; cognitive rehabilitation; cognitive impairment; dual-task training; cognitive-motor interference; dual-task cost; adherence; multiple sclerosis; walking

Introduction

The conventional approach of physical and cognitive rehabilitation is mainly focused on single-task conditions. Over the past decade, research has devoted increasing attention to dual-task training (DTT) [1] owing to the ascertainment that traditional interventions may not prepare patients for adequately returning to community living (eg, household, school, family, work, and leisure activities). In fact, dual tasking constitutes a large portion of most activities of daily living; in real-life situations, people need to not only maintain balance and mobility skills, but also perform other cognitive or motor tasks at the same time (eg, walking while talking on the phone and rehearsing a shopping list, typing on the smartphone and talking, and preparing meals and talking) [2]. This integrated dual tasking can be defined as the concurrent performance of two tasks that can be executed independently and measured separately and that have distinct goals. This requires adaptive under- or over-additive neural activation in related brain areas [3].

The simultaneous performance of motor and cognitive tasks can be difficult and can lead to worse performance in the motor or cognitive domain or both domains. This cognitive-motor interference (CMI) occurs when performance in a motor or cognitive task decreases on performing a dual task (DT) as compared with performing a single task, which is the so-called dual-task cost (DTC). CMI has been investigated in different neurological populations that usually experience physical and cognitive deficits, including individuals with stroke [4], Parkinson disease [5], and Alzheimer disease [6]. Findings in these populations showed a disproportionate effect of concurrent cognitive tasks on mobility when compared with healthy controls. Moreover, divided attention deficits may prevent neurological individuals from allocating appropriate attentional resources to balance and gait, consequently reducing adaptability to challenging environments (eg, obstacles and uneven paths) and contributing to fall risk [7,8]. Several results indicate that for these neurological conditions, motor and cognitive deficits can be minimized with not only focused single-task training but also targeted DTT [9-11].

Declines during simultaneous performance of cognitive and motor functions are commonly observed in multiple sclerosis (MS) [12,13]. For example, people with MS showed a greatly increased postural sway and a large decrement in variability of anteroposterior and mediolateral sway velocity while executing a simple arithmetic task during balance maintenance as compared with controls [14,15]. Similarly, they showed increased stride time and decreased walking speed during walking under several cognitive conditions (ie, talking) [16-18]. Recently, findings regarding cognitive-motor performances were reported during upper limb tasks [19]. Consolidated evidence of MS rehabilitation regarding walking and cognition (ie, attention, information processing speed, executive function, and long-term memory) [20] independently suggested that the effectiveness on these tasks could be further improved with targeted interventions based on DTT [21]. However, there are still very limited results on the effects of DT rehabilitation strategies in people with MS [22-24], and thus, more clinical and research efforts are required [25]. Standardized research methodology and innovative training programs directed toward meeting the demands of "real-life" situations lack evidence for MS.

Variabilities in task duration, type and complexity of the cognitive task, and training modality (single, consecutive, or integrated dual tasking) limit the availability of standardized testing and training protocols, comparisons across studies, and translation in clinical practice [11]. Therefore, moving toward protocols involving computerized tools is almost mandatory to allow larger sample size inclusion, more standardized protocols for assessment and rehabilitation training, more reliable outcome measures, and easier implementation of multicenter and multilanguage studies.

Moreover, to date, most technological solutions on the market are able to provide multisensory feedback and modulate exercise complexity according to the patient's capacity. However, rarely,

they implement tests and exercises suited and tested for neurological rehabilitative interventions, as well as appropriate interfaces to plan assessment and training in both the cognitive and motor domains independently or together. Moreover, the high costs and lack of portability severely limit their use in clinical practice. For these reasons, the increasing availability of portable devices with adequate memory and calculus performance prompted us to develop a new solution for DT investigations based on a tablet app.

Owing to these considerations, a mobile tablet-based app was proposed, designed, and developed to assess CMI, deliver DT exercises, and investigate DTT effects in people with MS. It allows the application and objective quantification of standardized assessment and rehabilitation interventions, as such opening the "black box" of rehabilitation content.

In this paper, we describe the design and development of CMI-APP, an adaptive and interactive technology tablet-based app, as well as the results of a multicenter pilot study involving people with MS that was performed in several European centers to evaluate the feasibility of and adherence to a rehabilitation program based on CMI-APP, the perceived exertion during the training, and the subjective experience regarding the training. This is fundamental in preparation for future large-sample studies examining CMI and the effectiveness of DTT interventions with CMI-APP in people with MS.

Methods

Study Centers

The participating centers were as follows: Rehabilitation Research Center (REVAL) and Expertise Centre for Digital Media of Hasselt University in Belgium; Italian Multiple Sclerosis Society (AISM) Rehabilitation Service of Genoa and Foundation of AISM (FISM) in Italy; Smart ICT of the PXL University College of Hasselt in Belgium; Rehabilitation and MS Center Overpelt in Belgium; National Multiple Sclerosis Center Melsbroek in Belgium; Masku Neurological Rehabilitation Centre in Finland; AZ Klina, campus De Mick, rehabilitation, Brasschaat in Belgium; Centre Hospitalier Universitaire de Liège in Belgium; and Multiple Sclerosis Center, Sheba Medical Center, Tel-Hashomer in Israel. The study was approved by the Ethics Committee of CHU Liège, Belgium, as well as the local ethics committee of each participating center.

The Expertise Centre for Digital Media of Hasselt University and Smart ICT of the PXL University College of Hasselt developed CMI-APP in collaboration with REVAL. The Centre Hospitalier Universitaire de Liège and Masku Neurological Rehabilitation Centre were involved in the test-retest reliability study of the assessment module of CMI-APP [18]. In addition, therapists from the MS rehabilitation centers in Belgium were involved in the development part of the study. For the multicenter pilot study, five centers recruited people with MS. These centers were AISM; Multiple Sclerosis Center, Sheba Medical Center; National Multiple Sclerosis Center Melsbroek; AZ Klina, campus De Mick; and Rehabilitation and MS Center Overpelt.

Based on clinical experience and the knowledge of researchers and therapists, the types and difficulties of cognitive and motor exercises were discussed during several meetings. If there was uncertainty about the duration of an exercise, a literature search was performed and the approach was tried out in clinical practice. The responses of participants to the Dual Task Questionnaire of Evans et al and their advice in the test-retest reliability study [18,26], as well as the long experience of working with people with MS helped us to identify the needs and appropriate exercises.

Development of CMI-APP

Overview

The initial design and development of CMI-APP was started in 2015, with start-up support from the Flemish Multiple Sclerosis Society. The concept of the app was discussed in a group of rehabilitation scientists and clinicians with MS expertise (ie, physiotherapy and neuropsychology), people with MS, and computer scientists from the Expertise Centre for Digital Media of Hasselt University (Belgium) and Smart ICT of PXL University College (Belgium). The multidisciplinary team regularly met and collaborated by using essential techniques of user-centered design and development, such as iterative development and evaluation of intermediate prototypes. The core development period was from January to December 2016, when the version of the app used in the study was finalized. In the second step, cultural adaptations and translations to other languages of the partners involved in the project (Italian, Hebrew, Finnish, and French beside Dutch and English) were performed to prepare for an international multicenter approach. This was supported by the European network for MS rehabilitation, Rehabilitation in Multiple Sclerosis, and Swedish PROMOBILIA foundation. In particular, in order to maintain consistency among the partner languages, the development team, information technology specialist, and representative of each of the partners continuously interacted for accurate translation and adaption of the text on the objects of the graphical user interface (GUI) (eg, labels and buttons) and for the production of the auditory files used in CMI-APP. To allow more flexibility, the app was designed to be easily extended to other languages.

CMI-APP has been developed for any Android tablet above API level 14 (version 4.0), using C# (Microsoft Corp, Redmond, Washington, USA) as the programming language and Unity and Visual Studio (Microsoft Corp) as development tools. These common platforms facilitate accessibility for the centers and therapists involved in the study and are good choices for possible further development and deployment of the app in rehabilitation practice after the study.

The GUI of CMI-APP was implemented through three different but related modules. The main menu (Multimedia Appendix 1) allows the therapist to add new patients or therapists and to retrieve previously created people for assessment and training. Patients and therapists are added with unique codes (ie, "Patient code" and "Therapist code"). Additional information about the selected patient (eg, visual problems) and the current session (eg, bad sleep) can be added as a note. Furthermore, two numerical text boxes are provided to add the baseline number

of steps in a predefined temporal range (eg, 1 minute) during walking and stepping on the spot, which are both assessed at the first evaluation (eg, through a pedometer). The main menu also allows language selection (currently, Dutch, Italian, French, Hebrew, and Finnish). Finally, there are two buttons "Start assessment" and "Start exercises" to access the modules for assessment and training, respectively.

Assessment Module

Assessment tests for three different cognitive functions (working memory, information processing speed, and sustained attention) at various difficulty levels are implemented in CMI-APP. These are "Titrated digit span backwards," "Auditory vigilance with alphabets," and "Serial counting backwards by 7." In "Menu–Assessment," these three types of tests are available for selection to be performed on their own or in combination with motor tasks. Cognitive tasks were chosen considering that working memory, information processing speed, and attention are among the most affected cognitive domains in MS and considering the results of pilot studies and feasibility during walking [18,27]. Currently, according to clinical and experimental experience, the following four common walking activities, which are carried out in daily life but differ in motor complexity and require attention or adaptation, are included in the testing protocol: walking at a self-selected speed, walking at a self-selected speed while carrying a cup filled with water, walking at a self-selected speed while stepping over various obstacles (eg, 10-cm height, 10-cm width, soft material, and every 3 m in a straight line), and walking crisscross at a self-selected speed from one cone to another (eg, every 2 m with a fixed 80-cm width in between). The motor tasks (actual single or dual motor tasks) were chosen according to the findings of previous studies investigating reliability in persons with neurological conditions during various walking tasks [18,27-31]. It is suggested to perform walking on a 30-m quiet walkway that is free of obstacles and has marked start and turning lines (eg, 80 cm). Before the execution of these tasks, the walking activities should be tried in order to perform them without uncertainty. For all motor tasks, the therapist should demonstrate how to walk over the walkway, and participants should try to walk on a part of the walkway. The different complexities of motor tasks allow for personalization of the difficulty level depending on the individual's ability, the need to train for specific problems, and the disease progression. Thus, the performance of a patient can be assessed under a total of 19 conditions (three single cognitive conditions, four single or dual motor conditions, and 12 dual cognitive-motor conditions) (Multimedia Appendix 2). Each test lasts for 60 seconds, and the result is stored only if it is successfully performed. For safety, it is suggested for the examiner to walk close to but behind the participant. Moreover, it is suggested to put the tablet in a case with a sling wearable over the shoulder, so that, if needed, the therapist can drop the tablet (it will hang on the therapist's neck) without damaging it and catch the patient.

The order in which the blocks of single cognitive, single or dual motor, or dual cognitive-motor tasks are presented, as well as the sequence of each separate task within one block is optimally randomized. To make the assessment easier and more reliable at different time points, the order automatically remains the same for the patient. Multiple conditions should be evaluated for a complete assessment of CMI. In fact, usually, DT performance is assessed through one DT condition/paradigm, which is mostly quantified as motor DTC. DTC may suggest whether and how attention resources, executive functions, and working memory affect a motor task (ie, motor DTC) or whether and how walking affects cognitive tasks (ie, cognitive DTC) [32-35]. Considering that different cognitive or motor tasks compete for cognitive or motor resources to varying extents, using only one cognitive or motor task may not be sufficient to explain CMI in its entirety [18].

Nevertheless, a therapist can decide to administer only a reduced subset of conditions (cognitive, motor, or cognitive-motor). Descriptions of the cognitive tasks are provided below.

Titrated Digit Span Backwards

This mental tracking task requires sustained attention, working memory, and information processing speed. Patients listen to a titrated string of digits (eg, 3-2-5-7-9), which is presented at a rate of one per second, as commonly used in standard neuropsychological tests [36]. The digit order in the string is automatically and randomly generated by an app routine that follows ordered sampling without replacement for the digits 1-9 (eg, for a string length of three digits, the number of orders is 9×8×7=504).

Subsequently, they are requested to repeat the string in reverse order. Before the test, the therapist can define a personalized sequence length with the procedure activated through the button "Assess span length" (Multimedia Appendix 3). Four trials are performed at each sequence length starting from a length of three digits. If three out of four trials at a given length are correct, the patient is considered to have passed the test for that specific sequence length, and the length is increased by one digit. Each patient's digit span length is determined as the largest sequence length for which the patient succeeds in at least three out of four trials. The interface of the digit span test is similar to that of the determination of the titrated span length and is activated through the button "Start the exercise" (Multimedia Appendix 3). Before starting the test, the therapist can ultimately set the digit span length of the patient. After the therapist types the digits responded by the subject and pushes the button "Enter," the next sequence is delivered.

Auditory Vigilance With Alphabets

This reaction time task requires processing speed, with detection of underlying attention deficits. In this test, patients listen to 60 seconds of recorded letters at the presented rate of one letter per 2.5 seconds and have to say aloud "yes" every time they hear the two target letters indicated before starting the trial (a total of 24 letters, of which 10 are target letters). Target letters were chosen as not very common or very rare in everyday speech and not easily confused with other letters (each country has its own version based on some common rules). Each time, the order of 24 letters is automatically and randomly generated by an app routine that firstly extracts a combination with replacement of 14 sequence nontarget letters (ie, 24 sequence letters − 10 sequence target letters) from 24 alphabet nontarget letters (ie, 26 alphabet letters − 2 alphabet target letters) and

secondly randomly combines the 14 extracted nontarget letters with the 10 target letters.

The therapist only has to push the button "YES" when the participants says "yes." False positive (wrong answers) and negative answers (omissions) are automatically counted as incorrect (Multimedia Appendix 4).

Serial Counting Backwards by 7

This mental tracking task requires sustained attention and information processing speed. In this test, patients have to consecutively subtract 7 starting from a given number (different numbers at each measurement time point; the starting number is automatically and randomly selected by an app routine in the range 101-199). For example, take 7 away from 101 (value 94), take 7 away from 94 (value 87), and so on. However, if the patient makes a mistake, but he/she correctly goes on from it (eg, 101, 95, 88, and so on), it is only counted as one mistake, and subsequent numbers are considered accurate. The therapist has to type the responded number and push "Enter" to save it (Multimedia Appendix 5). The number of correct subtractions is automatically counted.

Training Module

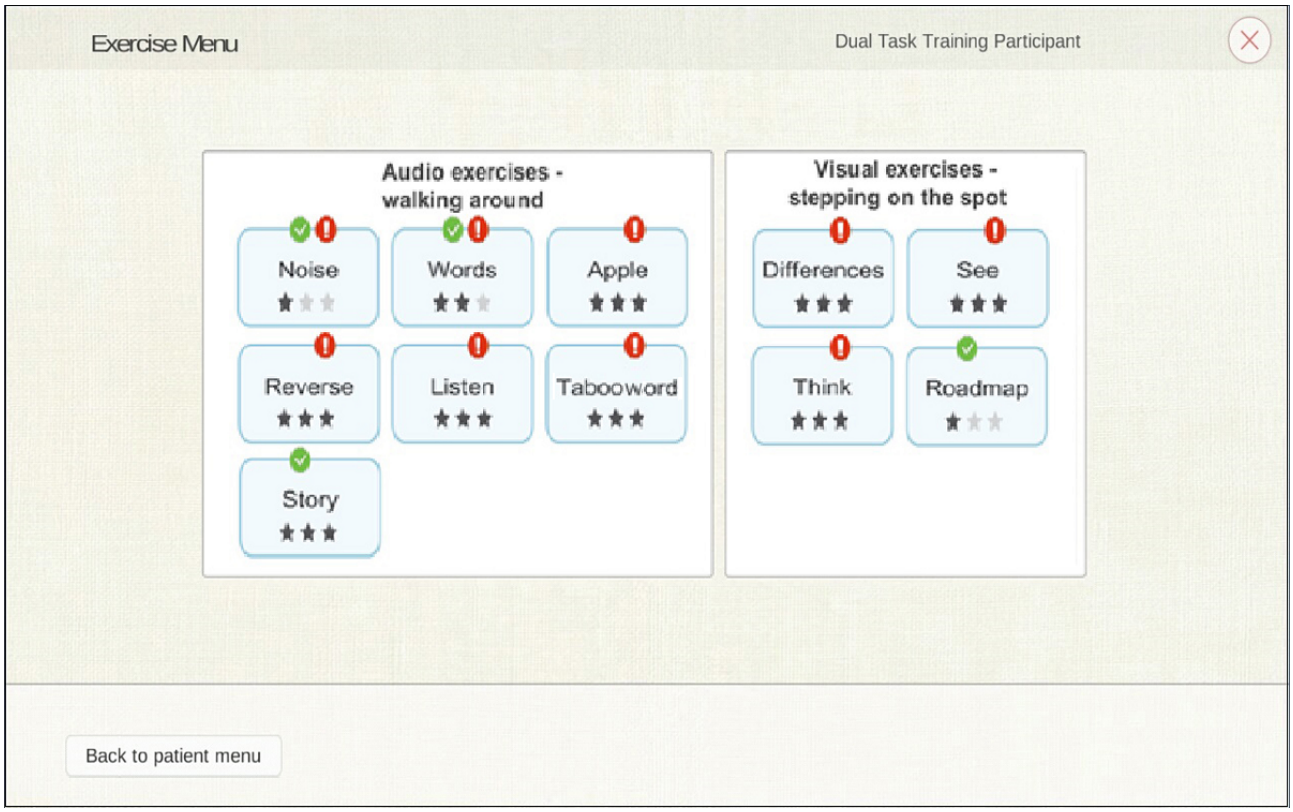
CMI-APP offers easy access to DTT interventions for people with MS and therapists, as it is conceptualized in such a way that cognitive exercises are combinable with selected motor tasks.

Training exercises for 12 different cognitive functions (sustained attention, text comprehension, verbal fluency, auditory discrimination, visual discrimination, working memory, information processing speed, auditory memory, visual memory, verbal analog reasoning, visual analog reasoning, and visual spatial planning) are implemented in CMI-APP. Detailed descriptions of the 11 exercises can be found in Table 1. "Noise," "Words," "Apple," "Reverse," "Listen," "Tabooword," and "Story" can be performed while walking, whereas "Differences," "See," "Think," and "Roadmap" are better suited during stepping on the spot (Figure 1).

Table 1. Cognitive exercises implemented in CMI-APP.

Exercise type	Cognitive function	Description
Exercises with auditory stimuli while walking		
Noise	Auditory discrimination	Recognizing two to four target noises over different sounds/noises.
Words	Working memory and information processing speed	After hearing a given word, formulating a new word with the first, last, second, or fourth letter of the given word.
Apple	Sustained attention	Reaction to one or two target word(s) over semantically equal or semantically different words.
Reverse	Working memory and information processing speed	After hearing a given word of three to seven or more letters, spelling the word in reverse (letter by letter).
Listen	Auditory memory	Each time a new word in a list of words is heard, saying if the word was already heard or not
Tabooword	Verbal fluency	Describing a target word (without using one or three forbidden taboo words) while following some rules.
Story	Text comprehension	After hearing a story, responding to three multiple choice questions about the story.
Exercises with visual stimuli while stepping on the spot		
Differences	Visual discrimination	While seeing two images, saying if the images are the same or different in a given time.
See	Visual memory	After seeing a smiley, saying which smiley is just seen among three presented.
Think	Verbal and visual analog reasoning	Making associations between pictures, solving assignments, and completing logical sequences.
Roadmap	Visual spatial planning	After seeing a roadmap with locations, roundabouts, houses, and trees, at each intersection, saying which direction to go to reach the given destination.

Figure 1. Selection and start of dual-task training. The interface is split into two parts as follows: audio exercises mainly executable by walking around (ie, "Apple," "Listen," "Noise," "Reverse," "Story," "Tabooword," and "Words") and visual exercises mainly executable by stepping on the spot (ie, "Differences," "Road map," "See," and "Think"). The number of dark stars indicates the exercise difficulty level (three stars indicate level 3). When the exercise is performed in the current session, it is marked with a green check mark. When the exercise is performed in the previous session, it is marked with a red exclamation mark.



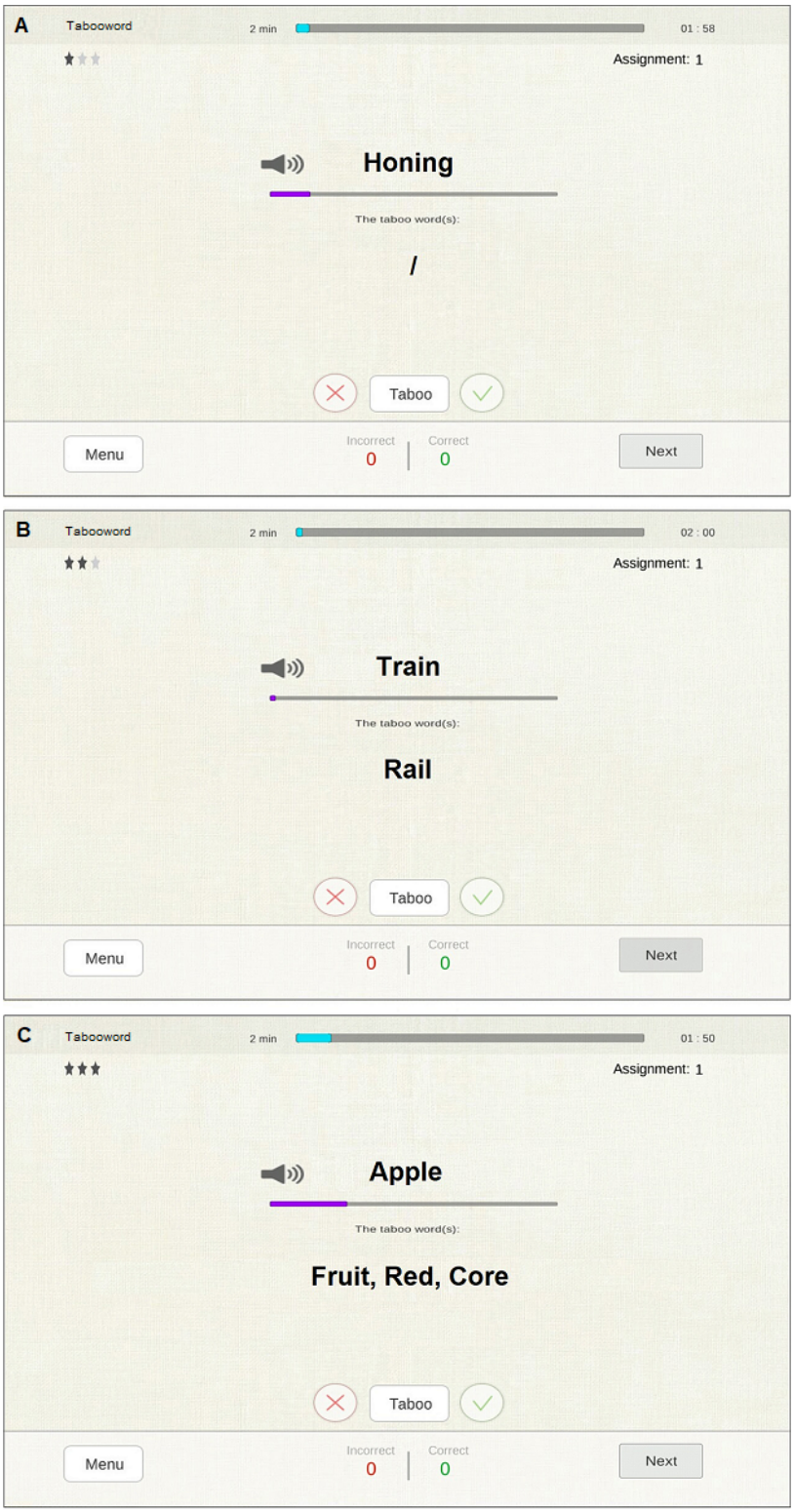
As examples, we present the following three exercises below: "Tabooword" for verbal fluency, "Differences" for visual discrimination, and "Roadmap" for visual spatial planning.

Tabooword

In this exercise, the patient hears a word (ie, the guess word) that he/she then has to describe to the therapist. However, the patient has to follow some rules and specifically cannot (1) use the word itself or parts of the word; (2) use words that are derived from the word; (3) use gestures and noises; (4) use abbreviations, initials, or clues like "sounds like" and "rhymes with;" and (4) use the taboo words indicated before exercise start. For example, if the guess word is "apple" and taboo words are "fruit," "red," and "core," the possible solutions are "eat healthy," "snow white ate an," and "Jonagold, Granny Smith."

At the low difficulty level, the patient has to describe a word in 20 seconds and there are no taboo words. At the medium level, the patient has to take into account only one taboo word and the description is required in 30 seconds. At the high level, three taboo words are presented and the test lasts 40 seconds (Figure 2).

Figure 2. Tabooword exercise. (A) Low difficulty level (20 seconds for description, no taboo words); (B) medium level (30 seconds for description, one taboo word); (C) high level (40 seconds for description, three taboo words).

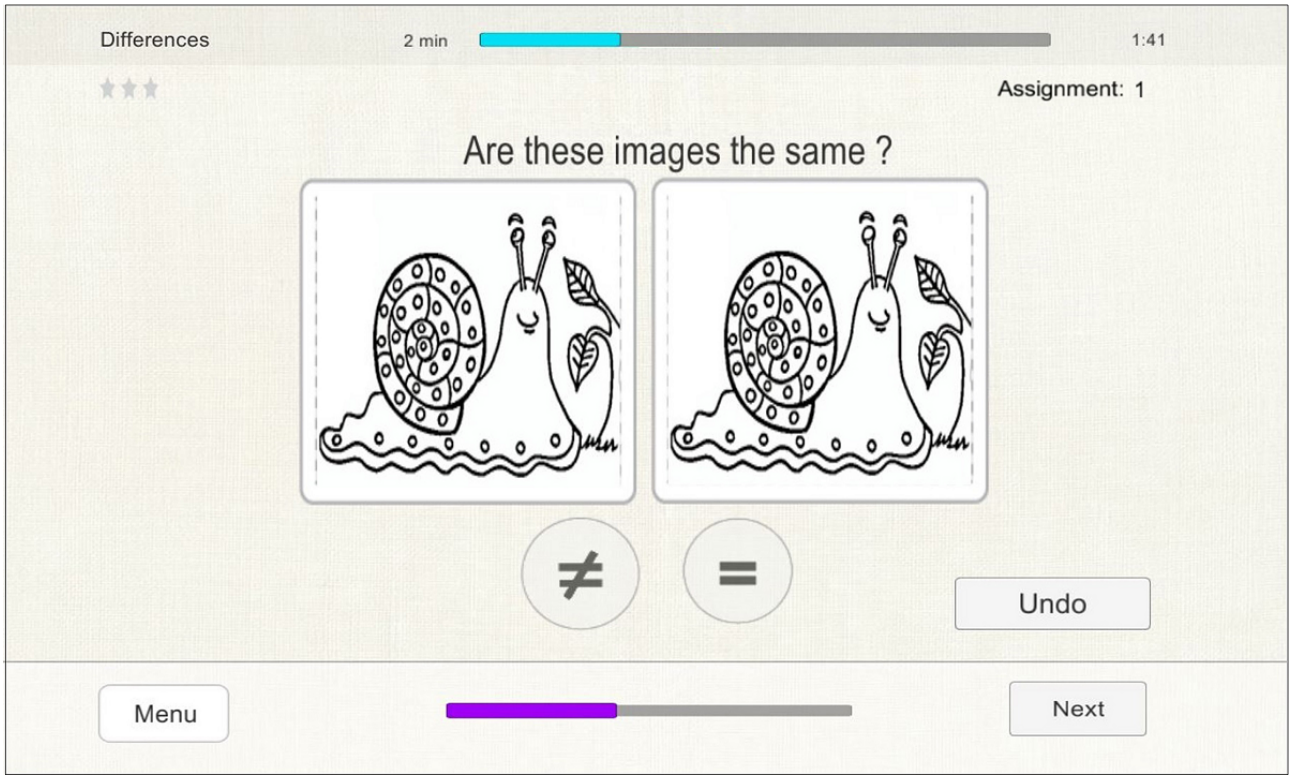


The therapist judges the clarity of the description by pressing score buttons as follows: green for a good description within the time limit and red for a bad description. The "Tabooword" button has to be pressed when the patient breaks one of the five rules (eg, uses the word apple to describe an apple tree). The "Tabooword" button can be used three times before the full assignment is scored as wrong; thus, each time the "Tabooword" button is used, the patient receives minus one-third of the score. When the therapist is satisfied with the score before the time limit is reached, the next assignment can be manually provided. Feedback on the marked mistakes can be delivered before advancing to the next assignment. When the timer runs out, the assignment is automatically scored as incorrect. When the patient wants to skip the assignment without trying, the therapist can press the "Next" button without using the score buttons.

Differences

In this exercise, the patient is shown two pictures and has to tell the therapist whether these pictures are the same or different (Figure 3). Specifically, at the low difficulty level, the patient has 15 seconds to judge the equality and pictures contain more than one difference. At the medium level, the patient has 20 seconds to judge the equality and pictures contain only one difference. At the high level, the patient has 30 seconds to judge the equality and pictures contain one small difference. The patient does not need to refer to the differences and needs to only state whether the pictures are the same or different. If the patient answers "different," the therapist has to press the button with the different mark (≠). If the patient answers "same," the therapist has to press the button with the same mark (=). After entering the answer, the "≠" and "=" buttons disappear and the therapist has to proceed to the next assignment by pressing the button "Next." The same button can be used if the patient wants to skip the current assignment. If time runs out, the assignment is marked as incorrect.

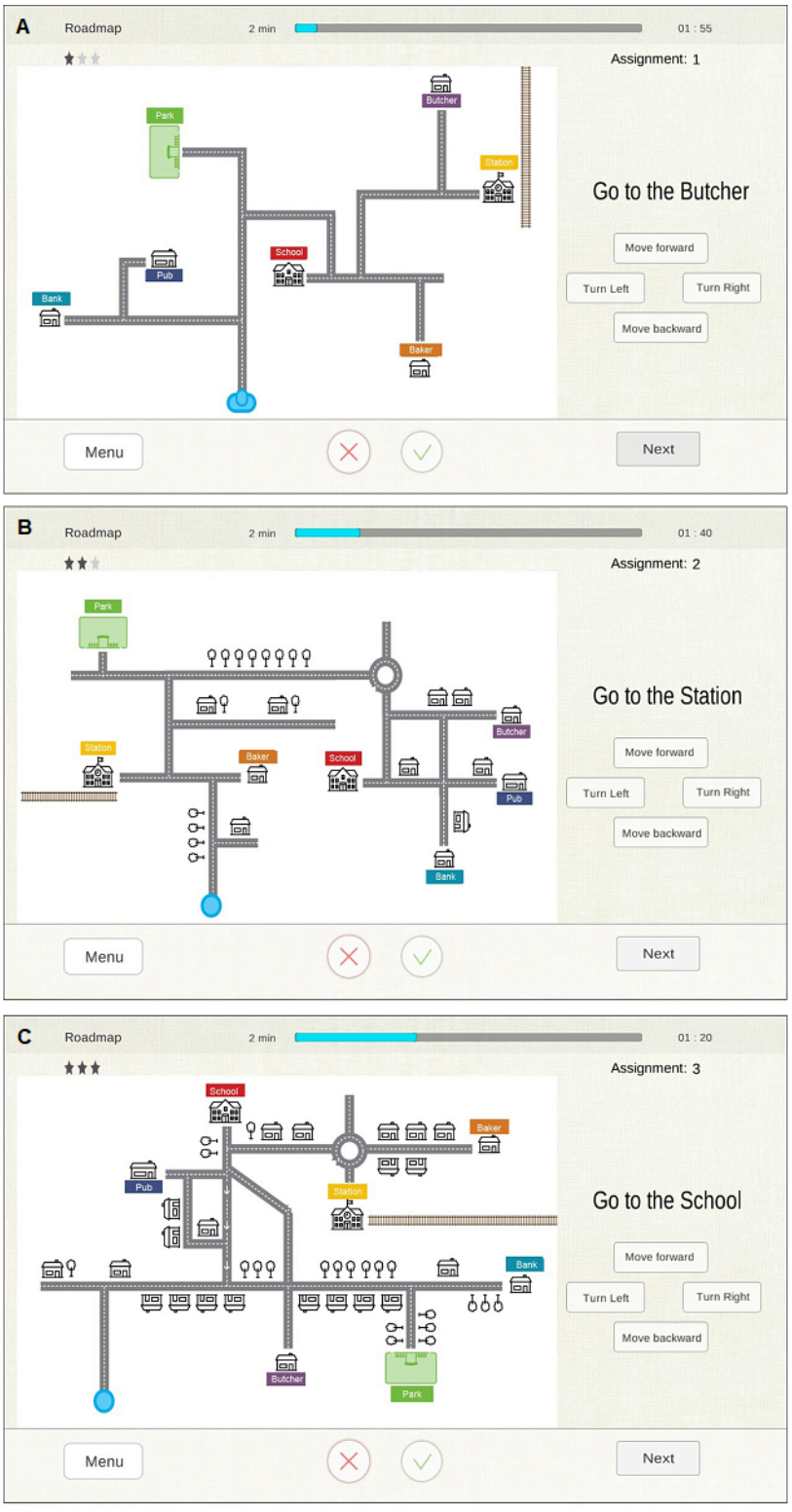
Figure 3. Differences exercise. Two pictures of the high difficulty level are matched.



Roadmap

In this exercise, the patient is shown a roadmap with seven locations (eg, butcher, park, and school). The current location of the patient is indicated by a blue icon (person symbol) (Figure 4). On the right side, the patient can see the destination of the route (eg "Go to the butcher"). The patient should tell the therapist at each intersection which direction he/she wants to go (ie, forward, backwards, to the left, or to the right) in order to move in the direction of the destination. The therapist has to enter this on the tablet, and the patient's location will change on the map. The patient has to continuously pay attention to the orientation in order to correctly indicate the direction in which he/she wants to go. The low difficulty level is implemented without map distractors and with a head as the indicator (blue icon) to help with orientation. The medium level involves roundabouts, houses, and trees as distractors on the map and a dot as the indicator (blue icon; invisible person's orientation). The high level involves roundabouts, houses, trees, and one-way streets as distractors on the map and a dot as the indicator (blue icon; invisible person's orientation).

Figure 4. Roadmap exercise. (A) Low difficulty level (without map distractors and with a head as the blue icon to help with orientation); (B) medium level (with roundabouts, houses, and trees as distractors on the map and a dot as the blue icon [invisible person's orientation]); (C) high level (with roundabouts, houses, trees, and one-way streets as distractors on the map and a dot as the blue icon [invisible person's orientation]).



The patient reads the assignment/target location and starts navigating by telling the therapist which direction he/she wants to go on the map. The therapist enters the patient's direction by pressing the corresponding button. The therapist does not need to remember or compensate for the orientation of the patient on the map. However, when the patient states a direction, he/she has to take into account his/her own mental orientation. For this reason, in order to better stimulate visual spatial planning, it is suggested to not allow changes in physical orientation. When the patient reaches the target location, considers the assignment complete, or wants to stop the assignment, the therapist scores the assignment by using the score buttons. The red cross indicates that the patient failed the assignment (ie, standing at the wrong location or making mistakes during the route). The

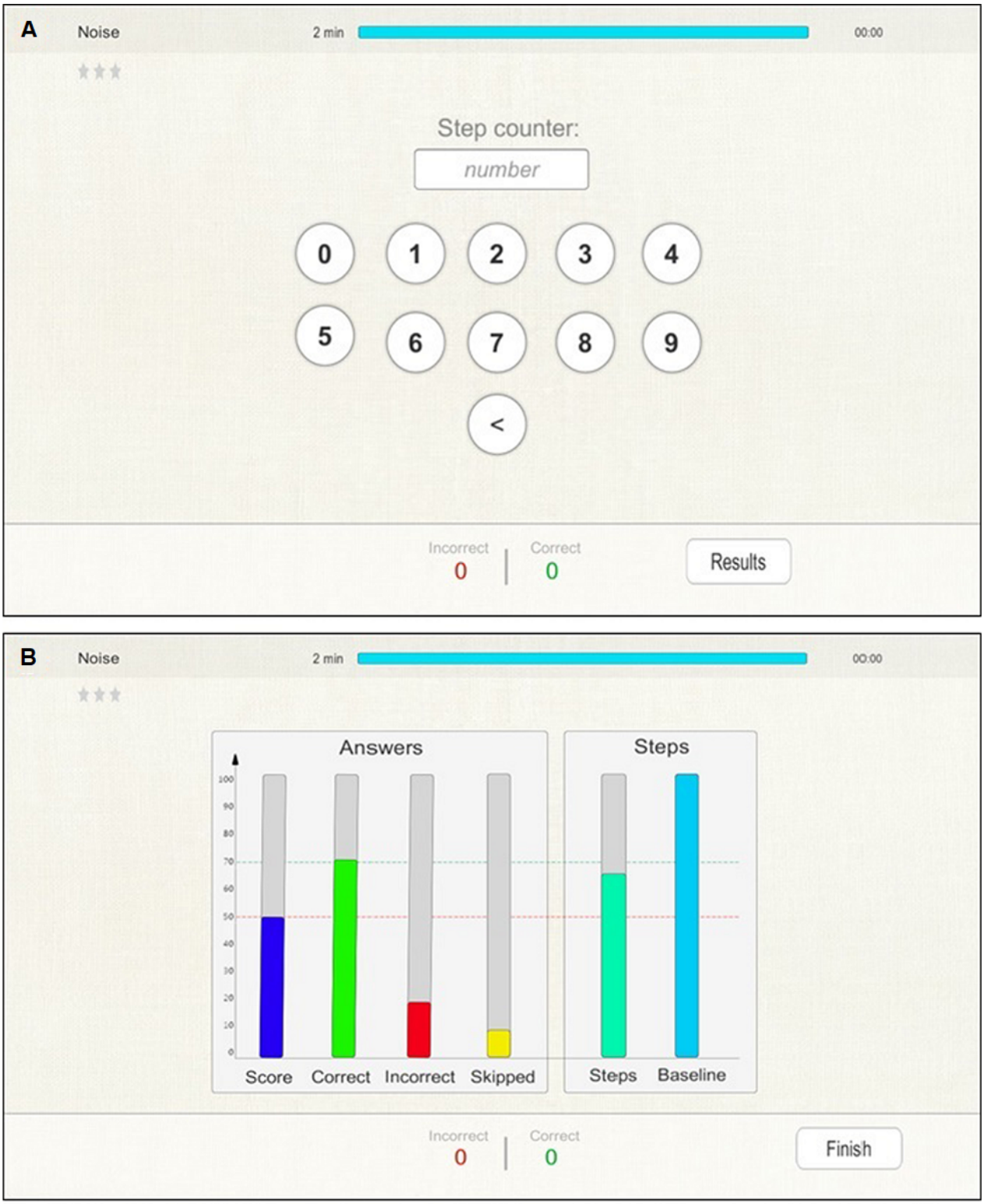
green mark indicates that the patient passed the assignment (ie, correct location and correct route). After entering a score, the score buttons disappear.

Feedback

At the end of each exercise, the therapist is obligated to enter the number of steps counted during walking or during stepping on the spot (Figure 5). In order to make counting of the steps and training administration feasible, a pedometer was used. After entering the number of steps, the "Results" button allows to continue to the visual overview of the results from the current exercise. The interface offers feedback on the cognitive and motor performances during dual tasking. In particular, the cognitive performance is displayed with colored bars; the first representing a percentage composite score calculated according to the numbers of correct, incorrect, and skipped answers, and the others representing the percentages of correct, incorrect, and skipped answers for the given assignments. Two more colored bars represent the percentages of the steps per minute during the exercise and at baseline. It should be taken into account that if the number of steps is smaller than at baseline, the baseline bar is 100% and the steps bar is the percentage of steps with reference to the baseline. On the other hand, if the number of steps is greater than at baseline, the steps bar is 100% and the baseline bar is the percentage of steps per minute during the exercise. To provide more complete information to the therapist, the absolute values of the correct and incorrect answers during the exercise are displayed at the bottom. A red line indicates a percentage performance of 50%, and values below 50% for cognitive score and steps are considered for level down in the next session. A green line indicates a percentage performance of 70%, and values above 70% for cognitive score and steps are considered for level up in the next session (Figure 5).

A visual overview of the results could be displayed to the patient if it is considered a valuable way to stimulate his/her cognitive and physical performances.

Figure 5. Feedback exercise. (A) The interface to enter the number of steps and the "Results" button to continue to the visual overview of the results as shown in (B). (B) Six colored bars represent the percentage composite score of correct, incorrect, and skipped answers; the percentages of correct, incorrect, and skipped answers for the given assignments; and the percentages of the steps per minute during the exercise and at baseline. A red line indicates a percentage performance of 50%, and values below 50% for cognitive score and steps are considered for level down in the next session. A green line indicates a percentage performance of 70%, and values above 70% for cognitive score and steps are considered for level up in the next session.



Routines for Auditory Files, Logs, and Data Storage

Auditory files (ie .wav format) are delivered by customized routines. During training exercises, if the setup includes the use of a wireless headset microphone (eg, Logitech H800 USB Wireless Headset with Noise Cancelling Microphone; Logitech, Lausanne, Switzerland), the patient's responses can be recorded. Similarly, during assessment, logs with assignments, patient's answers, response accuracy, and .wav files were recorded. The audio recording could be used to later calculate the percentage of correct answers in case of suspicion of typographical errors by the therapist. Data recorded during the CMI assessment and

DTT were sent via wireless service from each center to a central server in order to store data for subsequent analyses.

Instruction Booklet

Although the cognitive tasks implemented in CMI-APP could be useful for both outpatient and at-home cognitive rehabilitation, DTT involving walking and stepping on the spot (to be safe for neurological conditions) should be performed in a clinical setting and delivered by a trained therapist. For this reason, to make the use of CMI-APP more practical, safer, and widespread, we provide an instruction booklet for assessment and training. Owing to the user-friendly and tailor-made interface of CMI-APP and the easiness for learning how the cognitive, motor, and cognitive-motor tasks have to be executed, the booklet is a valid alternative to a training course for therapists. In particular, the booklet shows the screens encountered in CMI-APP with annotations and instructions necessary to complete each screen for each assessment or training exercise.

Study design

Overview

In order to evaluate the feasibility and adherence to a rehabilitation program based on CMI-APP, perceived exertion during the training, and subjective experience regarding the training and the training activities, a pilot test was performed in people with MS. The information presented here is a small part of a larger study on the assessment and DTT of CMI in people with MS. The whole study consisted of an assessment study with a test-retest design and an intervention study as described previously [18,37]. The intervention study was a multicenter, randomized, two-arm, controlled trial consisting of the integrated DTT as provided by the CMI-APP and a single mobility training group. The results of this study on DT and cognitive and motor performances are published in another manuscript [37]. Here, the development and technical details of the CMI-APP are described. Furthermore, a subsample of the intervention study sample receiving DTT was analyzed for adherence, perceived exertion, and subjective experience regarding the training, and the data are presented.

Patients

A total of 15 people with MS (10 women and five men) were recruited from the participating centers (five from Belgium, seven from Italy, and three from Israel). The inclusion criteria were as follows: (1) diagnosis of MS according to McDonald criteria [38]; (2) all types of MS; (3) age between 18 and 65 years; (4) Expanded Disability Status Scale (EDSS) [39] score ≥2 and ≤5, as determined by neurologists or trained clinicians, which includes ability to walk without a waking aid or rest for 200 m; (5) no relapse within the last 30 days; (6) no changes in disease-modifying treatment and no corticoid therapy within the last 50 days; and (7) appropriate cognitive capacity measured by the Mini-Mental State Examination (MMSE) (score ≥26) [40]. The exclusion criteria were as follows: (1) other medical conditions interfering with mobility (eg, acute/subacute fractures and pregnancy >20 weeks); (2) other neurological diagnoses (eg, stroke and Parkinson disease) or MS-like syndromes (eg, neuromyelitis optica); (3) inability to understand and execute simple instructions; (4) problems with hearing or vision interfering with assessment or training (even after adjustment with hearing aids or glasses); and (5) ongoing DT training, other interfering physical therapy, or cognitive training/neuropsychological rehabilitation (eg, balance and walking rehabilitation, occupational therapy, cognitive rehabilitation, and speech rehabilitation).

All participants in this pilot study provided informed consent. The study was conducted in accordance with the Declaration of Helsinki (1964) [41].

Training Program

The participants performed outpatient DTT of 20 sessions, with a frequency of five times over 2 weeks (two or three times in the first week and three or two times in the second week) during 8 weeks. Each session lasted 45 to 60 minutes with a total DTT time of 30 minutes. DTT consisted of the execution of the cognitive exercises implemented in CMI-APP (Table 1) while participants walked or stepped on the spot. Training was performed in a dedicated rehabilitation room and supervised by a therapist specifically trained to tailor make the cognitive exercises with difficulty progression according to feedback performance of the previous training session, safety, quality consideration, and patient preference regarding the exercises. Progression to a higher difficulty level was allowed if the number of steps per minute during the exercise was ≥70% of baseline and if the accuracy of answers in the exercise was ≥70%. Moreover, the therapist decided the level increment if the safety and quality judgements of participant performance were respected. Regression to a lower difficulty level occurred if the number of steps per minute during the exercise was <50% of baseline or the accuracy of answers in the exercise was <50%, or for safety/quality reasons according to therapist judgement. The therapist counted the steps using a simple user-friendly pedometer (SW200 Digi-Walker Pedometer, Yamax, Bridgnorth, UK), which could be compared to those recorded at the first evaluation.

For each exercise, participants always started from the first level provided by CMI-APP. If the performance met the criteria for progression, the second level was adopted. Participants were advanced to the third level or downgraded to the first level if the exercise performance was good (advanced) or bad (downgraded) thrice (not necessary consecutively). To avoid overloading the participants, only five out of 12 cognitive functions while walking at usual pace or stepping on the spot were trained per session (eg, working memory, information processing speed, auditory memory, visual discrimination, and visual memory trained with four exercises, such as “Words,” “Listen,” “Differences,” and “See”). Subjects were instructed and encouraged to perform both tasks as good as possible and were free to prioritize. This might allow people to decide unconsciously which task to prioritize, as in everyday life. This aspect was considered in line with the study by Silsupadol et al [42] showing that in elderly adults, variable-priority training (ie, no instructions to prioritize either the motor or cognitive task) was more effective for improvement in mobility or cognitive outcomes under DT conditions than fixed-priority

training (ie, instructions to prioritize either the motor or cognitive task).

Outcomes

Adherence is the extent to which the number of training sessions completed by the patient corresponds with the number of sessions of DTT. The patients were involved in a single-arm pilot study, and they could choose to not execute the sessions and eventually drop out. Any adverse effects (eg, falling and pain) during the training period were also recorded.

After each training session, perceived exertion during the training was assessed using the Borg 15-point Ratings of Perceived Exertion (RPE) scale [43]. The scale score ranges from 6 to 20, where 6 indicates "no exertion at all (rest)" and 20 indicates "maximal exertion." It should reflect how heavy or strenuous the exercise was according to the patient, combining all sensations and feelings of physical stress, effort, and fatigue.

After the 20 sessions of the training program, a 30-item questionnaire, the Intrinsic Motivation Inventory (IMI) [44], was administered to assess the patient's subjective experience regarding the trained activities. This instrument, through several subscales, assesses the participant's interest/enjoyment, perceived competence, effort/importance, felt pressure and tension, value/usefulness, and perceived choice while performing the training. The possible answers in the multiple-choice questions ranged from 1 ("not at all true") to 7 ("very true"). However, several reverse items were present. For these items, the response score was subtracted from 8, and the result was used as the item score. The items of each subscale were interest/enjoyment (score range 5-35); perceived competence (score range 5-35); effort/importance (score range 5-35); pressure/tension (score range 5-35); value/usefulness (score range 6-42); and perceived choice (score range 4-28). The subscale scores were obtained by summing the scores of the items of each subscale; higher scores indicated positive subjective experience.

Moreover, at the end of the training, an open-ended survey on training perception was administered to both the patient and therapist. The patient and therapist had to refer explicitly to what were the strong and weak aspects of the proposed training, and patients were asked if they prioritized tasks during the exercises.

Results

The mean age of the patients was 52.6 years (SD 8.6, range 34.9-63.7), and the mean disease duration was 9.4 years (SD 8.4, range 0.8-25.1). Among the 15 patients, nine had a relapsing-remitting form of MS and six had a progressive form. The mean EDSS score was 3.6 (SD 1.1, range 2-5). Among the participants, 40% (6/15) had a bachelor's or master's education level, 27% (4/15) had a tertiary education level, and 33% (5/15) had an upper secondary or lower education level. The mean body mass index was 26.2 (SD 4.3). All patients showed appropriate cognitive capacity (mean MMSE score 28.7, SD 1.3; range 26-30), although they reported presence of DT interference (mean DT screening list 4.7, SD 2.7; range 0-9).

Most patients performed 20 training sessions (median 20, IQR 16-20), with a mean adherence of 91% (18.1/20). No adverse effects of the DTT were reported in any of the participants. On average, participants perceived the DTT as "somewhat hard" as shown by a mean RPE score of approximately 13 (mean 12.6, SD 1.9; range 8.8-16.1).

As recorded with the IMI, participants in general enjoyed the training (IMI interest/enjoyment: mean 27.5, SD 5.1) and felt that it was valuable and, to some extent, important (IMI value/usefulness: mean 31.1, SD 9.5; IMI effort/importance: mean 23.5, SD 7.8), without feelings of pressure (IMI pressure/tension: mean 8.2, SD 3.7). Additionally, participants had feelings of competence (IMI perceived competence: mean 27.1, SD 5.3). They did not always feel that they could choose the exercises (IMI perceived choice: mean 19.3, SD 6.2), probably because the choice was usually made by the therapist and because some exercises have limited differentiation among several difficulty levels (eg, the participant rapidly reached the most difficult level or the step between two consecutive difficulty levels was too large).

In general, the patients with MS and therapists were positive about the DTT program. The patients perceived it as useful, challenging, interesting, and fun. They made the following statements: "it is useful, because people can train something that is related to daily life activities," "it is challenging and interesting, because of the combination of both tasks, walking and cognition," "more levels make the training very challenging, and when I perceived that the performances were improving, I was very satisfied," and "the work on walking and memory is a positive aspect of the training." The therapists also indicated that they want to keep using the system, because of "its novelty," "the similarity to daily living," and "the feeling that patients enjoy the training and already perceive improvements during training." The main weak aspect, reported by both patients and therapists, was related to the too small differentiation among difficulty levels of some exercises that could make the training boring.

Discussion

Principal Findings

Research focused on finding new ways to administer standardized DT assessments and DTT rehabilitation interventions of high quality, make them more effective, and ensure high adherence to treatments is mandatory. For this reason, rehabilitation researchers (ie, physicians, therapists, and computer scientists) should define, design, and develop new tools that are able to assess CMI and deliver DTT. Portable and low-cost technology-based products, such as mobile phones and tablets, are the main candidates for these aims [45,46]. In fact, the benefits of adopting electronic devices instead of traditional pen and paper tools depend on several factors, such as dynamic presentation of the stimuli (eg, speed and difficulty levels according to an individual's specific needs and progression in training), more reliable recording of cognitive and perceptual performance (eg, reaction time), standardization of the test and training environment (eg, reduced or null errors in administration), availability of faster feedback and behavioral

information (eg, the time spent on each item), and reduced time in delivering and scoring exercises [37,47]. Moreover, owing to the unavoidable requirement of the execution of motor and cognitive tasks at the same time, the use of traditional computerized tools could limit or make both CMI assessment and DTT impossible. Computer-based cognitive rehabilitation systems that are usually implemented on laptops and desktop computers are not adaptable for DT exercises requiring, for example, walking or stepping on the spot, although they have been shown to be effective in improving cognitive functions [48-51]. Moreover, virtual reality or exergaming may be beneficial to improve DT performance; however, these devices are quite expensive and therefore not available in all clinical settings. Furthermore, they may not be sufficiently adaptive for people with MS having moderate-to-severe disability.

Owing to these considerations, CMI-APP, a tool based on economic, accessible, and widely-used technology (ie, tablets), was proposed to assess CMI, deliver DT exercises, and investigate DTT effects in people with MS. The tablet-based app CMI-APP implements exercises suited and tested for neurological rehabilitative interventions and is conceptualized in such a way that cognitive exercises are combinable with selected motor tasks, with presentation in an easy-to-use GUI to plan assessment and training in both cognitive and motor domains.

Here, we described the design and development of CMI-APP, as well as the results of a multicenter pilot study involving people with MS that was performed to evaluate adherence to a rehabilitation program based on CMI-APP, perceived exertion during the training, and subjective experience regarding the training.

From a technical point of view, CMI-APP provides an interactive, adaptive, user-friendly, and tailor-made interface that can help the therapist to better focus on a patient's safety and quality of performance during DTT, without bothering about inventing new exercises, assignments, and the correctness of provided answers. Moreover, owing to these aspects of design and development, more standardized assessments of CMI and DTT and the proposal of well-designed randomized controlled trials are strongly warranted. Feedback on performance, according to the implemented cutoff values, and the variety of exercises were taken into account during the design of the app in order to improve the engagement and motivation of the patient and support the progression of the difficulty level [37]. Additionally, a clear overview of what has really been trained (ie, dosage and content) is provided through easily accessible training output logs. In addition, because several cognitive domains are trained, particular impairments in daily life may be more easily and timely identified.

The results show that this new system was very well received by patients with MS, as deduced by the high adherence to the treatment. In fact, 91% of all scheduled training sessions were completed by the patients, suggesting that this tool could be proposed for a DTT intervention in people with MS. Importantly, no adverse effects of the DTT were recorded.

On average, participants perceived the DTT as somewhat difficult. Nevertheless, participants reported that they enjoyed the training and were interested in practicing the exercises again because they considered the exercises valuable and important for preserving or improving DT performance. This result seems to be confirmed by reports recorded with an open-ended survey on training perception, revealing that training was perceived as useful, challenging, interesting, and fun. Although the patients felt that they were only relatively involved in the choice of the exercises and the difficulty level more suitable for their own abilities, they felt having competence in the execution of the proposed exercises. Probably, as a consequence, the level of stress due to the training was low, as shown by the low level of perceived pressure and tension. Although a larger sample-size study could definitively shed light on the clinical effects of CMI-APP, we are confident of the reliability of the observed usability results because our sample size matches that of previous studies on the usability of apps for MS cognitive rehabilitation [45].

Accordingly, all the therapists involved in the DTT of the 15 participants had positive feedback on CMI-APP, considered it very user friendly, and had full interest in implementing the proposed training in the clinical routine, including self-use by patients, if the aspects that they reported as weak (eg, limited differentiation among difficulty levels) were addressed in a future release of the app.

Although adherence to the treatment was very high, perception of the DTT was tolerable, and subjective experience regarding the trained activities was high, improvements in CMI-APP should be considered for use in research programs and for translation into clinical practice. In particular, it is recommended to have a wider variation in diverse assignments (eg, more words, stories, and pictures), as well as a better differentiation between the difficulty levels and an increase in the number of exercises over the diverse cognitive domains (eg, more linguistic tasks, such as verbal fluency and alternating alphabets).

CMI-APP has been recently used to assess the test-retest reliability of the 12 CMI paradigms, and the results were recently published [18]. The highest reliability was found for the motor DTC under all walking conditions (walking, walking with a cup, and walking over obstacles) in healthy controls, but the strongest was for walking alone in people with MS. Cognitive DTC appeared to not be reliable in either healthy controls or people with MS. These findings will be taken into account in future developments of CMI-APP.

Before using CMI-APP in clinical practice, large-sample studies that examine CMI and the effectiveness of a DTT intervention in people with MS, establishment of the optimal dosage per exercise, and adaptation of the thresholds for progression and regression of the difficulty level are mandatory.

Moreover, use extension to other neurological pathologies, such as stroke, Parkinson disease, and Alzheimer disease, that are shown to benefit from DTT [9-11] could be proposed and realized owing to the easy adaptability of the system to disease-dependent requirements in terms of technical aspects and exercise features.

Limitations

There are some study limitations. First, not all the cognitive functions trainable with CMI-APP show corresponding exercises for assessment and not all motor conditions available for assessment were suggested to be adopted during the training, limiting the possibility to evaluate task-specific learning. For example, despite the use of different exercises based on vision in the training module, no visually-based exercises were included in the assessment module of CMI-APP. Even if visuospatial tests are very difficult and unsafe for execution during walking, future versions of CMI-APP should allow the execution of tasks based on vision during walking, eventually suggesting convenient set-up adaptation to preserve safety (eg, projection on the wall in front of the patient during walking on a treadmill) [52]. However, we want to clarify that the first version of CMI-APP was developed to evaluate if a carry-over general effect of training was present on CMI, and consequently, no assessment exercises specific for the training exercises were strictly required.

Second, for use in both clinical and research settings, advise to therapists will be implemented to stimulate them to train patients to walk over obstacles, carry a cup, and walk crisscross under DT conditions; moreover, the simple walking motor training condition allows the assessment of potential transfer involving more complex motor tasks (ie, walking over obstacles, carrying a cup, and walking crisscross).

Finally, to better understand the usefulness of CMI-APP and its potentiality in the market [53], a technological acceptance model [54] and client satisfaction scale [55] should be considered and administered to both patients and therapists in a study with a larger sample size.

Conclusion

In this study, we demonstrated that CMI-APP is a tool that is safe, highly usable, motivating, and well accepted by people with MS for motor-cognitive DTT. In fact, the participants in this multicenter pilot study perceived the exercises implemented in CMI-APP as interesting, valuable, and useful to stimulate motor-cognitive abilities usually involved in daily activities. Moreover, the feeling of competence and the absence of perceived pressure are aspects that could improve the self-efficacy of people with MS [56].

Owing to these results, we are now ready for a large-sample study that examines the effectiveness of a DTT intervention with CMI-APP in people with MS, using specific clinical outcomes for motor, cognitive, and motor-cognitive performances. Moreover, despite the market presence of technological solutions providing multisensory feedback and having the ability to modulate exercise complexity, we think that positive results from a randomized controlled trial on the use of this app implementing exercises specifically suited for people with MS will suggest the most effective feedback events (eg, feedback on cadence during exercise) [37] and will play a key role in the actual exploitation of the app in the field of MS as a tool for cognitive-motor rehabilitation. However, if researchers of other neurological pathologies show interest in CMI-APP, new customized versions will be provided.

Acknowledgments

The study was partially funded by an unrestricted educational grant from Novartis Pharma AG to Rehabilitation in Multiple Sclerosis, Swedisch PROMOBILIA foundation, and Flemish MS Liga. The authors would like to thank Marijke Vandermaesen for assistance in the development of CMI-APP.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The “main menu” of CMI-APP. The interface allows selection of language; addition of a new patient code (entry of patient code, number of baseline steps measured in walking/stepping within one minute, and additional information of the patient) or selection of an already added patient from the drop down menu; creation of a new therapist code (entry of therapist code) or selection of an already added therapist from the drop down menu for the current session; addition of a session note for the current session (optional information); and advancement to the cognitive-motor interference assessments or the dual-task training exercises.
[PNG File , 2760 KB - mhealth_v8i4e15344_app1.png]

Multimedia Appendix 2

“Menu–Assessment” of CMI-APP usable by the therapist to select the next test for the cognitive-motor interference assessment. The performances of a patient can be assessed under a total of 19 conditions as follows: three single cognitive conditions, "DigitSpan–None," "Subtraction–None," and "Vigilance–None;" four single/dual motor conditions, "None–Cup," "None–Obstacles," "None–Crisscross," "None–Walk;" 12 dual cognitive-motor conditions, "DigitSpan–Cup," "Subtraction–Cup," "DigitSpan–Crisscross," "Subtraction–Obstacles," "Subtraction–Crisscross," "Vigilance–Crisscross," "Vigilance–Walk," "DigitSpan–Walk," "Vigilance–Obstacles," "DigitSpan–Obstacles," "Subtraction–Walk," and "Vigilance–Cup".
[PNG File , 3566 KB - mhealth_v8i4e15344_app2.png]

Multimedia Appendix 3
The cognitive task “Titrated digit span backwards” is shown. (A) Explanation of the task and the two buttons for the selection of span length determination or exercise start; (B) Interface to set the personalized span length; (C) interface for the therapist in which he/she can find the heard series of digits to repeat backwards and the numeric keypad to type what the patient responded. An answer is correct if all the numbers are correctly repeated backward.
[PNG File , 4948 KB - mhealth_v8i4e15344_app3.png]

Multimedia Appendix 4
The cognitive task “Auditory vigilance with alphabets” is shown. (A) Explanation of the task and button to start the exercise; (B) Interface for the therapist in which the current given letter is displayed and the button “YES” is present to allow the therapist to record when the patient says "yes." An incorrect answer is automatically counted as false positive (ie, a "yes" when there is no target letter) or false negative (ie, answer omission when there is the target letter). A correct answer is automatically counted in other cases (ie, a "yes" when there is the target letter and omission when there is no target letter).
[PNG File , 3712 KB - mhealth_v8i4e15344_app4.png]

Multimedia Appendix 5
The cognitive task “Serial counting backwards by 7” is shown. (A) Explanation of the task and the button to start the exercise; (B) Interface for the therapist in which the starting number is displayed and the numeric keypad to type what the patient responded is present. A correct answer is automatically displayed in green, whereas an incorrect answer is displayed in red. If the patient makes an error, he/she is requested to start again from the last correct subtraction referred.
[PNG File , 3708 KB - mhealth_v8i4e15344_app5.png]

References

1. He Y, Yang L, Zhou J, Yao L, Pang MY. Dual-task training effects on motor and cognitive functional abilities in individuals with stroke: a systematic review. Clin Rehabil 2018 Jul;32(7):865-877. [doi: 10.1177/0269215518758482] [Medline: 29473480]
2. Bowen A, Wenman R, Mickelborough J, Foster J, Hill E, Tallis R. Dual-task effects of talking while walking on velocity and balance following a stroke. Age Ageing 2001 Jul;30(4):319-323. [doi: 10.1093/ageing/30.4.319] [Medline: 11509310]
3. Leone C, Feys P, Moumdjian L, D'Amico E, Zappia M, Patti F. Cognitive-motor dual-task interference: A systematic review of neural correlates. Neurosci Biobehav Rev 2017 Apr;75:348-360. [doi: 10.1016/j.neubiorev.2017.01.010] [Medline: 28104413]
4. Wang X, Pi Y, Chen B, Chen P, Liu Y, Wang R, et al. Cognitive motor interference for gait and balance in stroke: a systematic review and meta-analysis. Eur J Neurol 2015 Mar;22(3):555-e37 [FREE Full text] [doi: 10.1111/ene.12616] [Medline: 25560629]
5. Canning CG, Ada L, Woodhouse E. Multiple-task walking training in people with mild to moderate Parkinson's disease: a pilot study. Clin Rehabil 2008 Mar;22(3):226-233. [doi: 10.1177/0269215507082341] [Medline: 18285432]
6. Coelho FG, Andrade LP, Pedroso RV, Santos-Galduroz RF, Gobbi S, Costa JL, et al. Multimodal exercise intervention improves frontal cognitive functions and gait in Alzheimer's disease: a controlled trial. Geriatr Gerontol Int 2013 Jan;13(1):198-203. [doi: 10.1111/j.1447-0594.2012.00887.x] [Medline: 22686565]
7. Yogev G, Giladi N, Peretz C, Springer S, Simon ES, Hausdorff JM. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? Eur J Neurosci 2005 Sep;22(5):1248-1256. [doi: 10.1111/j.1460-9568.2005.04298.x] [Medline: 16176368]
8. Cocchini G, Della Sala S, Logie RH, Pagani R, Sacco L, Spinnler H. Dual task effects of walking when talking in Alzheimer's disease. Rev Neurol (Paris) 2004 Jan;160(1):74-80. [Medline: 14978396]
9. Liu Y, Yang Y, Tsai Y, Wang R. Cognitive and motor dual task gait training improve dual task gait performance after stroke - A randomized controlled pilot trial. Sci Rep 2017 Jun 22;7(1):4070 [FREE Full text] [doi: 10.1038/s41598-017-04165-y] [Medline: 28642466]
10. Geroin C, Nonnekes J, de Vries NM, Strouwen C, Smania N, Tinazzi M, et al. Does dual-task training improve spatiotemporal gait parameters in Parkinson's disease? Parkinsonism Relat Disord 2018 Oct;55:86-91. [doi: 10.1016/j.parkreldis.2018.05.018] [Medline: 29802080]
11. Fritz NE, Cheek FM, Nichols-Larsen DS. Motor-Cognitive Dual-Task Training in Persons With Neurologic Disorders: A Systematic Review. J Neurol Phys Ther 2015 Jul;39(3):142-153 [FREE Full text] [doi: 10.1097/NPT.0000000000000090] [Medline: 26079569]
12. Wajda DA, Sosnoff JJ. Cognitive-motor interference in multiple sclerosis: a systematic review of evidence, correlates, and consequences. Biomed Res Int 2015;2015:720856 [FREE Full text] [doi: 10.1155/2015/720856] [Medline: 25839039]
13. Leone C, Patti F, Feys P. Measuring the cost of cognitive-motor dual tasking during walking in multiple sclerosis. Mult Scler 2015 Feb;21(2):123-131. [doi: 10.1177/1352458514547408] [Medline: 25178543]

14. Porosińska A, Pierzchała K, Mentel M, Karpe J. Evaluation of postural balance control in patients with multiple sclerosis – effect of different sensory conditions and arithmetic task execution. A pilot study. Neurologia i Neurochirurgia Polska 2010;44(1):35-42. [doi: 10.1016/s0028-3843(14)60405-9] [Medline: 20358484]
15. Negahban H, Mofateh R, Arastoo AA, Mazaheri M, Yazdi MJ, Salavati M, et al. The effects of cognitive loading on balance control in patients with multiple sclerosis. Gait Posture 2011 Oct;34(4):479-484. [doi: 10.1016/j.gaitpost.2011.06.023] [Medline: 21802303]
16. Hamilton F, Rochester L, Paul L, Rafferty D, O'Leary CP, Evans JJ. Walking and talking: an investigation of cognitive-motor dual tasking in multiple sclerosis. Mult Scler 2009 Oct;15(10):1215-1227. [doi: 10.1177/1352458509106712] [Medline: 19667011]
17. Nilsagård Y, Denison E, Gunnarsson L, Boström K. Factors perceived as being related to accidental falls by persons with multiple sclerosis. Disabil Rehabil 2009;31(16):1301-1310. [doi: 10.1080/09638280802532639] [Medline: 19479575]
18. Veldkamp R, Romberg A, Hämäläinen P, Giffroy X, Moumdjian L, Leone C, et al. Test-Retest Reliability of Cognitive-Motor Interference Assessments in Walking With Various Task Complexities in Persons With Multiple Sclerosis. Neurorehabil Neural Repair 2019 Aug 26;33(8):623-634. [doi: 10.1177/1545968319856897] [Medline: 31240994]
19. Raats J, Lamers I, Baert I, Willekens B, Veldkamp R, Feys P. Cognitive-motor interference in persons with multiple sclerosis during five upper limb motor tasks with different complexity. Mult Scler 2019 Nov;25(13):1736-1745. [doi: 10.1177/1352458518808194] [Medline: 30417718]
20. Korakas N, Tsolaki M. Cognitive Impairment in Multiple Sclerosis: A Review of Neuropsychological Assessments. Cogn Behav Neurol 2016 Jun;29(2):55-67. [doi: 10.1097/WNN.0000000000000097] [Medline: 27336803]
21. Beer S, Khan F, Kesselring J. Rehabilitation interventions in multiple sclerosis: an overview. J Neurol 2012 Sep;259(9):1994-2008. [doi: 10.1007/s00415-012-6577-4] [Medline: 22772357]
22. Peruzzi A, Cereatti A, Della Croce U, Mirelman A. Effects of a virtual reality and treadmill training on gait of subjects with multiple sclerosis: a pilot study. Mult Scler Relat Disord 2016 Jan;5:91-96. [doi: 10.1016/j.msard.2015.11.002] [Medline: 26856951]
23. Sosnoff JJ, Wajda DA, Sandroff BM, Roeing KL, Sung J, Motl RW. Dual task training in persons with Multiple Sclerosis: a feasability randomized controlled trial. Clin Rehabil 2017 Oct;31(10):1322-1331. [doi: 10.1177/0269215517698028] [Medline: 28933609]
24. Monjezi S, Negahban H, Tajali S, Yadollahpour N, Majdinasab N. Effects of dual-task balance training on postural performance in patients with Multiple Sclerosis: a double-blind, randomized controlled pilot trial. Clin Rehabil 2017 Feb;31(2):234-241. [doi: 10.1177/0269215516639735] [Medline: 27006419]
25. Motl RW, Sandroff BM, DeLuca J. Exercise Training and Cognitive Rehabilitation: A Symbiotic Approach for Rehabilitating Walking and Cognitive Functions in Multiple Sclerosis? Neurorehabil Neural Repair 2016 Jul;30(6):499-511. [doi: 10.1177/1545968315606993] [Medline: 27261483]
26. Evans JJ, Greenfield E, Wilson BA, Bateman A. Walking and talking therapy: improving cognitive-motor dual-tasking in neurological illness. J Int Neuropsychol Soc 2009 Jan;15(1):112-120. [doi: 10.1017/S1355617708090152] [Medline: 19128534]
27. Trenova AG, Slavov GS, Manova MG, Aksentieva JB, Miteva LD, Stanilova SA. Cognitive Impairment in Multiple Sclerosis. Folia Med (Plovdiv) 2016 Sep 01;58(3):157-163. [doi: 10.1515/folmed-2016-0029] [Medline: 27760012]
28. Lowrey CR, Watson A, Vallis LA. Age-related changes in avoidance strategies when negotiating single and multiple obstacles. Exp Brain Res 2007 Sep;182(3):289-299. [doi: 10.1007/s00221-007-0986-0] [Medline: 17551718]
29. Curuk E, Goyal N, Aruin AS. The Effect of Motor and Cognitive Tasks on Gait in People with Stroke. J Stroke Cerebrovasc Dis 2019 Nov;28(11):104330. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104330] [Medline: 31439522]
30. Monticone M, Ambrosini E, Fiorentini R, Rocca B, Liquori V, Pedrocchi A, et al. Reliability of spatial-temporal gait parameters during dual-task interference in people with multiple sclerosis. A cross-sectional study. Gait Posture 2014 Sep;40(4):715-718. [doi: 10.1016/j.gaitpost.2014.06.015] [Medline: 25086800]
31. Yang L, He C, Pang MY. Reliability and Validity of Dual-Task Mobility Assessments in People with Chronic Stroke. PLoS One 2016;11(1):e0147833 [FREE Full text] [doi: 10.1371/journal.pone.0147833] [Medline: 26808662]
32. Kalron A, Dvir Z, Achiron A. Walking while talking--difficulties incurred during the initial stages of multiple sclerosis disease process. Gait Posture 2010 Jul;32(3):332-335. [doi: 10.1016/j.gaitpost.2010.06.002] [Medline: 20594850]
33. Sosnoff JJ, Boes MK, Sandroff BM, Socie MJ, Pula JH, Motl RW. Walking and thinking in persons with multiple sclerosis who vary in disability. Arch Phys Med Rehabil 2011 Dec;92(12):2028-2033. [doi: 10.1016/j.apmr.2011.07.004] [Medline: 22133253]
34. Sosnoff JJ, Socie MJ, Sandroff BM, Balantrapu S, Suh Y, Pula JH, et al. Mobility and cognitive correlates of dual task cost of walking in persons with multiple sclerosis. Disabil Rehabil 2014;36(3):205-209. [doi: 10.3109/09638288.2013.782361] [Medline: 23597000]
35. Nogueira LA, Santos LT, Sabino PG, Alvarenga RM, Thuler LC. Walking execution is not affected by divided attention in patients with multiple sclerosis with no disability, but there is a motor planning impairment. Arq Neuropsiquiatr 2013 Aug;71(8):521-526 [FREE Full text] [doi: 10.1590/0004-282X20130080] [Medline: 23982009]
36. Wechsler D. WAIS-III Administration and Scoring Manual. San Antonio, TX: Psychological Corporation; 1997.

37. Veldkamp R, Baert I, Kalron A, Tacchino A, D'hooge M, Vanzeir E, et al. Structured Cognitive-Motor Dual Task Training Compared to Single Mobility Training in Persons with Multiple Sclerosis, a Multicenter RCT. J Clin Med 2019 Dec 10;8(12) [FREE Full text] [doi: 10.3390/jcm8122177] [Medline: 31835502]
38. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011 Feb;69(2):292-302 [FREE Full text] [doi: 10.1002/ana.22366] [Medline: 21387374]
39. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
40. Beatty WW, Goodkin DE. Screening for cognitive impairment in multiple sclerosis. An evaluation of the Mini-Mental State Examination. Arch Neurol 1990 Mar;47(3):297-301. [doi: 10.1001/archneur.1990.00530030069018] [Medline: 2310313]
41. Rickham PP. Human Experimentation: Code of Ethics of the World Medical Association (Declaration of Helsinki). Can Med Assoc J 1964 Sep 12;91(11):619 [FREE Full text] [Medline: 20327943]
42. Silsupadol P, Lugade V, Shumway-Cook A, van Donkelaar P, Chou L, Mayr U, et al. Training-related changes in dual-task walking performance of elderly persons with balance impairment: a double-blind, randomized controlled trial. Gait Posture 2009 Jun;29(4):634-639 [FREE Full text] [doi: 10.1016/j.gaitpost.2009.01.006] [Medline: 19201610]
43. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
44. Ryan RM. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology 1982;43(3):450-461. [doi: 10.1037/0022-3514.43.3.450]
45. Tacchino A, Pedullà L, Bonzano L, Vassallo C, Battaglia MA, Mancardi G, et al. A New App for At-Home Cognitive Training: Description and Pilot Testing on Patients with Multiple Sclerosis. JMIR Mhealth Uhealth 2015 Aug 31;3(3):e85 [FREE Full text] [doi: 10.2196/mhealth.4269] [Medline: 26323749]
46. Pedullà L, Brichetto G, Tacchino A, Vassallo C, Zaratin P, Battaglia MA, et al. Adaptive vs. non-adaptive cognitive training by means of a personalized App: a randomized trial in people with multiple sclerosis. J Neuroeng Rehabil 2016 Oct 04;13(1):88 [FREE Full text] [doi: 10.1186/s12984-016-0193-y] [Medline: 27716336]
47. Noyes JM, Garland KJ. Computer- vs. paper-based tasks: are they equivalent? Ergonomics 2008 Sep;51(9):1352-1375. [doi: 10.1080/00140130802170387] [Medline: 18802819]
48. Mattioli F, Flavia M, Stampatori C, Zanotti D, Parrinello G, Capra R. Efficacy and specificity of intensive cognitive rehabilitation of attention and executive functions in multiple sclerosis. J Neurol Sci 2010 Jan 15;288(1-2):101-105. [doi: 10.1016/j.jns.2009.09.024] [Medline: 19825502]
49. Shatil E, Metzer A, Horvitz O, Miller A. Home-based personalized cognitive training in MS patients: a study of adherence and cognitive performance. NeuroRehabilitation 2010;26(2):143-153. [doi: 10.3233/NRE-2010-0546] [Medline: 20203380]
50. Gich J, Freixanet J, García R, Vilanova JC, Genís D, Silva Y, et al. A randomized, controlled, single-blind, 6-month pilot study to evaluate the efficacy of MS-Line!: a cognitive rehabilitation programme for patients with multiple sclerosis. Mult Scler 2015 Sep;21(10):1332-1343. [doi: 10.1177/1352458515572405] [Medline: 25716880]
51. Bonavita S, Sacco R, Della Corte M, Esposito S, Sparaco M, d'Ambrosio A, et al. Computer-aided cognitive rehabilitation improves cognitive performances and induces brain functional connectivity changes in relapsing remitting multiple sclerosis patients: an exploratory study. J Neurol 2015 Jan;262(1):91-100. [doi: 10.1007/s00415-014-7528-z] [Medline: 25308631]
52. Wollesen B, Voelcker-Rehage C, Regenbrecht T, Mattes K. Influence of a visual-verbal Stroop test on standing and walking performance of older adults. Neuroscience 2016 Mar 24;318:166-177. [doi: 10.1016/j.neuroscience.2016.01.031] [Medline: 26808774]
53. Paglialonga A, Lugo A, Santoro E. An overview on the emerging area of identification, characterization, and assessment of health apps. J Biomed Inform 2018 Jul;83:97-102 [FREE Full text] [doi: 10.1016/j.jbi.2018.05.017] [Medline: 29852318]
54. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
55. Attkisson CC, Zwick R. The client satisfaction questionnaire. Evaluation and Program Planning 1982 Jan;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
56. Motl RW, Balto JM, Ensari I, Hubbard EA. Self-efficacy and Walking Performance in Persons With Multiple Sclerosis. J Neurol Phys Ther 2017 Apr;41(2):114-118. [doi: 10.1097/NPT.0000000000000172] [Medline: 28263255]

Abbreviations

CMI: cognitive-motor interference
DT: dual task
DTC: dual-task cost
DTT: dual-task training
EDSS: Expanded Disability Status Scale
GUI: graphical user interface
IMI: Intrinsic Motivation Inventory

MMSE: Mini-Mental State Examination
MS: multiple sclerosis
RPE: Borg 15-point Ratings of Perceived Exertion

Edited by G Eysenbach; submitted 09.07.19; peer-reviewed by M Pau, T Raffegeau, J Sosnoff, F Lanfranchi, E Sadeghi-Demneh; comments to author 29.09.19; revised version received 22.01.20; accepted 26.02.20; published 16.04.20.

Please cite as:
Tacchino A, Veldkamp R, Coninx K, Brulmans J, Palmaers S, Hämäläinen P, D'hooge M, Vanzeir E, Kalron A, Brichetto G, Feys P, Baert I
Design, Development, and Testing of an App for Dual-Task Assessment and Training Regarding Cognitive-Motor Interference (CMI-APP) in People With Multiple Sclerosis: Multicenter Pilot Study
JMIR Mhealth Uhealth 2020;8(4):e15344
URL: http://mhealth.jmir.org/2020/4/e15344/
doi:10.2196/15344
PMID:32343258

Original Paper

Exploring Patients' Intentions for Continuous Usage of mHealth Services: Elaboration-Likelihood Perspective Study

Xitong Guo[1,2], PhD; Shuqing Chen[2], MSc; Xiaofei Zhang[3], PhD; Xiaofeng Ju[2], PhD; Xifu Wang[1], MSc

[1]Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, China

[2]School of Management, Harbin Institute of Technology, Harbin, China

[3]Business School, Nankai University, Tianjin, China

Corresponding Author:
Shuqing Chen, MSc
School of Management
Harbin Institute of Technology
No 2, Yikuang Str, Nangang District
Harbin, 150001
China
Phone: 86 451 8641 4022
Email: chenshuqinghit@gmail.com

Abstract

Background: With the increasingly rapid development of Web 2.0 technologies, the application of mobile health (mHealth) care in the field of health management has become popular. Accordingly, patients are able to access consulting services and effective health information online without temporal and geographical constraints. The elaboration-likelihood model (ELM) is a dual-process persuasion theory that describes the change of attitudes and behavior.

Objective: In this study, we drew on the ELM to investigate patients' continuous usage intentions regarding mHealth services. In addition, we further examined which route—central or peripheral—has a stronger impact on a patient's usage of health care management.

Methods: To meet these objectives, five hypotheses were developed and empirically validated using a field survey to test the direct and indirect effects, via attitude, of the two routes on continuous usage intention.

Results: We found that patients' perceived mHealth information quality and perceived mHealth system quality had a positive effect on their personal attitudes. The results revealed that social media influence had a positive effect on a patient's attitude toward mHealth services. In particular, our findings suggest that a patient's health consciousness has a positive effect on the relationship between social media influence and attitude.

Conclusions: This study contributes to the mHealth services literature by introducing the ELM as a referent theory for research, as well as by specifying the moderating role of health consciousness. For practitioners, this study introduces influence processes as policy tools that managers can employ to motivate the uptake of mHealth services within their organizations.

(JMIR Mhealth Uhealth 2020;8(4):e17258) doi:10.2196/17258

KEYWORDS

mHealth services; health consciousness; elaboration-likelihood model; health behavior; patients' continuous usage

Introduction

Background

Advances in Web 2.0 technology have resulted in the application of mobile health (mHealth) care to the field of health management with increasing popularity. According to the Mobile Medical Survey Report of 2016, there were about 90,000 mHealth apps on the iOS platform in the United States in 2015. Compared with that of 2013, the growth rate was 106%. In October 2015, the China State Council implemented the Internet Plus policy, which enhanced the status of internet technology in health care. In October 2016, the outline of the Healthy China 2020 plan was proposed to establish a healthy concept to realize the health care needs of the entire nation. In January 2017, the Chinese government proposed the 13th Five-Year Plan for Health Promotion and Work and established the activities and plans for health promotion. In February 2017, the Chinese government proposed a long-term plan for the prevention and

treatment of chronic diseases in China, which fully expressed the concept of the management of chronic diseases in daily lives. Studies have shown that health management apps have a positive impact on the improvement of personal health [1,2]. However, realistically, in all health management apps, the actual usage rate is rather low and the effects of health management are not obvious [3]. In our research, we ask the following questions: What are the reasons for the low utilization of health management apps? What factors affect users' low compliance with mHealth apps and health behavior changes?

Prior literature suggests that the availability and ease of the use of health management apps play a crucial role in a patient's choice of whether or not to use an app [4]. In 2009, Sykes et al [5] indicated that the professional support provided by a health management app to patients exerts a positive impact on the use of the app. Another study found that the patient's personal intrinsic motivation has a positive effect on his or her health behavior changes [6]. Chen also explored the attitude toward organic food among the Taiwanese, which is related to health consciousness, environmental attitudes, and the mediating effects of a healthy lifestyle [7]. In fact, he found that health consciousness and the environment are the two most important motives for purchasing organic food. In addition, healthy lifestyle acts as a positive mediating effect between health consciousness and environmental attitudes, as well as the consumer's attitudes toward organic foods [7]. Most of the previous studies have shown that system and information service quality, one's motivation and ability, as well as social influence could affect the usage of apps. However, there is no clear evidence on the types of information that are most effective in influencing a patient's continuous usage intention and whether these influences are temporally persistent.

Bhattacherjee and Sanford found that the influence processes for information technology (IT) acceptance are based on an elaboration-likelihood model (ELM) [8]. This theory enhanced our understanding of influence processes in IT acceptance based on the ELM. Indeed, the ELM includes two types of routes: central and peripheral routes are based on the type of information processed by a given user. The ELM also explains how one route may cause more influence than the other and how each route can have a long-term impact on users' behavior changes [8,9]. Therefore, in this study, we conducted in-depth research on the central route (ie, intrinsic motivation) and peripheral route (ie, external motivation) of the ELM as well as on health awareness through multifactor perspectives at the individual, organizational, and social levels. We also determined which route—the central or the peripheral route—has a stronger impact on a patient's use of health management apps. In particular, we planned to answer to the following research questions:

1. Question 1: What are the factors that affect a patient's continuous usage intention of mHealth services?
2. Question 2: How does health consciousness affect a patient's continuous usage intention of mHealth services?
3. Question 3: Which factor has a stronger effect on a patient's continuous usage intention of mHealth services?

Seeking the answers to these three questions can enable us to better understand the factors that affect patients' continuous usage of mHealth.

To meet the aforementioned objectives, a survey was conducted to test the main and mediated effects of mHealth services on a patient's continuous usage intention. Our participants were users of mHealth apps in China. We tested the proposed model using structural equation modeling (SEM) with a partial least squares (PLS) estimation.

This study provides several theoretical contributions. First, we expanded the ELM and applied it to the health sector. Second, we explored the central and peripheral routes of the ELM, each with a different motivation and ability regarding the effects of patients' continuous usage intention. Third, we redefined the cognitive-attitude-change process based on the original ELM.

Moreover, this study also contributes to the practical understanding of mHealth services in several important ways. First, as a result of the verification of our model, hospital administrators are provided with a theoretical basis, which enables them to better understand a patient's psychological characteristics and accordingly construct a more comprehensive informatization process. Second, our model can provide support for health management entrepreneurs in terms of the design of health management apps so that they can clearly identify the problems and directions facing them. Third, the patients themselves can also gain a clear understanding of their inner tendencies, so that they can identify health management systems that best suit their personal characteristics through many health management apps and, thus, better manage their own health problems.

The rest of this paper proceeds as follows. In the next section (Theoretical Background), we present a brief overview of prior research and identify gaps in health behavior through mHealth service management. In the following section (Research Model and Hypotheses), we describe the key constructs and relationships in the ELM and present five hypotheses. The proposed hypotheses are tested using a survey based on a dataset of the app users of the mHealth service. This is followed by a discussion of our findings and their implications for information service research and practice. We end with some conclusions.

Theoretical Background

mHealth Services

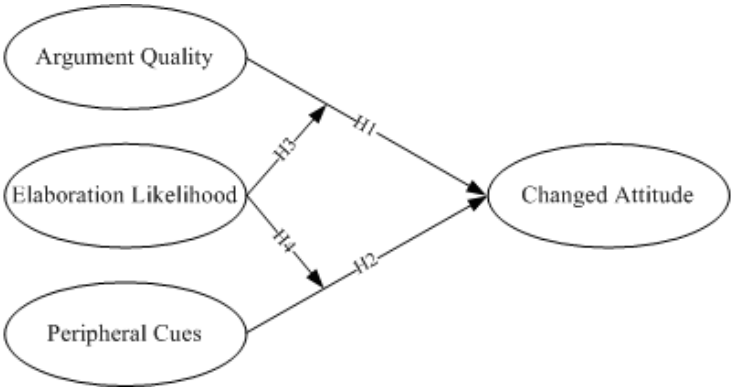
mHealth services, as a new health management approach to break time and space constraints, can employ communication technology and mobile devices to provide personalized health management, real-time information tracking, and other services for patients [10]. They can also provide services for all patients without temporal and geographical constraints. Compared with traditional health management approaches, mHealth services enable users to access unlimited data and more effective disease control approaches. With an inflow of medical capital, policy support, and technology development, mHealth services have undergone rapid development. According to the Mobile Medical Survey Report of 2017 [11], the stickiness of mHealth app users has generally been seen as being on the rise since 2016, and mHealth services are gradually becoming one of the ways for

patients to access health management. Because of their mobility, traceability, and high levels of interaction, mHealth apps provide solutions for effective health management. Thus, mHealth apps alleviate the health care pressures and economic burdens of patients; improve the well-being of individuals, families, and society; and achieve the maximization of medical values. On the one hand, mHealth services provide effective information for patients, which could enable them to perform effective health self-management without the constraints of time and location. On the other hand, mHealth services provide effective communication channels for patients, family members, and doctors and thereby enable patients to receive emotional and professional support from family members and doctors.

The Elaboration-Likelihood Model

The ELM was developed by Petty and Cacioppo in 1986 [9]. It is a dual-process persuasion theory that describes the change of attitudes. The ELM mainly includes the central route and the peripheral route to persuasion. The central route to persuasion—where elaboration is high—is probably due to the careful and thoughtful consideration of the true value of the information provided by a person to support advocacy. The central route involves a high level of message elaboration, where the individual receiving the message generates a large amount of cognitions about the arguments. The result of changing attitudes will be relatively long-lasting, resistant, and predictive of behavior. In contrast, the peripheral route to persuasion—where elaboration is low—is caused by a person's connection with positive or negative clues in the stimulus or by simple reasoning about the merits of the claimed position. The prompts that an individual receives in a peripheral route are usually independent of the logical quality of the stimulus. These tips will relate to factors such as the credibility or appeal of the source or the quality of the message. The likelihood of elaboration will be determined by an individual's motivation and ability to evaluate the argument being presented. The central and peripheral routes of attitude change are typically operationalized in ELM research using the argument quality and peripheral cues constructs, respectively, as shown in Figure 1.

Figure 1. Original model for the elaboration-likelihood model. H1: Hypothesis 1; H2: Hypothesis 2; H3: Hypothesis 3; H4: Hypothesis 4.



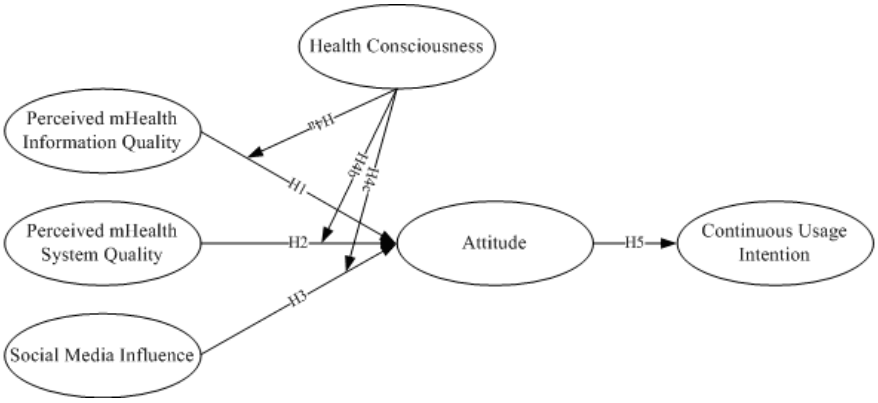
Prior scholars have widely implied that the ELM applies in e-commerce, advertising, and other fields. Cheung et al examined four information cues to evaluate the credibility of online reviews drawing on the ELM [12]. Ho and Bodoff also integrated the ELM and consumer search theory to illustrate how depth and breadth influence a user's attitude toward a personalization agent and item selection [13]. In addition, Cao et al investigated patients' selection decisions based on the ELM and the service quality theory; they also examined the moderating effects of disease risk and disease knowledge on patients' consulting intention [14]. Although the above studies demonstrated the effectiveness of the ELM in user behavior, user purchase, and patients' consulting intention in the online health community, to the best of our knowledge, there is no study involving patients' continuous usage intention toward mHealth services.

Research Model and Hypotheses

Overview

Based on the aforementioned theoretical foundations, our research model is proposed as shown in Figure 2. In the model, the central routes (ie, perceived mobile information quality and perceived mobile system quality) and the peripheral route (ie, social media influences) are considered as antecedents of the model; to these, we add the moderating effect (ie, health consciousness) on the attitude, which in turn influences a patient's continuous usage intention.

Figure 2. The conceptual research model based on the elaboration-likelihood model. H1: Hypothesis 1; H2: Hypothesis 2; H3: Hypothesis 3; H4a: Hypothesis 4a; H4b: Hypothesis 4b; H4c: Hypothesis 4c; H5: Hypothesis 5.



Perceived mHealth Information and System Quality Based on a Patient's Attitudes

Previous studies have shown that the quality of e-services has a positive effect on customers' attitudes [15]. Wixom and Todd studied the impact of service quality on user satisfaction and divided service quality into information quality and system quality in the e-service context [16]. Cao et al revealed that service quality and e-word-of-mouth both had positive effects on patients' selection decisions [14]. Similar to e-service, mHealth service, as a new service platform and channel, has attracted considerable attention. Information quality and design standards in mHealth service apps can have an impact on patients' perceptions.

Thus, we hypothesize the following:

1. Hypothesis 1: Patients perceive that the information quality of an mHealth service has a positive effect on their personal attitudes.
2. Hypothesis 2: Patients perceive that the quality of the systems of an mHealth service has a positive effect on their personal attitudes.

Social Media Influence on Patients' Attitudes

In addition to focusing on app-provided information, patients will also take into consideration external information, which influences their decision making. The prior study proposed that entertainment, sociality, information, and trust positively influence WeChat users' attitudes and users' trust [17]. Treviño et al put forward that medium symbolism, message equivocality, distance between message partners, perceived media richness, number of message recipients, and perceived message recipients' attitudes are the factors that affect the user's media choices [18]. Erkan and Evans also found that the key factors of social media that influence consumers' purchase intentions were quality, credibility, usefulness and adoption of information, need for information, and attitude toward information [19]. Therefore, they take into consideration the peripheral route, which includes the persuasive effects of information conveyed by social media on patients.

Thus, we hypothesize the following: Social media influence has a positive impact on a patient's attitude (Hypothesis 3).

The Moderation Effects of Health Consciousness

In addition to the effects of the central and the peripheral routes, an individual's health consciousness is related to his or her ability to receive information. Studies have shown that the higher the level of a patient's health consciousness, the more concerned that patient would be about his or her health [20]. Scholars have shown that people with high health-literacy rates have quick access to information and strong judgment [21]. In addition, a previous study has illustrated that social cognitive factors and all the perceived social influence variables significantly improved eHealth literacy [22]. Hence, we propose that personal health consciousness plays a moderating role in the ELM.

Thus, we hypothesize the following:

1. Hypothesis 4a: A patient's health consciousness has a positive moderating effect on his or her relationship between the perceived information quality and the contents and attitudes of mHealth services.
2. Hypothesis 4b: A patient's health consciousness has a positive moderating effect on the relationship between the perceived service quality and the systems and attitudes of mHealth services.
3. Hypothesis 4c: A patient's health consciousness has a negative moderating effect on the relationship between social media influence and attitudes.

The Attitudes and Continuous Usage Intention of Patients

Prior studies have inferred that a patient's attitude has a positive effect on his or her usage intention. Thus, we hypothesize the following: A patient's attitude has a positive effect on his or her continuous usage intention (Hypothesis 5). The descriptions and definitions of the constructs of our hypotheses are presented in Table 1.

Table 1. The descriptions and definitions of the constructs.

Construct	Description
Perceived mHealth information quality	A patient's perception of the quality of information provided by the mHealth management apps
Perceived mHealth system quality	A patient's perception of the quality of the system provided by the mHealth management apps
Social media influence	The persuasive effects of the information conveyed by social media on a patient's attitude to the mHealth management apps: this is not a component of the app itself
Health consciousness	A patient's attention to his or her personal health issues or the motivation to protect them
Attitude	A patient's attitude toward the mHealth management apps
Continuous usage intention	The willingness of patients to continue to use their health management apps

Methods

Overview

To test our model, we conducted a survey. This method was chosen because it is a way to access perceived information quality, perceived system quality, social media influence, attitude, and continuous usage intention; as well, it can enhance the broader applications of the research findings [23].

Sample of the Normal Distribution Test

Bentler and Chou stressed that data can be structured for analysis on the premise that the sample dataset satisfies the normal distribution [24]. Although the results of our study indicate that the software program SmartPLS (SmartPLS GmbH) is suitable for nonnormal distribution, the results are better under conditions of a normal distribution. The condition by which the sample data satisfies a normal distribution occurs after the descriptive test of each constructed measurement item index, when the absolute value of the skewness is less than 2, and while the absolute value of the kurtosis is less than 5.0. We used SPSS Statistics for Windows, version 23.0 (IBM Corp), for statistical analysis of our sample data. The main output indicators were measured (ie, mean, SD, minimum, maximum, skewness, and kurtosis). The specific results are shown in Table 2.

We can perceive from this table that the largest value of skewness of 1.433 is less than 2, while the largest value of kurtosis of 2.269 is less than 5. Therefore, our survey data basically conforms to the above normal distribution and can be used for further structural equation analysis.

Table 2. Sample of the normal distribution test.

Item	Sample size	Min[a]	Max[b]	Mean (SD)	Skewness	Kurtosis
PIQ[c]1	255	1	7	5.52 (0.917)	-0.759	1.723
PIQ2	255	1	7	5.42 (1.123)	-0.589	0.457
PIQ3	255	1	7	5.54 (1.086)	-0.777	1.203
PIQ4	255	1	7	5.57 (1.154)	-0.958	1.335
PIQ5	255	2	7	5.13 (1.077)	-0.527	0.399
PSQ[d]1	255	2	7	5.39 (1.013)	-0.586	0.442
PSQ2	255	1	7	5.35 (1.151)	-0.644	0.483
PSQ3	255	1	7	5.37 (1.082)	-0.822	1.239
PSQ4	255	2	7	5.44 (1.138)	-0.697	0.396
PSQ5	255	1	7	5.06 (1.247)	-0.402	0.071
SMI[e]1	255	1	7	4.80 (1.469)	-0.663	-0.031
SMI2	255	1	7	4.85 (1.466)	-0.533	-0.125
SMI3	255	1	7	4.92 (1.456)	-0.642	-0.137
SMI4	255	1	7	4.93 (1.444)	-0.626	-0.207
HC[f]1	255	2	7	5.45 (1.244)	-0.953	0.612
HC2	255	3	7	5.73 (1.170)	-0.671	-0.432
HC3	255	1	7	5.83 (1.024)	-0.956	1.492
HC4	255	1	7	5.00 (1.290)	-0.639	0.325
HC5	255	2	7	5.79 (1.126)	-1.051	0.955
HC6	255	2	7	5.92 (1.145)	-0.947	0.475
AT[g]1	255	2	7	5.54 (1.071)	-0.612	0.244
AT2	255	2	7	5.46 (1.064)	-0.683	0.552
AT3	255	1	7	5.14 (1.289)	-0.692	0.215
AT4	255	1	7	5.35 (1.213)	-0.949	1.191
AT5	255	2	7	5.56 (1.175)	-0.832	0.617
CIU[h]1	255	1	7	5.87 (1.267)	-1.364	1.996
CIU2	255	1	7	5.69 (1.367)	-1.169	0.982
CIU3	255	1	7	5.76 (1.307)	-1.433	2.269
CIU4	255	1	7	5.71 (1.305)	-1.302	1.868

[a]Min: minimum.

[b]Max: maximum.

[c]PIQ: perceived information quality.

[d]PSQ: perceived system quality.

[e]SMI: social media influence.

[f]HC: health consciousness.

[g]AT: attitude.

[h]CIU: continuous usage intention.

Study Setting and Demographic Details of Participants

The survey questionnaire was distributed to the participants via WJX [25], a questionnaire website that is the oldest and most-used online survey software and serves as an examination and voting platform in China. Since its launch in 2006, users have posted more than 28.22 million questionnaires and have collected more than 1.88 billion responses while maintaining a growth rate of more than 100% yearly. The platform's users have covered more than 90% of universities and research institutions in China, and the platform is a well-known portal

for questionnaires, examinations, and voting systems that are trusted by leading companies in various industries.

The survey was posted on the questionnaire website. The objectives of the survey and the requirements for the recruitment of our participants were introduced at the beginning of the survey, and each participant would receive an incentive of RMB 15 (approximately US $2.19). The participants were from all over China and were users of mHealth management apps, such as Haodf, XYWY, or Guahao, which are large-scale mHealth apps in China. The data collection process was divided into two phases. In the first phase, a pilot analysis was conducted and the measurement model (eg, the reliability, validity, common method bias, and multicollinearity of the constructs) was assessed to ensure its appropriateness. In the second phase, we collected the data. A total of 262 questionnaires were distributed. Of these, 255 contained acceptable responses, yielding a response rate of 97.3%. Of these respondents, 55.3% (141/255) were male and 79.6% (203/255) were under the age of 40 years. Most of the participants were employed (241/255, 94.5%). Among the participants, 67.1% (171/255) had graduated with a bachelor's degree.

The demographics of the respondents are shown in Table 3. A 7-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), was used in our research. The scores indicate different levels of health consciousness for the two groups in our sample and we believe this approach was appropriate for our research.

Table 3. Demographics of the respondents.

Measure and category	Value (N=255), n (%)
Age in years	
<40	203 (79.6)
40-50	45 (17.6)
51-60	6 (2.4)
>60	1 (0.4)
Gender	
Female	114 (44.7)
Male	141 (55.3)
Education	
Junior middle school or below	1 (0.4)
High school	4 (1.6)
Special secondary school	14 (5.5)
Junior college	38 (14.9)
Bachelor's degree	171 (67.1)
Master's degree or above	27 (10.6)
Work status	
Working	241 (94.5)
Retired	14 (5.5)

Measurements

The measurement of the majority of constructs was adopted from prior relevant studies (see Multimedia Appendix 1). Slight modifications were necessary to make the text suitable for the research context, and all measures used a 7-point Likert scale. Sustained participation was measured using the instrument of continuous usage intention that was adopted from Bhattacherjee and Sanford [8] and from Petty and Cacioppo [9]. Perceived mHealth information quality was measured using methods by Park et al [26] and by Bhattacherjee and Sanford [8]. Perceived mHealth system quality was measured using methods by Cheong and Park [27] and by Brady and Cronin [28]. Social media influence was measured using methods by Keery et al [29]. Attitude was measured using methods by Kim [30] and by Yang and Yoo [31]. The six items used to measure health consciousness were adopted from Mai and Eisenberg [32] and from Chen [7].

To enhance the validity of the measures of the constructs, we followed Moore and Benbasat [33] when adapting the measurements for our study. All the measures were adapted from prior research along with their sources (see Multimedia Appendix 1). Other control variables, such as age, gender, education, and work experience, were measured using a single-item measure. To verify the adapted survey items, individual meetings were held with university colleagues and postgraduate students to discuss the following: (1) the appropriateness of the questionnaire items, (2) the possibility of ambiguity in the questionnaire items, and (3) the appearance and layout of the survey instrument. Based on the feedback received, a revised questionnaire was developed. This was then

sent to the same individuals for a second review. A minor revision was made based on their further suggestions.

Data Analysis and Results

Analysis Strategy

We tested the hypothesized relationships among the constructs using SEM with the software program SmartPLS 3.0 (SmartPLS GmbH) [12,34]. A two-stage analytical procedure was used to analyze the data [35]. First, we assessed the validity of the measurement model and then examined its structure. PLS is a powerful component-based method that has been widely used in prior research [36]. It does not require multivariate normal distribution and has a minimal sample size requirement as compared to other SEM packages (eg, linear structural relationships [LISREL], analysis of moment structure [AMOS], and equation structural program [EQS]) [34,37,38]. In addition, it simultaneously estimates the structural and measurement models [36]. As our sample size is relatively small, we chose PLS to analyze the data [34]. Specifically, SmartPLS was used to conduct the data analysis [39].

Measurement Model

As essential prerequisites for achieving valid results, the reliability, convergent validity, and discriminant validity of the measurement model were assessed. Reliability was assessed by applying the Cronbach alpha and composite test for reliability. As shown in Table 4, Cronbach alpha values range from .718 to .881 and the composite reliability values range from .797 to .918, both of which exceed the recommended value of .70, thus confirming their reliability.

Table 5 shows that most of the item loadings were above .70, thus indicating convergent validity [40]. Moreover, the factor loadings of each construct were much greater than the cross-loadings on other constructs, and correlations of the constructs were much smaller than the square root of the *average variance extracted* of each construct, thus indicating discriminant validity [41], as shown as Table 6.

Table 4. The results of confirmatory factor analysis: construct reliability and validity (N=255).

Indicator (abbreviation)	Cronbach alpha	rho_A	Composite reliability	Average variance extracted
Attitude (AT)	.774	.777	.855	.797
Continuous usage intention (CUI)	.881	.881	.918	.737
Health consciousness (HC)	.795	.701	.811	.717
Perceived information quality (PIQ)	.729	.734	.830	.751
Perceived system quality (PSQ)	.718	.720	.797	.767
Social media influence (SMI)	.866	.868	.909	.714

Table 5. Item loadings and cross-loadings for each construct (N=255).

Construct (abbreviation) and items	Item loadings and cross-loadings					
	AT	CUI	HC	PIQ	PSQ	SMI
Attitude (AT)						
AT1	.789	.471	.394	.402	.290	.407
AT3	.711	.387	.363	.309	.347	.299
AT4	.791	.453	.306	.350	.386	.336
AT5	.796	.427	.311	.330	.333	.274
Continuous usage intention (CUI)						
CUI1	.460	.862	.295	.334	.222	.290
CUI2	.504	.872	.285	.352	.248	.198
CUI3	.491	.860	.271	.391	.220	.244
CUI4	.481	.839	.232	.350	.228	.311
Health consciousness (HC)						
HC2	.298	.277	.709	.223	.148	.141
HC3	.255	.172	.723	.159	.129	.094
HC4	.398	.203	.736	.194	.241	.189
HC5	.299	.257	.709	.206	.120	.181
Perceived information quality (PIQ)						
PIQ1	.364	.331	.253	.777	.406	.200
PIQ2	.335	.327	.220	.716	.464	.218
PIQ3	.279	.237	.114	.714	.442	.166
PIQ4	.356	.328	.205	.760	.428	.223
Perceived system quality (PSQ)						
PSQ1	.302	.210	.180	.437	.741	.201
PSQ2	.346	.257	.207	.483	.771	.167
PSQ5	.337	.136	.137	.398	.746	.286
Social media influence (SMI)						
SMI1	.350	.263	.146	.261	.275	.857
SMI2	.342	.199	.179	.107	.227	.829
SMI3	.373	.296	.185	.251	.203	.825
SMI4	.385	.263	.219	.296	.273	.868

Table 6. Correlation matrix (N=255).

Construct (abbreviation)	Correlation					
	AT	CUI	HC	PIQ	PSQ	SMI
Attitude (AT)	.772					
Continuous usage intention (CIU)	.564	.858				
Health consciousness (HC)	.446	.315	.719			
Perceived information quality (PIQ)	.453	.416	.273	.742		
Perceived system quality (PSQ)	.438	.268	.232	.584	.753	
Social media influence (SMI)	.430	.303	.217	.274	.290	.845

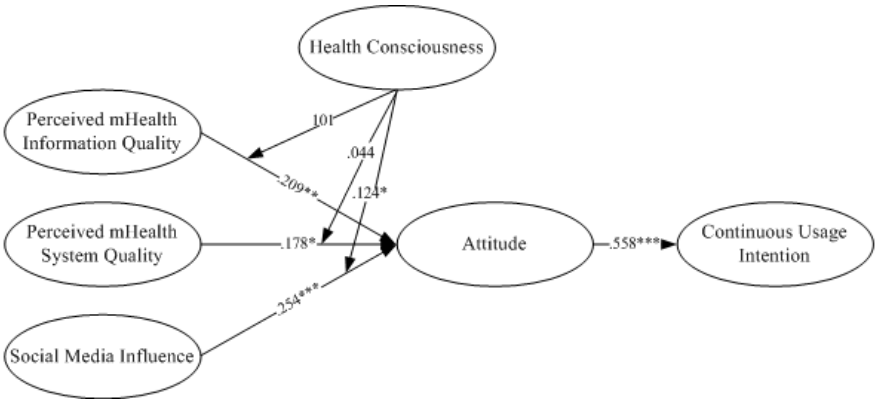
Common-Method Bias Testing

As our data were collected from self-report surveys from the participants, common-method bias may threaten the validity of the results [42]. We used a modified-marker variable analysis to test the common-method bias in our model following Rönkkö and Ylitalo [43]; three items in our dataset, which have low correlation with the items in this study, were used to measure the marker variable. Next, the marker variable was incorporated into the model with its impacts on the endogenous variables. The results showed that the marker variable had no significant impacts on perceived mHealth information quality, perceived mHealth system quality, social media influence, attitude, and continuous usage intention; as well, hypothesized relationships had no significant differences regardless of whether or not the marker variable was introduced into the model. The results indicate that common-method bias had little impact on the results of our study [43].

Structural Model

An analysis of the structural model formed the second stage in the SEM. The significance of each path coefficient was calculated by bootstrapping with 5000 samples using the replacement method. In addition, SmartPLS supports the function of testing, whether the moderator of a relationship is significant or not in the model. We were able to choose this factor as a moderator in the tool directly and the test of the entire model would include the analysis of the moderator. In Figure 3, we present the results of the structural model. The model explained 32.9% of the variance in intention to continue usage and 43.4% of the variance in attitude. Hypothesis 1 posits that perceived information quality influences the attitude toward continuous usage intention. From Figure 3, it can be observed that the path coefficient is .209 (*P*=.003), thus supporting Hypothesis 1. Moreover, Hypothesis 2, which states that the perceived system quality affects the attitude toward continuous usage intention, is also confirmed (beta=.178, *P*=.03). The positive effect of social media influence on the attitude toward continuous usage intention is also supported (beta=.254, *P*<.001), thereby confirming Hypothesis 3. Furthermore, the effects of the control variables—age, gender, and education—on the continuous usage intention have been found to be insignificant.

Figure 3. The conceptual research model based on the elaboration-likelihood model and results of the partial least squares analysis. **P*<.05; ***P*<.01; ****P*<.001.



We also tested the moderation effects of health consciousness between perceived information quality, perceived system quality, social media influence, and attitude. Following prior research, three steps were required to test the mediation effects [44,45]. In Step 1, we treated perceived information quality, perceived system quality, and social media influence as independent variables and attitude as the dependent variable; we found a significant relationship between them: beta=.250, *P*<.001; beta=.205, *P*=.02; and beta=.302, *P*<.001, respectively. In Step 2, we built a model that added health consciousness as an independent variable and attitude as the dependent variable, after which we found a significant effect: beta=.196, *P*=.003; beta=.179, *P*=.03; beta=.261, *P*<.001; and beta=.294, *P*<.001, respectively. In Step 3, we built a moderation model and found that the effects of social media influence and health consciousness on continuous usage intention were significant: beta=.124, *P*=.04. The respective effects of perceived information quality, perceived system quality, and health consciousness on continuous usage intention were insignificant. Thus, health consciousness partially moderated the impact of perceived information quality, perceived system quality, and social media influence on the continuous usage intention. Table 7 shows the results of the hypotheses testing.

Table 7. Results of the hypotheses testing.

Indicator (abbreviation)	Model I		Model II		Model III		Hypothesis test
	Path coefficient	*P* value	Path coefficient	*P* value	Path coefficient	*P* value	
Perceived information quality (PIQ)	.250	.001	.196	.003	.209	.003	Hypothesis 1 was supported
Perceived system quality (PSQ)	.205	.02	.179	.03	.178	.03	Hypothesis 2 was supported
Social media influence (SMI)	.302	<.001	.261	<.001	.254	<.001	Hypothesis 3 was supported
Attitude (AT)	.558	<.001	.558	<.001	.558	<.001	Hypothesis 5 was supported
Health consciousness (HC)	N/A[a]	N/A	.294	<.001	.269	<.001	N/A
PIQ × HC	N/A	N/A	N/A	N/A	.101	.17	Hypothesis 4a was not supported
PSQ × HC	N/A	N/A	N/A	N/A	.044	.56	Hypothesis 4b was not supported
SMI × HC	N/A	N/A	N/A	N/A	.124	.04	Hypothesis 4c was reverse supported
Age	.054	.25	.055	.24	.055	.25	N/A
Gender	.028	.62	.028	.60	.029	.60	N/A
Education	-.090	.28	.013	.83	.013	.83	N/A
Work	.013	.83	.090	.28	.090	.28	N/A
R^2	.333	N/A	.411	N/A	.434	N/A	N/A
R^2 adjusted	.325	N/A	.401	N/A	.418	N/A	N/A

[a]Not applicable.

Discussion

Principal Findings

This study examined the factors that affect the continuous usage intention of mHealth services based on the ELM and resulted in several findings. First, patients perceived mHealth services to have positive information quality; we also observed that the system quality of mHealth services had a positive effect on patients' personal attitudes, indicating that when they perceived the provided mHealth apps to have a higher quality, they would change their attitudes. Our results are consistent with findings in prior studies [46]. Second, the results reveal that social media influence does have a positive effect on a patient's attitude toward mHealth services, which is consistent with findings in existing research [47]. It indicates that effective social media influence can help patients improve their attitudes toward mHealth services.

Third, our study purports that patients' health consciousness does moderate the influence of perceived information quality, perceived system quality, and social media on their attitudes. In particular, our findings suggest that a patient's health consciousness has a positive effect on the relationship between social media influence and attitude, in contrast with Hypothesis 4c. This may be because patients with high health consciousness are more likely to pay attention to the management of personal health; this could lead them to discover new and effective health management modes and to try and continue to find more useful functions for health management, in order to manage their own health. However, patients with low health consciousness may not have a strong awareness of personal health management, and information about health promotion may be directly ignored. Boontarig also illustrated that the participant with a lower level of health consciousness is more concerned about value and convenience in supporting the use of the system. They also elaborated that the participant's personality as a factor affecting a user's perception is important in technology adoption and must be considered when developing and designing these services [48]. Thus, we infer that the effect of social media influence on patients' attitudes toward mHealth services may be mediated through the patients' health consciousness. Moreover, the moderation effects of health consciousness on the relationship between the perceived information quality of the content of mHealth services and the perceived system quality of mHealth services, as well as attitudes toward them, were found to be insignificant. This means that the patient's health consciousness has no moderating effect on the central route (ie, perceived mHealth information quality and perceived mHealth system quality) and the patient's attitude toward mHealth services. This may be because patients pay more attention to the quality of information and the system when using mHealth services and this has nothing to do with personal health consciousness. As long as the quality of mHealth services is verified, patients will choose to use mHealth services. Biduski at el illustrated that the determining aspect in user experience is whether the app features meet users' health needs [49]. As long as the quality of mHealth services can meet the expectations of patients, patients will continue to use mHealth services.

Theoretical Implications

To the best of our knowledge, this is the first study to use the ELM to explore the factors that affect a patient's continuous usage intention on mHealth services based on multidimensional perspectives. Specifically, this study contributes to the mHealth service literature in three aspects.

First, in our research, we expanded the ELM and applied it to the mHealth care domain. Previous researchers have applied the ELM theory to many fields, including e-commerce, consumer behavior advertising, health care, and so on. Rollins and Bhutada illustrated that more highly involved consumers had more positive attitudes, behavioral intentions, and greater information-searching behavior [50]. Sher and Lee studied the effects of consumer skepticism on online shopping based on the ELM [51]. Angst and Agarwal also found that privacy concerns plays a significant role in the adoption of electronic health records [52]. Based on the ELM, this study explained the mechanism of influencing patients' continuous usage intention toward mHealth services.

Second, we explored the central and peripheral routes, with their different motivations and abilities and their effects on user behavior. The central route (ie, perceived mHealth information quality and perceived mHealth system quality) is the factor that patients are most concerned with, which directly affect whether patients accept the mHealth services mode. As well, the peripheral route (ie, social media influence) can affect patients' attitudes toward mHealth services through effective communication channels, such as TV shows, magazines, or advertisements, then change their continuous usage intention of mHealth services. However, due to individual differences (ie, patients' health consciousness), the process and response of these things are not the same, so patients' health consciousness plays a significant role between the central route, the peripheral route, and patients' attitudes toward mHealth services.

Third, we redefined the cognitive-attitude-change process based on the original ELM. The continuous usage behavior of mHealth services needs to go through two stages: the first stage is a process from cognitive-to-attitude change and the second stage is a process from the attitude change to continuous usage, which can be well explained by the ELM theory.

Practical Implications

Our research provides relevant practical implications. First, through the verification of this model, hospital administrators are provided with theoretical support, which enables them to better grasp the patient's psychological characteristics and make their information construction more comprehensive in the process of informatization. Second, for health management entrepreneurs, this model can provide support for the design of health management apps; it can help entrepreneurs clearly identify the problems they encounter and help them decide which directions to take. Third, the patients themselves can gain a clear understanding of their inner tendencies so they can identify health management systems that suit their characteristics within many health management apps, in order to better manage their own health care.

Limitations and Directions for Future Research

Some limitations of this study need to be considered. First, our data was collected through online questionnaires. Although the participants were from all over the country, and though they were representatives of each province, there were also certain limitations. In a future field study, we will need to use a health management software app designed by our laboratory so that we can measure the users' actual usage. Second, we measured continuous usage intention rather than actual usage. Although many theories, such as the theory of reasoned action [53] and the theory of planned behavior [54], and empirical studies have shown that behavioral intention is a reliable proxy for actual human behavior, future studies need to explore how and which factors influence actual usage behavior. Finally, we tried to conduct group experiments for health consciousness—high and low groups—to test whether there were different effects on continuous usage intention when patients experienced different degrees of health consciousness, but because of data constraints, the results did not differ significantly. Hence, we will measure and analyze the effects of this variable in the future.

Conclusions

This study explored the impacts of mHealth services on patient's continuous usage intention from three dimensions, including personal intrinsic factors, organizational factors, and social factors. We found the following: (1) the influence of the individual's internal motivation on his or her health behavior is positively correlated, (2) the influence of external factors, such as social media, on a patient's health behavior is positive, and (3) a patient's health consciousness plays a positive role in changing a person's health continuous usage behavior. The findings of this article are important contributions to the field for both scholars and practitioners.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (71531007, 71471048, 71871074, 71871073, and 71801062).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement items.

[DOCX File , 34 KB - mhealth_v8i4e17258_app1.docx]

References

1. Chen S. The design of personalized artificial intelligence diagnosis and the treatment of health management systems simulating the role of general practitioners. In: Proceedings of the 6th International Conference for Smart Health. 2018 Oct 26 Presented at: 6th International Conference for Smart Health; July 2-3, 2018; Wuhan, China. [doi: 10.1007/978-3-030-03649-2_3]
2. Ventola CL. Mobile devices and apps for health care professionals: Uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
3. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
4. Calisir F, Altin Gumussoy C, Bayraktaroglu AE, Karaali D. Predicting the intention to use a Web-based learning system: Perceived content quality, anxiety, perceived system quality, image, and the technology acceptance model. Hum Factors Ergon Manuf Serv Ind 2014 Aug 28;24(5):515-531. [doi: 10.1002/hfm.20548]
5. Sykes TA, Venkatesh V, Gosain S. Model of acceptance with peer support: A social network perspective to understand employees' system use. MIS Q 2009;33(2):371-393. [doi: 10.2307/20650296]
6. Feldman JM, Lynch JG. Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. J Appl Psychol 1988;73(3):421-435. [doi: 10.1037/0021-9010.73.3.421]
7. Chen M. Attitude toward organic foods among Taiwanese as related to health consciousness, environmental attitudes, and the mediating effects of a healthy lifestyle. Br Food J 2009 Feb 15;111(2):165-178. [doi: 10.1108/00070700910931986]
8. Bhattacherjee A, Sanford C. Influence processes for information technology acceptance: An elaboration likelihood model. MIS Q 2006;30(4):805. [doi: 10.2307/25148755]
9. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York, NY: Springer Verlag; 1986.
10. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: An overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
11. 2017-2023 In-depth Investigation and Future Development Report of China's Chronic Disease Management Industry. URL: http://www.cninfo360.com/yjbg/yyhy/qt/20170414/566492.html [accessed 2020-02-14]
12. Cheung C, Sia C, Kuan K. Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. J Assoc Inf Syst 2012 Aug;13(8):618-635. [doi: 10.17705/1jais.00305]
13. Ho SY, Bodoff D. The effects of Web personalization on user attitude and behavior: An integration of the elaboration likelihood model and consumer search theory. MIS Q 2014 Feb 2;38(2):497-520. [doi: 10.25300/misq/2014/38.2.08]
14. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: Empirical study from elaboration likelihood perspective. Comput Human Behav 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
15. Ayo CK, Oni AA, Adewoye OJ, Eweoya IO. E-banking users' behaviour: E-service quality, attitude, and customer satisfaction. Int J Bank Mark 2016 May 16;34(3):347-367. [doi: 10.1108/ijbm-12-2014-0175]
16. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. Inf Syst Res 2005 Mar;16(1):85-102. [doi: 10.1287/isre.1050.0042]
17. Lien CH, Cao Y. Examining WeChat users' motivations, trust, attitudes, and positive word-of-mouth: Evidence from China. Comput Human Behav 2014 Dec;41:104-111. [doi: 10.1016/j.chb.2014.08.013]
18. Treviño LK, Webster J, Stein EW. Making connections: Complementary influences on communication media choices, attitudes, and use. Organ Sci 2000 Apr;11(2):163-182. [doi: 10.1287/orsc.11.2.163.12510]
19. Erkan I, Evans C. The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. Comput Human Behav 2016 Aug;61:47-55. [doi: 10.1016/j.chb.2016.03.003]
20. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
21. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: A systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
22. Paek H, Hove T. Social cognitive factors and perceived social influences that improve adolescent eHealth literacy. Health Commun 2012;27(8):727-737. [doi: 10.1080/10410236.2011.616627] [Medline: 22452551]
23. Pedhazur EJ, Kerlinger FN. Multiple Regression in Behavioral Research. New York, NY: Holt, Rinehart and Winston; 1973.
24. Bentler PM, Chou C. Practical issues in structural modeling. Sociol Methods Res 2016 Jun 30;16(1):78-117. [doi: 10.1177/0049124187016001004]
25. WJX. URL: https://www.wjx.cn/ [accessed 2020-02-16]

26. Park KS, Kim NJ, Hong JH, Park MS, Cha EJ, Lee TS. PDA-based point-of-care personal diabetes management system. In: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2006 Apr 10 Presented at: 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; September 1-4, 2005; Shanghai, China p. 1-7. [doi: 10.1109/iembs.2005.1617299]
27. Cheong JH, Park M. Mobile internet acceptance in Korea. Internet Res 2005 Apr;15(2):125-140. [doi: 10.1108/10662240510590324]
28. Brady MK, Cronin JJ. Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. J Mark 2018 Oct 10;65(3):34-49. [doi: 10.1509/jmkg.65.3.34.18334]
29. Keery H, van den Berg P, Thompson J. An evaluation of the Tripartite Influence Model of body dissatisfaction and eating disturbance with adolescent girls. Body Image 2004 Sep;1(3):237-251. [doi: 10.1016/j.bodyim.2004.03.001] [Medline: 18089156]
30. Kim SS. The integrative framework of technology use: An extension and test. MIS Q 2009;33(3):513-537. [doi: 10.2307/20650307]
31. Yang H, Yoo Y. It's all about attitude: Revisiting the technology acceptance model. Decis Support Syst 2004 Oct;38(1):19-31. [doi: 10.1016/S0167-9236(03)00062-9]
32. Mai Y, Eisenberg A. Self-assembly of block copolymers. Chem Soc Rev 2012 Sep 21;41(18):5969-5985. [doi: 10.1039/c2cs35115c] [Medline: 22776960]
33. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
34. Chin WW, Marcolin BL, Newsted PR. A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. Inf Syst Res 2003 Jun;14(2):189-217. [doi: 10.1287/isre.14.2.189.16018]
35. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychol Bull 1988;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
36. McLure Wasko M, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q 2005;29(1):35-57. [doi: 10.2307/25148667]
37. Bentler PM. Theory and Implementation of EQS: A Structural Equations Program. Los Angeles, CA: BMDP Statistical Software; 1985.
38. Bentler PM, Wu EJ. EQS for Windows User's Guide. Version 5. Temple City, CA: Multivariate Software; 1995.
39. Ringle CM, Wende S, Will A. Smart PLS 2.0 M3. Hamburg, Germany: University of Hamburg; 2005. URL: https://www.smartpls.com/ [accessed 2020-02-25]
40. Gefen D, Straub D. A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. Commun Assoc Inf Syst 2005;16:91-109. [doi: 10.17705/1cais.01605]
41. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
42. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
43. Rönkkö M, Ylitalo J. PLS marker variable approach to diagnosing and controlling for method variance. In: Proceedings of the International Conference on Information Systems (ICIS). 2011 Presented at: International Conference on Information Systems (ICIS); December 4-7, 2011; Shanghai, China.
44. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
45. Komiak SYX, Benbasat I. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006;30(4):941-960. [doi: 10.2307/25148760]
46. Zhang X, Yan X, Cao X, Sun Y, Chen H, She J. The role of perceived e-health literacy in users' continuance intention to use mobile healthcare applications: An exploratory empirical study in China. Inf Technol Dev 2017 Mar 09;24(2):198-223. [doi: 10.1080/02681102.2017.1283286]
47. Schivinski B, Dabrowski D. The effect of social media communication on consumer perceptions of brands. J Mark Commun 2014 Jan 20;22(2):189-214. [doi: 10.1080/13527266.2013.871323]
48. Boontarig W. Effect of personality factors on attitude towards the adoption of health information via online social networking. In: Proceedings of the 20th International Computer Science and Engineering Conference (ICSEC). 2017 Presented at: 20th International Computer Science and Engineering Conference (ICSEC); December 14-17, 2016; Chiang Mai, Thailand. [doi: 10.1109/icsec.2016.7859897]
49. Biduski D, Bellei EA, Rodriguez JP, Zaina LA, De Marchi AC. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. Comput Human Behav 2020 Mar;104:106169. [doi: 10.1016/j.chb.2019.106169]
50. Rollins B, Bhutada N. Impact of celebrity endorsements in disease-specific direct-to-consumer (DTC) advertisements. Int J Pharm Healthc Mark 2014 May 27;8(2):164-177. [doi: 10.1108/ijphm-05-2013-0024]

51. Sher PJ, Lee S. Consumer skepticism and online reviews: An Elaboration Likelihood Model perspective. Soc Behav Pers 2009 Feb 01;37(1):137-143. [doi: 10.2224/sbp.2009.37.1.137]
52. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. MIS Q 2009;33(2):339. [doi: 10.2307/20650295]
53. Hill RJ, Fishbein M, Ajzen I. Belief, attitude, intention and behavior: An introduction to theory and research. Contemp Sociol 1977 Mar;6(2):244. [doi: 10.2307/2065853]
54. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

Abbreviations

AMOS: analysis of moment structure
ELM: elaboration-likelihood model
EQS: equation structural program
IT: information technology
LISREL: linear structural relationships
mHealth: mobile health
PLS: partial least squares
SEM: structural equation modeling

Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by F Liu, E Bellei; comments to author 15.01.20; revised version received 30.01.20; accepted 01.02.20; published 06.04.20.

Please cite as:
Guo X, Chen S, Zhang X, Ju X, Wang X
Exploring Patients' Intentions for Continuous Usage of mHealth Services: Elaboration-Likelihood Perspective Study
JMIR Mhealth Uhealth 2020;8(4):e17258
URL: https://mhealth.jmir.org/2020/4/e17258
doi:10.2196/17258
PMID:32250277

Review

Clinical Applications of Mobile Health Wearable–Based Sleep Monitoring: Systematic Review

Elise Guillodo[1], MD; Christophe Lemey[2,3], MD; Mathieu Simonnet[2], PhD; Michel Walter[3], MD, PhD; Enrique Baca-García[4], MD, PhD; Vincent Masetti[5], MD; Sorin Moga[6], PhD; Mark Larsen[7], PhD; HUGOPSY Network[8]; Juliette Ropars[9,10], MD; Sofian Berrouiguet[2,3], MD

[1]Urci Mental Health Department, Brest, France
[2]IMT Atlantique, Lab-STICC, F-29238 Brest, Brest, France
[3]EA 7479 SPURRBO, Université de Bretagne Occidentale, Brest, France
[4]Fundacion Jimenez Diaz, Madrid, Spain
[5]Clinea Psychiatry France, Paris, France
[6]IMT Atlantique, Brest, France
[7]Black Dog Institute, University of New South Wales, Sydney, Australia
[8]Please see Acknowledgements for list of collaborators
[9]Laboratoire de Traitement de l'Information Médicale, INSERM, UMR 1101, Brest, France
[10]Department of Child Neurology, University Hospital of Brest, Brest, France

Corresponding Author:
Elise Guillodo, MD
Urci Mental Health Department
Brest Medical University Hospital
Brest, 29200
France
Phone: 33 0298223333
Email: elise.guillodo@chu-brest.fr

Abstract

Background: Sleep disorders are a major public health issue. Nearly 1 in 2 people experience sleep disturbances during their lifetime, with a potential harmful impact on well-being and physical and mental health.

Objective: The aim of this study was to better understand the clinical applications of wearable-based sleep monitoring; therefore, we conducted a review of the literature, including feasibility studies and clinical trials on this topic.

Methods: We searched PubMed, PsycINFO, ScienceDirect, the Cochrane Library, Scopus, and the Web of Science through June 2019. We created the list of keywords based on 2 domains: wearables and sleep. The primary selection criterion was the reporting of clinical trials using wearable devices for sleep recording in adults.

Results: The initial search identified 645 articles; 19 articles meeting the inclusion criteria were included in the final analysis. In all, 4 categories of the selected articles appeared. Of the 19 studies in this review, 58 % (11/19) were comparison studies with the gold standard, 21% (4/19) were feasibility studies, 15% (3/19) were population comparison studies, and 5% (1/19) assessed the impact of sleep disorders in the clinic. The samples were heterogeneous in size, ranging from 1 to 15,839 patients. Our review shows that mobile-health (mHealth) wearable–based sleep monitoring is feasible. However, we identified some major limitations to the reliability of wearable-based monitoring methods compared with polysomnography.

Conclusions: This review showed that wearables provide acceptable sleep monitoring but with poor reliability. However, wearable mHealth devices appear to be promising tools for ecological monitoring.

(JMIR Mhealth Uhealth 2020;8(4):e10733) doi:10.2196/10733

KEYWORDS

sleep; eHealth; telemedicine; review; medicine; wearable electronic devices

Introduction

Sleep Disorders

Sleep disorders are a major public health issue. Nearly 1 in 2 people experience sleep disturbances during their lifetime [1] with a potential harmful impact on well-being and physical and mental health [2]. The International Classification of Sleep Disorders distinguishes the following 6 categories: insomnia, sleep-related breathing disorders, central hypersomnia, circadian rhythm disorders, parasomnias, and sleep-related motor disorders [3]. For example, insomnia is characterized by complaints about the duration and quality of sleep, difficulty falling asleep, nocturnal awakenings, early awakening, or nonrecuperative sleep [4]. This symptomatology must be present at least three times a week for at least 1 month, with negative consequences the next day. Sleep and mental health are highly related, with many mental health problems also being associated with sleeping disorders [5]. Traditionally, sleeping disorders have been viewed as a consequence of mental health disorders, and evidence also suggests that sleeping disorders can contribute to the development of new mental health problems [6].

Sleep Monitoring

Normal sleep is characterized by a succession of 4 to 6 cycles lasting approximately 90 min. Each of these cycles consists of slow-wave phases and rapid eye movement (REM) sleep, which are related to the slowdown and activation of the central nervous system. During REM sleep, or stage 5, REMs are observed and muscle tone is abolished. The early-night cycles are especially rich in deep, slow sleep, and the latter cycles are dominated by REM sleep [7]. The duration of normal sleep varies between 6 and 10 hours depending on several factors, the most important of which are age and genetics.

Normal and pathological sleep can be explored either subjectively, that is, by asking the subject, or objectively, using sensors. An epidemiological study conducted in 2013 with over 1000 participants found the prevalence of subjective insomnia to be 15%, whereas the objective prevalence measured by polysomnography (PSG) was 32% [8]. To date, PSG remains the gold standard for objectively assessing sleep characteristics. The polysomnograph plots a hypnogram, integrating data from several sensors: an electroencephalogram (EEG), an electromyogram (EMG), an electrooculogram (EOG), thoracic movement (from belts on the chest and abdomen), airflow measures, oximetry, and an electrocardiogram (ECG). The sleep stages are scored according to standard visual criteria based on the EEG, EOG, and EMG sensors [5]. The assessment must be carried out under controlled conditions in the laboratory for 8 to 12 hours. An automated hypnogram analysis is possible but still requires manual integration of data [7]. Successful recording of the PSG over the course of the recording and the analysis of the results must be carried out by a clinician with expertise in sleep pathologies and brain disorders. Although PSG is considered the *gold standard*, it is an examination with limitations: it can be cumbersome for the patient, is not very accessible, and is not being realized in ecological conditions.

Mobile Health Wearables

The internet has increased the possibilities for improved patient monitoring. The integration of mobile phones and wearable tools into medical practice has been heralded as the electronic health and mobile health (mHealth) era [9]. These tools can be used to self-monitor or self-assess, allowing individuals to better understand their behavior and body and therefore their health. Aspects of daily life are particularly targeted, with measures of diet, physical activity, or sleep. These self-measurements can be tracked and analyzed with the objective of modifying individual behaviors, including using educational approaches. We therefore observed an association between the concepts of self-monitoring or self-tracking and empowerment, with greater patient involvement and better autonomy. Finally, these devices also allow clinicians to access and review clinical data in real time [10].

Wearables and Sleep Monitoring

The most frequent sensor embodied into commercially available wearables for sleep monitoring is the actimeter. The actimeter uses an accelerometer worn on the wrist and thus detects the movements of the limb [11]. The use of the actimeter has increased because it is easy to use and allows recordings over periods of time longer than a single night of PSG. However, this assessment method has some limitations. Indeed, according to the numerous comparative studies with PSG, it has been shown that the actimeter hardly detects sleepiness, underestimates the latency of falling asleep, and overestimates the number of microawakenings compared with the reference examination. Finally, this device does not provide information on the stages of sleep. Actimetry is therefore limited to subjects with circadian rhythm disturbances and to evaluation of total sleep time (TST) [12]. Some devices use electroencephalographic and electro-oculographic recordings [13]. However, these devices, still not widely used, require the positioning of several electrodes and are therefore impractical for home use by the patient [12].

Other devices measure heart rate and rely on the variability in the heart rate to identify the stages of sleep. Indeed, this variability is higher during paradoxical sleep or nocturnal awakenings and lower during slow sleep [14] because of the sympathetic or parasympathetic action modulations of the autonomic nervous system [12]. These devices are available in different forms, such as watches, chest bands, electrodes, and monitors on the mattress or pillow, but still have poor results [11].

Overall, wearables are promising sleep-monitoring methods and allow for the recording of several nights, whereas PSG assesses only a single night of recording [11]. A total of 3 reviews have examined the potential features of wearable devices for sleep monitoring [15-17]. However, none of these reviews used a systematic review method to report recent clinical research results. Another review recently assessed the efficiency of actigraphy for evaluating mood disorders [18] and activity [19] but did not have any specific focus on sleep monitoring using mHealth wearable methods. Our hypothesis was that the use of wearables was described in the scientific literature. We

therefore conducted a review of the literature on the use of mHealth wearable devices for sleep assessment.

Methods

Objectives and Databases

This literature review aims to identify published articles focusing on wearable-based sleep recording in human participants.

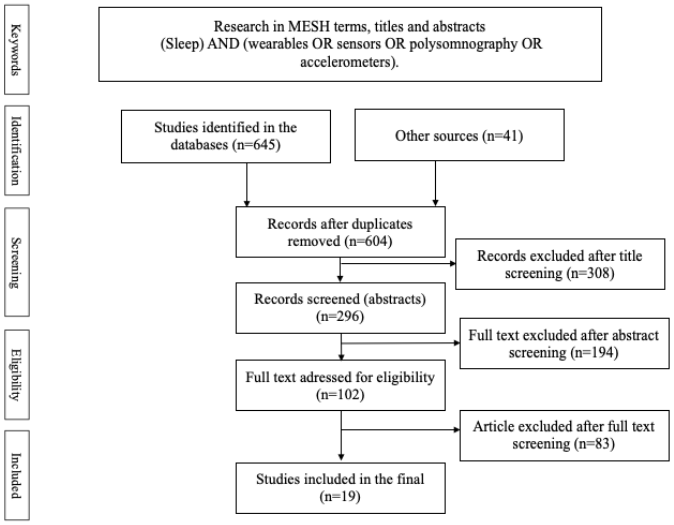
We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [20] to identify, select, and critically appraise relevant research while minimizing bias.

Selection Criteria

The literature search was conducted in June 2019 in the PubMed, PsycINFO, Science Direct, Cochrane Library, Scopus, and Web of Science databases. The keywords used were chosen from the terms used in the health terminology of the biomedical reference thesaurus or MeSH terms. The search was conducted using *AND* and *OR* logistic operators in the MeSH terms, titles, and summaries (Figure 1). The keywords and search strategy we used were (sleep) AND (wearables OR sensors OR polysomnography OR actigraphy).

We included randomized controlled trials and nonrandomized studies. We excluded studies without a clinical population, theoretical articles, editorials, and viewpoints without practical results. We excluded articles with an exclusive focus on technical aspects, sleep-monitoring devices that were not connected to the internet, articles presenting monitoring procedures that were not performed ecologically (ie, at home), and unstructured narrative reviews. We also excluded narrative reviews or any article reporting results in an under 18-year-old population.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart. MESH: Medical Subject Headings.



Data Extraction and Analysis

The analysis of the articles was conducted in two stages. As a first step, a census based on the review of titles and abstracts of scientific papers meeting the inclusion criteria was conducted. Two review authors (SB and EG) independently assessed all studies retrieved against the inclusion criteria. Disagreements were resolved by a third review author (JR). During the literature review process, relevant studies were categorized using a 2-step approach. We first performed a review of the titles and abstracts of all publications that were identified as relevant according to the inclusion criteria. Abstracts were then categorized by the type of methodology used, health condition, applications, and purposes. The full texts of all publications that were not excluded during the title and abstract review stage were checked. Publications that met all inclusion criteria comprised the final sample. The full texts of all publications that were not excluded after the analysis of titles and abstracts were reviewed. All studies meeting the inclusion criteria were included.

Results

Identification of the Articles

The steps of the literature review research and analysis are summarized in Figure 1. The initial search identified 645 articles. After the removal of duplicate articles, screening based on the titles resulted in the removal of 308 articles. A total of 194 articles were excluded after review of the abstracts. After review of the full text, 83 additional articles were excluded because they did not meet the inclusion criteria. Thus, 19 articles meeting the inclusion criteria were included in the final analysis—an overview of these studies is shown in Multimedia Appendix 1.

Design and Size

Of the 19 studies in this review, 58% (11/19) were comparison studies with the gold standard, 21% (4/19) were feasibility studies, 16% (3/19) were population comparison studies, and 5% (1/19) assessed the impact of sleep disorders in the clinic. The samples were heterogeneous in size, ranging from 1 to 15,839 patients.

Analysis of Results

A review of the full text of the articles revealed 4 categories. We identified feasibility studies, population comparison studies, studies comparing mHealth wearables to PSG to identify sleep stages, and a study describing the impact of sleep on clinical outcomes.

Feasibility Studies

In their study, Baron et al [21] aimed to outline the theoretical foundation and iterative process of designing the *Sleep Bunny*, a technology-assisted sleep extension intervention including a mobile phone app, a wearable sleep tracker, and brief telephone coaching. The population comprised 6 adults with short sleep duration (<7 hours), testing the application with the sleep tracker, and the telephone coaching once a week over 4 weeks. The survey, based on open-ended questions, asked participants to provide comments for general feedback on the content and layout of the app. In conclusion, users enjoyed the wearable sleep tracker and found the app visually pleasing but suggested improvements to the notification and reminder features.

The team of Castiglioni et al [22] studied the feasibility of wearing a *MagIC-SCG* sternal wrist device during high-altitude sleep, which is conducive to hypoxia. Their device recorded an ECG, respiratory movements, sternal accelerations, and oxygen saturation. The study demonstrated the feasibility of recording and using the equipment in high-altitude conditions.

The study by Di Rienzo et al [23] demonstrated the feasibility of estimating cardiac functions such as contraction and isovolumic relaxation times or ventricular ejection time during sleep. The data were transmitted in real time to an external device via a Bluetooth connection.

Kayyali et al [24], in the United States, investigated the feasibility of a wearable sleep recorder at the patient's home. Their device, *PSG @ home*, was placed in the thoracic region and recorded respiratory movements, oxygen saturation, airflow, snoring, body position, and an ECG. Their study demonstrated the feasibility of using this discrete device overnight at the subject's home.

Population Comparison

The purpose of the study by Fagherazzi et al [25] was to highlight the determinants involved in poor sleep. The authors calculated the 7-consecutive-night deep sleep/total sleep ratio of a large number of users of Withings wearable devices. They used an algorithm that used the data obtained from both the accelerometer and temperature sensor. A ratio indicating poor sleep was defined as below 0.40. Their findings showed that young men with elevated heart rate and high blood pressure were at higher risk for poor sleep quality.

Migliorini et al [26] compared sleep records between a healthy adult population and a patient with bipolar disorder. The monitoring was performed by the *Smartex* T-shirt equipped with sensors. The data collected were an ECG, respiratory activity. and movement via an accelerometer, allowing the stages of the sleep and an estimation of the percentage of paradoxical sleep to be obtained. The results showed a variability in the reduced heart rate in the individual with bipolar disorder, as well as an increase in the percentage of paradoxical sleep. These results need to be confirmed by a larger sample but seem to be an interesting way of identifying emergent depressive disorders.

The study by Sringean et al [27] compared the sleep of individuals with Parkinson disease in the homes with that of their spouses or partners as healthy controls to provide a quantitative analysis of nocturnal hypokinesia. Wearable sensors were worn on the trunk and limbs. Records included number, speed, acceleration, degree, and duration of movement/turnarounds, number of bed exits, and limb movements. The researchers noted the effectiveness of their system to record nocturnal movements, demonstrating the significant presence of nocturnal hypokinesia.

Comparison With Polysomnography

Comparisons with PSG were conducted using either commercially available devices or custom wearable devices developed specifically for the study.

Commercially Available Devices

The American team of De Zambotti et al [28] compared data from the Jawbone UP with PSG data collected simultaneously. The Jawbone UP is a wristband that, in its first version, records accelerometer data. Comparisons were made between TST, bedtime, sleep latency, and nighttime awakenings. It has been shown that the estimates of these parameters are in good agreement with PSG, a reference examination for sleep pathologies.

Kang et al [20] compared the commercial Fitbit Flex device with PSG in terms of the accuracy of detecting sleep epochs. They studied a population of 41 individuals with insomnia and 21 good sleepers. Participants wore the wearable electronic device while undergoing PSG for 1 night. The measures of interest in this study were TST, sleep efficiency (SE), sleep onset length, and wake after sleep onset (WASO). They concluded that the frequency of agreement was high in good sleepers but significantly low in those with insomnia.

In their pilot study, Looney et al [29] compared electroencephalographic recordings obtained with standard electrodes at the level of the scalp and those obtained with an intra-auricular device simultaneously during sleep. The lines were read blindly by an expert. The results showed a significant concordance between the two recordings.

Parak et al [30] compared the nightly heart rate recording of the connected watch *PulseOn* with the reference test, the ECG. The study, conducted at home, showed that the device correctly detected 99.57% of heartbeats, making it an accurate method during sleep.

Mantua et al [31] compared the data from 5 portable connected devices recording sleep with those of the gold standard, PSG. The devices studied were Actiwatch, Basis, Misfit Shine, Fitbit Flex, and Withings Pulse O2. The recordings were made simultaneously at the participant's home, with participants wearing the 5 devices on the wrists, and PSG was performed. Significant data loss was reported by *Fitbit Flex* and *Misfit Shine*. The correlation analysis allowed them to conclude that there was no significant difference in estimating TST between

PSG and each of the 5 devices. In addition, only *Actiwatch* had concordant data with the baseline SE test. The light sleep time differed between all devices. Finally, a correlation of deep sleep time was significant only for *Basis*.

Liang et al [32] aimed to examine the accuracy of Fitbit Charge 2 for measuring transition probabilities among wake, light sleep, deep sleep, and REM sleep under free-living conditions. A Fitbit Charge 2 and a medical device were used concurrently to measure a whole night's sleep in participants' homes. Sleep data were collected from 23 participants.

Fitbit had the tendency to overestimate the probability of staying in a sleep stage while underestimating the probability of transiting to another stage. SE>90% (P=.05) was associated with a significant increase in measurement error. A Pittsburgh Sleep Quality Index (PSQI)<5 and WASO<30 min could be associated with significantly decreased or increased errors, depending on the outcome sleep metrics.

Dafna et al [33] studied the use of a wearable respiratory sound–recording tool, with the aim of estimating respiratory rate by analyzing the audio signal. The data were compared with those of PSG. The authors concluded that their method was reliable and robust for estimating the respiratory rate. The device was described to be not intrusive and did not interfere with the subject's sleep.

Sano et al [34] compared EEG data with the Q Sensor Affectiva for the detection of waking and sleeping phases in a specialized hospital laboratory. Q Sensor Affectiva is a watch that records skin temperature, cutaneous conductance, and acceleration. In their conclusion, it appears that the combination of acceleration and skin temperature is most effective for the sleep/wake classification.

Sargent et al [35] evaluated the validity of a commercial wearable device, the Fitbit HR Charge, for measuring TST. This study showed that the Fitbit HR Charge overestimated TST for night-time sleep periods and for daytime naps.

Custom Devices

The team of Kuo et al [36] developed and evaluated a hand-held wrist-based sleep-recording tool based on actimetry. The wearable device was judged to be energy efficient and highly accurate in measuring SE, TST, sleep time, and nighttime awakenings. PSG measurements were taken simultaneously. The different variables were concordant and significantly correlated with TST and SE. According to the authors, this system is an interesting option for obtaining objective sleep data at the patient's home.

Rodriguez-Villegas et al [13] compared the effectiveness of a wireless system for the detection of apnea and hypopnea with that of PSG. The 17-gram device was placed on the skin of the anterior aspect of the neck. It recorded turbulence in the trachea using an acoustic chamber. Data were analyzed by blinded investigators. The tolerance of the device was greater than that of PSG. However, the results did not agree with the gold standard regarding the correct detection of hypopneas. In conclusion, this tool could be an adequate solution for the monitoring of apneas in ecological conditions but would not replace a complete recording in the sleep laboratory.

Impact of Sleep on Clinical Outcomes

In the study by Agmon et al [37], the impact of sleep on walking performance in institutionalized elderly individuals was measured using a connected watch and an accelerometer. SE, sleep latency, TST, and nocturnal awakenings were taken into consideration. The team demonstrated that a decrease in recovery sleep was significantly associated with a decrease in start-up speed and a greater variability in walking during double tasks.

Discussion

Principal Findings

This review of the literature shows an increasing interest in the use of wearable devices for sleep assessment. Overall, our review shows that mHealth wearable–based sleep monitoring is feasible but not reliable. Existing commercial technology might be attractive for both clinicians and patients, as shown by the excellent acceptance of mHealth wearable technologies we found. This acceptance has clearly influenced the feasibility of ecological sleep monitoring methods in the selected studies. However, we identified some major limitations to the reliability of wearable-based monitoring methods compared with PSG.

A Global Lack of Reliability

Our study and recent findings indicate that wearables are reliable monitoring tools compared with PSG. However, some recent findings have shown that these devices often over- or underestimate TST or total wake time. This lack of reliability might be partially explained by the power of the trials. For example, only 7 studies among 18 included 30 participants or more [38,39]. Furthermore, recent findings also emphasize that little is still known about physiological monitoring in ecological situations [40], which might explain some discrepancies in results obtained ecologically compared with the gold standard in sleep-recording laboratories. Furthermore, the most common recording methodology identified in this review was motion sensing via accelerometry, in which, recording limits are well established [16]. The research literature consistently shows that wrist accelerometry, even in healthy adults, has high sensitivity but low specificity for sleep detection. The study by Liang [32] showed that Fitbit Charge 2 underestimated sleep stage transition dynamics compared with the medical device. Fitbit had the tendency to overestimate the probability of staying in a sleep stage while underestimating the probability of transiting to another stage. SE>90% (P=.047) was associated with a significant increase in measurement error. PSQI<5 and WASO<30 min could be associated with significantly decreased or increased errors, depending on the outcome sleep metrics. A significant improvement in ecological sleep parameter detection might be provided by recent improvements in miniaturized sensors [41] and embodied data analysis methods.

Limitations of the Review Method

Although the studies selected for this review are recent, the rapid evolution of technologies in this area makes it difficult to

adjust research to keep pace with commercial releases. Since the start of the review process, additional articles may have been published. Moreover, while the conclusions are encouraging, most of them are pilot studies with small samples. Limits on scientific validity mean that these devices are not usable in the current clinical setting, and it may be premature to recommend them. Given the rapid progression of technologies, it does not seem unrealistic to think that more complete and validated devices will be available soon. Moreover, the heterogeneity of the population studied in this review makes it difficult to draw a general conclusion. Our review reflects the broad spectrum of usability of mHealth wearable devices in the field of sleep. Finally, it should be noted that this review of the literature does not provide any data on the use of these objects in the long term because the studies included were mostly short-term clinical trials [29] with devices that may have defects in the collection of data because of limited battery life, for example.

Data mining is the core of analysis and is used to explore clinical questions in large databases as those produced by mHealth wearables. The data mining process includes several steps, including data selection, data processing, and machine learning, to identify which factors may influence results. This review did not identify any studies describing the data mining techniques employed. This finding might be explained by the editorial policies of the clinically oriented articles our inclusion criteria selected. Another reason is that mHealth wearables are commercial products, and the analysis methods are patented. As an example, Fitbit devices that were used in several studies do not give access to the raw data gathered by Fitbit wearables, which make these devices hardly suitable for medical purposes. Consumer sleep devices contribute to the blurry boundary between sleep as a medical concept and sleep as *wellness* and the need for a framework to interpret consumer sleep device outputs.

Future Applications and Recommendations

Regarding the results of our review and as a proposal for future applications and development, some recommendations can be made.

Make Sleep Data Readily and Remotely Available for Nurses and Physicians

Our review shows that sleep can be monitored using wearables for an extensive panel of physical and mental conditions. Most of the research used commercially available devices that were linked to a mobile phone, increasing the networking capabilities and the user experience [28-30]. Collected data can be processed and transferred over the internet to a remote clinical back-end server for further analysis, assessment, decision making, and intervention. However, we noted that the potential to explore sleep remotely and in real time has been poorly reported. Recent research has specifically focused on comparing the reliability of wearables to monitor sleep with PSG. The ability to capture that data, apply machine learning to evolving trends, and alert patients, nurses, and physicians instantaneously is powerful. As sleep is a risk factor for many chronic diseases, the momentary tracking of everyday sleep quality of patients may be very useful for a wide range of clinical conditions, including mental health disorders, neurological disorders, and other chronic diseases. Thus, innovative procedures aiming to make (even simpler than PSG-like signals) outpatient sleep records accessible to clinicians are needed.

The Future of Wearable Sleep Monitoring: Long-Term Assessment

Sleep quality is a key component of health and well-being. Our review shows that most wearables lack the ability to monitor sleep with the same accuracy as PSG. Another important limitation to note is that sleep monitoring using wearables has been poorly explored in a long-term setting. The study duration did not exceed 1 month. However, the main advantage of wearable PSG is that the recording of sleep can be performed over a long period. This ability might help to strengthen or reveal the links between sleep quality and health outcomes, such as depression [42], respiratory problems [43], and epilepsy [44]. Thus, we recommend long-term sleep-monitoring studies.

Development of Specific Sleep-Monitoring Devices

Acceptance is one of the key components of implementation in the clinical setting of wearables. A recent review showed that designing an all-purpose wearable activity tracker (WAT) is unreasonable [19]. A variety of design concepts and data models should continue to emerge that align with the personal preferences of various groups of users. However, it is important to note that most commercially available wearables described in our review have been developed for activity tracking. Sleep monitoring is often presented as a secondary feature of activity trackers. Although some specific sleep-monitoring devices exist, the further development and assessment of devices aiming to specifically monitor sleep are needed. Furthermore, these devices should take more advantage of existing mHealth features, as goal-based gamification, continuous feedback, and social support seem to encourage healthy sleep behaviors.

Commercially Available Versus Custom Wearable Activity Trackers

Our reviews show that most research focuses on commercially available WATs. This important limitation reflects the lack of cooperation among device industries, information technology scientists, and clinical researchers, who might be tempted to implement commercially available wearables instead of developing expensive customized hardware devices. A major limitation of commercially available devices is the poor accessibility of data for analysis purposes, especially in the clinical population. However, analyzing the data generated by commercial wearables is feasible. These data sets are orders of magnitude larger than traditional research studies and can be accessed by researchers at a relatively low cost [45]. However, the consumer market of wellness claims is not necessarily adapted to clinical practice settings and, as a result, may reduce the adoption of these devices in clinical practice by both patients and clinicians. Overall, we believe that further studies should incorporate device developments to better fit long-term and reliable ecological sleep monitoring.

Conclusions

This review of the literature on mHealth wearable devices for sleep monitoring shows the growing interest in these new

technologies, as well as their wide application. Indeed, it was observed that the studies can reflect different specialties of medicine and that the populations studied varied. In addition, this interest is recent, with the majority of studies from 2014 or after. Qualitatively, the majority of devices were considered comfortable [30], easy to use [24], and to preserve the natural sleep of the user [29,33], making them good candidates for home monitoring and care [13,26,36]. In addition, the wearable devices have an economic advantage, and the preliminary results of this study show a good correlation with the reference examination [28]. Given the many benefits, we must consider mHealth wearable devices as promising tools for ecological sleep monitoring. Our review also highlighted some limitations that may help clinicians and researchers better identify current challenges in ecological sleep monitoring using wearables.

Acknowledgments

The HUGOPSY Network collaborators are Bénédicte Gohier, Michel Walter, Morgane Guillou, Christophe Lemey, Florian Stephan, Dominique Drapier, Gabriel Robert, Jean-Marie Batail, Marie Grall-Bronnec, Anne Sauvaget, Nemat Jaafari, Wissam El-Hage, Vincent Camus, Nicolas Ballon, Paul Brunault, and Thomas Desmidt. These Authors are all affiliated with Réseau HUGOPSY, Rennes, France.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the selected studies sorted by year.
[DOCX File , 23 KB - mhealth_v8i4e10733_app1.docx]

References

1. Tan M, Marra C. The cost of sleep disorders: no snoring matter. Sleep 2006 Mar;29(3):282-283. [doi: 10.1093/sleep/29.3.282] [Medline: 16553012]
2. Li SX, Lam SP, Zhang J, Yu MW, Chan JW, Chan CS, et al. Sleep disturbances and suicide risk in an 8-year longitudinal study of schizophrenia-spectrum disorders. Sleep 2016 Jun 1;39(6):1275-1282 [FREE Full text] [doi: 10.5665/sleep.5852] [Medline: 27091530]
3. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 10th Revision. Fifth Edition. Geneva, Switzerland: World Health Organization; 2016.
4. Kerkhofs M. EM consulte. 2015. Troubles du sommeil de l'adulte et de l'adolescent URL: https://www.em-consulte.com/en/article/946739 [accessed 2020-02-03]
5. Baglioni C, Nanovska S, Regen W, Spiegelhalder K, Feige B, Nissen C, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. Psychol Bull 2016 Sep;142(9):969-990 [FREE Full text] [doi: 10.1037/bul0000053] [Medline: 27416139]
6. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
7. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005 Apr;28(4):499-521. [doi: 10.1093/sleep/28.4.499] [Medline: 16171294]
8. Castro LS, Poyares D, Leger D, Bittencourt L, Tufik S. Objective prevalence of insomnia in the São Paulo, Brazil epidemiologic sleep study. Ann Neurol 2013 Oct;74(4):537-546. [doi: 10.1002/ana.23945] [Medline: 23720241]
9. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: transition to participatory and personalized medicine in mental health. J Med Internet Res 2018 Jan 3;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
10. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. A systematic review of wearable patient monitoring systems - current challenges and opportunities for clinical adoption. J Med Syst 2017 Jul;41(7):115. [doi: 10.1007/s10916-017-0760-1] [Medline: 28631139]
11. Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. ISRN Neurol 2012;2012:768794 [FREE Full text] [doi: 10.5402/2012/768794] [Medline: 23097718]
12. Muzet A, Werner S, Fuchs G, Roth T, Saoud JB, Viola AU, et al. Assessing sleep architecture and continuity measures through the analysis of heart rate and wrist movement recordings in healthy subjects: comparison with results based on polysomnography. Sleep Med 2016 May;21:47-56 [FREE Full text] [doi: 10.1016/j.sleep.2016.01.015] [Medline: 27448472]
13. Rodriguez-Villegas E, Chen G, Radcliffe J, Duncan J. A pilot study of a wearable apnoea detection device. BMJ Open 2014 Oct 3;4(10):e005299 [FREE Full text] [doi: 10.1136/bmjopen-2014-005299] [Medline: 25280802]

14. Viola AU, Simon C, Ehrhart J, Geny B, Piquard F, Muzet A, et al. Sleep processes exert a predominant influence on the 24-h profile of heart rate variability. J Biol Rhythms 2002 Dec;17(6):539-547. [doi: 10.1177/0748730402238236] [Medline: 12465887]
15. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011 Aug;15(4):259-267. [doi: 10.1016/j.smrv.2010.10.001] [Medline: 21237680]
16. Russo K, Goparaju B, Bianchi MT. Consumer sleep monitors: is there a baby in the bathwater? Nat Sci Sleep 2015;7:147-157 [FREE Full text] [doi: 10.2147/NSS.S94182] [Medline: 26604847]
17. Ko PT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
18. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: a systematic review and meta-analysis. J Affect Disord 2019 Jun 15;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
19. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
20. Kang S, Kang JM, Ko K, Park S, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009] [Medline: 28606497]
21. Baron KG, Duffecy J, Reid K, Begale M, Caccamo L. Technology-assisted behavioral intervention to extend sleep duration: development and design of the sleep bunny mobile app. JMIR Ment Health 2018 Jan 10;5(1):e3 [FREE Full text] [doi: 10.2196/mental.8634] [Medline: 29321122]
22. Castiglioni P, Meriggi P, Rizzo F, Vaini E, Faini A, Parati G, et al. Seismocardiography while sleeping at high altitude. Conf Proc IEEE Eng Med Biol Soc 2012;2012:3793-3796. [doi: 10.1109/EMBC.2012.6346793] [Medline: 23366754]
23. Di Rienzo M, Vaini E, Castiglioni P, Lombardi P, Parati G, Lombardi C, et al. Wearable seismocardiography for the beat-to-beat assessment of cardiac intervals during sleep. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6089-6091. [doi: 10.1109/EMBC.2014.6945018] [Medline: 25571386]
24. Kayyali HA, Weimer S, Frederick C, Martin C, Basa D, Juguilon JA, et al. Remotely attended home monitoring of sleep disorders. Telemed J E Health 2008 May;14(4):371-374. [doi: 10.1089/tmj.2007.0058] [Medline: 18570567]
25. Fagherazzi G, El Fatouhi D, Bellicha A, El Gareh A, Affret A, Dow C, et al. An international study on the determinants of poor sleep amongst 15,000 users of connected devices. J Med Internet Res 2017 Oct 23;19(10):e363 [FREE Full text] [doi: 10.2196/jmir.7930] [Medline: 29061551]
26. Migliorini M, Mendez MO, Bianchi AM. Study of heart rate variability in bipolar disorder: linear and non-linear parameters during sleep. Front Neuroeng 2011;4:22 [FREE Full text] [doi: 10.3389/fneng.2011.00022] [Medline: 22291638]
27. Sringean J, Taechalertpaisarn P, Thanawattano C, Bhidayasiri R. How well do Parkinson's disease patients turn in bed? Quantitative analysis of nocturnal hypokinesia using multisite wearable inertial sensors. Parkinsonism Relat Disord 2016 Feb;23:10-16. [doi: 10.1016/j.parkreldis.2015.11.003] [Medline: 26683743]
28. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiol Int 2015;32(7):1024-1028 [FREE Full text] [doi: 10.3109/07420528.2015.1054395] [Medline: 26158542]
29. Looney D, Goverdovsky V, Rosenzweig I, Morrell MJ, Mandic DP. Wearable in-ear encephalography sensor for monitoring sleep. Preliminary observations from nap studies. Ann Am Thorac Soc 2016 Dec;13(12):2229-2233 [FREE Full text] [doi: 10.1513/AnnalsATS.201605-342BC] [Medline: 27684316]
30. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:8099-8102. [doi: 10.1109/EMBC.2015.7320273] [Medline: 26738173]
31. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based Actigraphy and Polysomnography. Sensors (Basel) 2016 May 5;16(5):pii: E646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
32. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth 2019 Jun 6;7(6):e13384 [FREE Full text] [doi: 10.2196/13384] [Medline: 31172956]
33. Dafna E, Rosenwein T, Tarasiuk A, Zigel Y. Breathing rate estimation during sleep using audio signal analysis. Conf Proc IEEE Eng Med Biol Soc 2015;2015:5981-5984. [doi: 10.1109/EMBC.2015.7319754] [Medline: 26737654]
34. Sano A, Picard RW. Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data. Conf Proc IEEE Eng Med Biol Soc 2014;2014:930-933 [FREE Full text] [doi: 10.1109/EMBC.2014.6943744] [Medline: 25570112]
35. Sargent C, Lastella M, Romyn G, Versey N, Miller DJ, Roach GD. How well does a commercially available wearable device measure sleep in young athletes? Chronobiol Int 2018 Jun;35(6):754-758. [doi: 10.1080/07420528.2018.1466800] [Medline: 29750580]

36. Kuo CE, Liu YC, Chang DW, Young CP, Shaw FZ, Liang SF. Development and evaluation of a wearable device for sleep quality assessment. IEEE Trans Biomed Eng 2017 Jul;64(7):1547-1557. [doi: 10.1109/TBME.2016.2612938] [Medline: 28113301]
37. Agmon M, Shochat T, Kizony R. Sleep quality is associated with walking under dual-task, but not single-task performance. Gait Posture 2016 Sep;49:127-131. [doi: 10.1016/j.gaitpost.2016.06.016] [Medline: 27414039]
38. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
39. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
40. Brinton JE, Keating MD, Ortiz AM, Evenson KR, Furberg RD. Establishing linkages between distributed survey responses and consumer wearable device datasets: a pilot protocol. JMIR Res Protoc 2017 Apr 27;6(4):e66 [FREE Full text] [doi: 10.2196/resprot.6513] [Medline: 28450274]
41. Shustak S, Inzelberg L, Steinberg S, Rand D, Pur MD, Hillel I, et al. Home monitoring of sleep with a temporary-tattoo EEG, EOG and EMG electrode array: a feasibility study. J Neural Eng 2019 Apr;16(2):026024. [doi: 10.1088/1741-2552/aafa05] [Medline: 30566912]
42. Fan B, Wang T, Wang W, Zhang S, Gong M, Li W, et al. Long-term exposure to ambient fine particulate pollution, sleep disturbance and their interaction effects on suicide attempts among Chinese adolescents. J Affect Disord 2019 Nov 1;258:89-95. [doi: 10.1016/j.jad.2019.08.004] [Medline: 31400628]
43. Bachour P, Bachour A, Kauppi P, Maasilta P, Mäkitie A, Palotie T. Oral appliance in sleep apnea treatment: respiratory and clinical effects and long-term adherence. Sleep Breath 2016 May;20(2):805-812. [doi: 10.1007/s11325-015-1301-0] [Medline: 26754931]
44. Lucey BP, Leahy A, Rosas R, Shaw PJ. A new model to study sleep deprivation-induced seizure. Sleep 2015 May 1;38(5):777-785 [FREE Full text] [doi: 10.5665/sleep.4674] [Medline: 25515102]
45. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]

Abbreviations

ECG: electrocardiogram
EEG: electroencephalogram
EMG: electromyogram
EOG: electrooculogram
mHealth: mobile health
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
REM: rapid eye movement
SE: sleep efficiency
TST: total sleep time
WASO: wake after sleep onset
WAT: wearable activity tracker

Edited by G Eysenbach; submitted 20.06.19; peer-reviewed by H Mufti, D Bowie-DaBreo; comments to author 30.07.19; revised version received 04.10.19; accepted 22.10.19; published 01.04.20.

Please cite as:
Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, Moga S, Larsen M, HUGOPSY Network, Ropars J, Berrouiguet S
Clinical Applications of Mobile Health Wearable–Based Sleep Monitoring: Systematic Review
JMIR Mhealth Uhealth 2020;8(4):e10733
URL: https://mhealth.jmir.org/2020/4/e10733
doi:10.2196/10733
PMID:32234707

Original Paper

A Social Group-Based Information-Motivation-Behavior Skill Intervention to Promote Acceptability and Adoption of Wearable Activity Trackers Among Middle-Aged and Older Adults: Cluster Randomized Controlled Trial

Jing Liao[1,2], PhD; Hai-Yan Xiao[1], BSc; Xue-Qi Li[1], BSc; Shu-Hua Sun[3], MD; Shi-Xing Liu[3], MD; Yung-Jen Yang[4], MD, MSc; Dong (Roman) Xu[2], MPP, PhD

[1]Department of Medical Statistics & Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Sun Yat-sen Global Health Institute, School of Public Health and Institute of State Governance, Sun Yat-sen University, Guangzhou, China

[3]Division of Health Management, Shayuan Primary Health Care Center, Guangzhou, China

[4]Taiwanese Society of Geriatric Psychiatry, Taiwan, China

Corresponding Author:
Dong (Roman) Xu, MPP, PhD
Sun Yat-sen Global Health Institute
School of Public Health and Institute of State Governance
Sun Yat-sen University
#135 Xingang West Road
Guangzhou, 510275
China
Phone: 86 84112657
Email: xudong5@mail.sysu.edu.cn

Abstract

Background: Wearable activity trackers offer potential to optimize behavior and support self-management. To assist older adults in benefiting from mobile technologies, theory-driven deployment strategies are needed to overcome personal, technological, and sociocontextual barriers in technology adoption.

Objective: To test the effectiveness of a social group–based strategy to improve the acceptability and adoption of activity trackers by middle-aged and older adults.

Methods: A cluster randomized controlled trial was conducted among 13 groups of middle-aged and older adults (≥45 years) performing group dancing (ie, square dancing) as a form of exercise in Guangzhou from November 2017 to October 2018. These dancing groups were randomized 1:1 into two arms, and both received wrist-worn activity trackers and instructions at the baseline face-to-face assessment. Based on the Information-Motivation-Behavior Skill framework, the intervention arm was also given a tutorial on the purpose of exercise monitoring (Information), encouraged to participate in exercise and share their exercise records with their dancing peers (Motivation), and were further assisted with the use of the activity tracker (Behavior Skill). We examined two process outcomes: acceptability evaluated by a 14-item questionnaire, and adoption assessed by the uploaded step count data. Intention-to-treat analysis was applied, with the treatment effects estimated by multilevel models.

Results: All dancing groups were followed up for the postintervention reassessment, with 61/69 (88%) participants of the intervention arm (7 groups) and 56/80 (70%) participants of the control arm (6 groups). Participants' sociodemographic characteristics (mean age 62 years, retired) and health status were comparable between the two arms, except the intervention arm had fewer female participants and lower cognitive test scores. Our intervention significantly increased the participants' overall acceptability by 6.8 points (95% CI 2.2-11.4), mainly driven by promoted motivation (adjusted group difference 2.0, 95% CI 0.5-3.6), increased usefulness (adjusted group difference 2.5, 95% CI 0.9-4.1), and better perceived ease of use (adjusted group difference 1.2, 95% CI 0.1-2.4), whereas enjoyment and comfort were not increased (adjusted group difference 0.9, 95% CI –0.4-2.3). Higher adoption was also observed among participants in the intervention arm, who were twice as likely to have valid daily step account data than their controlled counterparts (adjusted incidence relative risk [IRR]=2.0, 95% CI 1.2-3.3). The average daily step counts (7803 vs 5653 steps/day for the intervention and control, respectively) were similar between the two arms (adjusted IRR=1.4, 95% CI 0.7-2.5).

Conclusions: Our social group–based deployment strategy incorporating information, motivation, and behavior skill components effectively promoted acceptability and adoption of activity trackers among community-dwelling middle-aged and older adults. Future studies are needed to examine the long-term effectiveness and apply this social engagement strategy in other group settings or meeting places.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOC-17013185; https://tinyurl.com/vedwc7h.

(JMIR Mhealth Uhealth 2020;8(4):e14969) doi:10.2196/14969

KEYWORDS

mobile health; group exercise; social influence; behavior change; cluster randomized controlled trial

Introduction

The use of mobile technology in health care is becoming increasingly common [1], especially technologies that can support self-management [2]. By quantifying personal health data, mobile devices can facilitate behavior tracking and optimization [3]. Wearable activity trackers represent a subset of consumer mobile devices that can monitor physical activity and collect fitness-related data [4]. These trackers have gained popularity because of their affordable, nonintrusive, and useful features. In addition, recent reviews have suggested their effectiveness in promoting physical activity and improving health [5,6].

Despite their potential effectiveness, the voluntary use of activity trackers among older adults for health purposes remains limited. Compared to younger generations, middle-aged and older adults generally are less technology savvy [7] and continue to lag behind in technology adoption [8,9]. To date, three interrelated categories of barriers in technology adoption among older adults have been identified [10,11]. Person-related factors such as low literacy, limited income, poor health, and declined cognition have been associated with low technology acceptability [12] and adoption [9]. Technology-related factors, particularly perceived ease of use and usefulness, have been shown to largely explain variations in intention to use [13]. Finally, the social context under which technology is being used has also emerged as an important factor limiting adoption [13], such that the lack of assistance and a supporting environment may significantly reduce older adults' intention and ability to use technology [10,14].

The complexity of older users' reactions to mobile technology requires studies targeting improving technology acceptability and adoption in this population [8]. A few small-scale trials on activity tracker usage indicated that adults over the age of 50 years perceived their physical activity self-tracking experiences as acceptable and useful [15-17], particularly for fostering awareness of and motivation for physical activity [16,17]. Although most older adults held a positive attitude toward activity trackers, training and deployment strategies are still essential to overcome usage barriers [18]. Training on the core functions of activity trackers [15] combined with physical activity education [19] has been typically provided to older users as part of the deployment intervention, mainly targeting personal and technological barriers. Although crucial, training alone seems insufficient to meet the unique needs of older adults [19]. Given that activity trackers are less disease-oriented and require frequent user engagement [16], social influence and enjoyment components may be particularly relevant for their adoption [18]. A handful of recent studies examined the impact of sociocontextual drivers. By encouraging social support and comparison via face-to-face group discussions [20], online communications with virtual team members [21], or a combination of offline and online interactions [22], these studies found small yet significant increases in activity tracker usage. Despite these promising findings, two of these studies were not randomized controlled trials (RCTs) [21,23], and the only RCT pilot conducted to date was not properly designed to account for the social clustering effect [20]. Therefore, effective strategies for promoting technology acceptability and adoption among older adults remain elusive.

Considering this previous evidence, we adopted a social group–based deployment strategy and tested its effectiveness to improve the acceptability and adoption of activity trackers by middle-aged and older adults. The social group–based deployment strategy was developed based on the Information-Motivation-Behavioral Skills (IMBS) framework [23]. The IMBS framework specifies three components of behavioral change:

- Information: is directly related to the performance of the given behavior and permits cognitively effortless behavior-related decision making;
- Motivation: includes personal attitudes toward the outcomes (personal motivation) and perceived social norms for engaging in the behavior (social motivation);
- Behavior Skills: are essential for performing the behavioral change through enhancement of individuals' skills and perceived self-efficacy.

We here report the results of phase 1 of a community-based cluster RCT (ChiCTR-IOC-17013185) to assess mobile technology–assisted interventions on the health of middle-aged and older adults. Mapping onto the IMBS framework, the phase 1 study aimed to address key personal, technological, and sociocontextual barriers simultaneously. We hypothesized that older adults who were facilitated by adequate information and behavior skills, and motivated by peers of their social groups would be more likely to accept and adopt wearable activity trackers than their counterparts exposed to these trackers independently.

Methods

Trial Design and Setting

This phase 1 report covers the trial conducted between November 2017 and October 2018, consisting of a 7-month recruitment and baseline assessment stage, a 3-month intervention stage, and the postintervention reassessment. Phase 2 focuses on health outcomes, whereas this phase 1 study prioritizes implementation outcomes. The study was carried out in Guangzhou, the capital city of Guangdong province, China. The trial development was guided by the CONSORT-eHEALTH Checklist [24] and CONSORT-Checklist for reporting a cluster RCT[25].

Participants

In light of the fact that the social influences embedded in an existing social network can better motivate behavior change and adherence [26], the Chinese middle-aged and older adults who routinely practice dancing as a form of physical exercise on squares or other public spaces in groups (ie, square dancing) were chosen as our study population. Their main sociodemographic characteristics have been reported in a related study [27]. Briefly, major public squares and parks of three old districts of Guangzhou, namely Yuexiu, Haizhu, and Liwan, were identified via an online map (Baidu map), considering their land area, visitor flow, and residential locations. Using a restricted randomizing sampling approach, 8 squares and parks per district were chosen at random. We recruited participants in the selected squares and parks using advertisements and flyers. Square dancing groups regularly practicing in the selected places were used as the sampling frame. Dancing groups were eligible for the current study if the dancing style was not ballroom dance, and the total group size was no less than 20 with more than half of the dancers aged 45 years and older. Square dancers of the eligible dancing group were (1) community residents of Guangzhou, (2) regularly practiced square dancing at least once per week in the past 12 months, (3) aged 45 years and older, and (4) agreed to participate in our study if recruited. Participants were excluded if they (1) had serious and uncontrolled diseases related to the heart, brain, lung, liver, and kidney, or any acute complications; and (2) had no smartphone devices (as the data recorded by the wearable activity trackers can only be uploaded to the cloud via a paired smartphone device). Participants were initially screened for eligibility via onsite interviews, and eligible participants were invited for health checkups at the local community health centers on a scheduled date. All participants read and signed the written informed consent form approved by the Institutional Review Board (No. L2016-004) of the School of Public Health of Sun Yat-sen University.

Random Allocation

All random allocation was performed at the cluster level, namely by the square dancing groups. After the recruitment of all eligible participants, a statistician otherwise not associated with the project allocated participants by their square dancing groups equally into two arms (1:1) following a simple randomization process. Although the participants were aware of the interventional nature of the study, they were blinded to their allocation status. Outcome assessors were blinded to the group assignments and were different from the researchers who conducted and monitored the interventions.

Intervention and Procedures

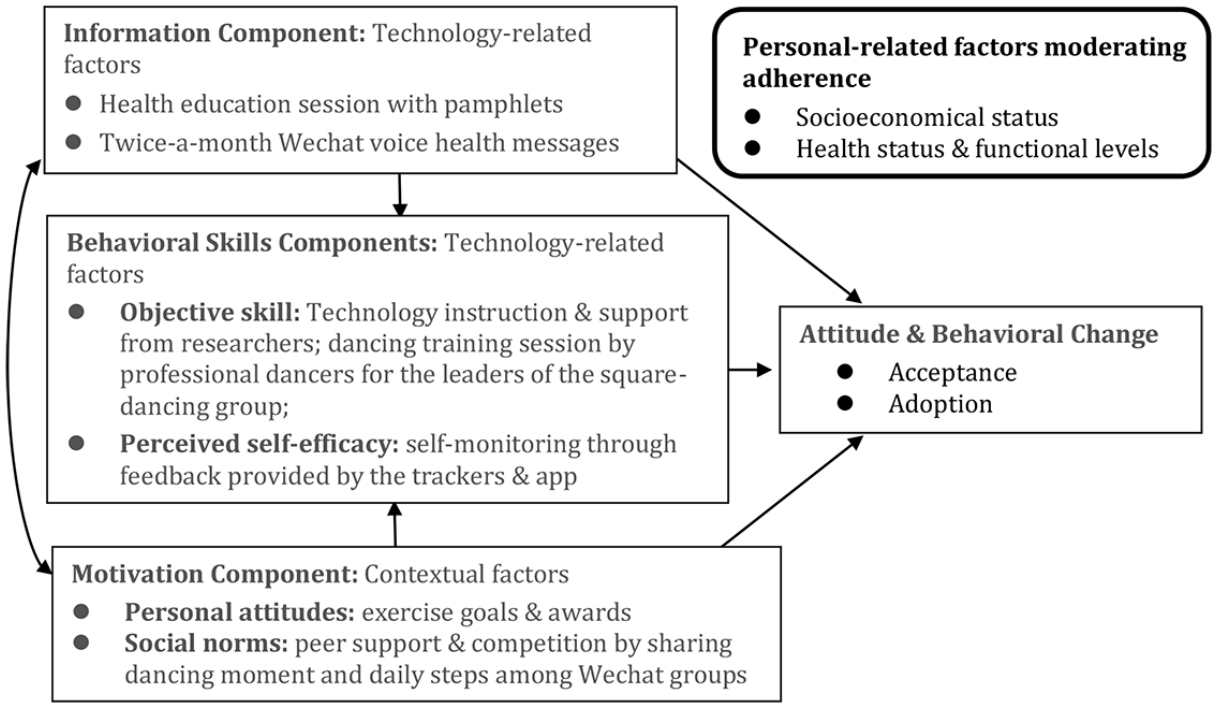
Overview of Intervention

Both the intervention and control arms were equipped with wrist-worn activity trackers free of charge at the baseline assessment. Lifesense MAMBO2 wristbands [28] were used in this study, as their features were deemed to be suitable for our intervention. These activity trackers can automatically record and display participants' step counts, heart rate, and exercising modes (eg, walking, dancing), and generate visualized daily physical activity reports on the paired smartphone via WeChat (the most popular communication and social media platform in China) or the Lifesense app available for ISO and Android systems. Participants' WeChat IDs were linked with their Lifesense accounts to enable physical activity data sharing among peers. Individual participant data were securely aggregated and stored in the managerial account, which were then exported for analysis.

Participants in both arms received a 30-minute demonstration on the core functions of these activity trackers (eg, how to wear, read the displays, and charge the trackers), and were instructed to wear them throughout the day until going to bed. They were encouraged to ask questions and to test these functions during the tracker setup with the researchers. One project facilitator was assigned to assist one dancing group, who conducted the initial setups, provided troubleshooting, and monitored participants' uploaded physical activity data via the managerial accounts.

The control arm received no other assistance, and these participants were left to use the activity trackers independently. The intervention arm was further assisted by the following three intervention components, grounded in the IMBS framework [23] (Figure 1).

Figure 1. The Information-Motivation-Behavioral skills (IMBS) framework for intervention components.



Information Component

Educational information on the benefits of using activity trackers to monitor daily physical activity was delivered in two ways. The first involved a single 1-hour health education session on the recommended age-specific physical activity intensity, duration, and frequency. The participants also received a pamphlet with cartoon and bullet-point messages. The second involved six booster educational voice messages on physical activity-related topics such as warming up and relaxing exercises, which were delivered twice a month via the participants' WeChat groups (defined further below). These messages consisted of a 1-minute voice message, along with the corresponding transcript and illustrations. The content of the education session and voice messages was designed based on the World Health Organization recommendations on physical activity for health [29] and the Chinese adult physical activity guideline [30].

Motivation Component

To motivate participants in using activity trackers for daily activity monitoring according to the physical activity recommendations, we teamed up participants of the same dancing group into their WeChat group, which was named after their dancing group. They were encouraged to share their daily physical activity progress recorded by the activity trackers with their peers in the WeChat group. At the end of each month, individual participants' physical activity rankings identified by their WeChat nicknames were announced within their own group alongside their group's overall physical activity ranking compared to that of other groups of the intervention arm. The participants and the groups with high rankings were awarded with corresponding points, which could be used to redeem gifts.

Behavior Skill Component

Skills on how to use these activity trackers for self-monitoring were delivered through the education sessions described above, assisted by our facilitators who provided in-time technology support and supervision. Additionally, a 2-hour dancing training session was delivered to the lead dancers of the intervention arm, regarding how to design their dancing exercises according to the physical activity recommendations with the assistance of the activity trackers. These lead dancers were identified as the pioneer of their own dancing group, who were able to reinforce behavioral change skills to their dancing peers.

Outcomes

The current study focused on the process outcomes, namely the acceptability and adoption of wearable activity trackers.

Acceptability was defined as users' subjective perception and experiences [6], and was evaluated by a 14-item user feedback questionnaire (Multimedia Appendix 1), adapted from a previous usability and acceptability study [15]. Rated on a 5-point Likert scale from 1 ("strongly disagree") to 5 ("strongly agree"), this questionnaire assessed users' acceptability in four main domains: enjoyment and comfort (3 items, range 3-15; Cronbach alpha=.85), motivation to use (4 items, range 4-20; Cronbach alpha=.83), usefulness (4 items, range 4-20; Cronbach alpha=.89), and perceived ease of use (3 items, range 3-15; Cronbach alpha=.76). A total score was calculated to indicate the users' overall experience (Cronbach alpha=.93, range 14-70).

Adoption was defined as users' interaction and usage behavior [6], and was evaluated objectively via the uploaded step count data in two ways: (1) the percentage of days with valid step records over individuals' follow-up days (average 90.7 days), and (2) the average daily step counts per person of these valid

step records. Daily step counts less than the 5th percentile of the study sample's daily step counts (ie, 1311 steps per day) were treated as invalid records and were removed, as these steps might represent nonwear and inappropriate use of the activity trackers [31].

Control Variables

Participants' demographic characteristics (age, gender, marital status, children), socioeconomic status (education, retirement status, income), as well as self-reported health and any doctor-diagnosed chronic disease were assessed by a self-reported questionnaire. Health-related quality of life was examined by the Short Form Health Survey (SF)-12 [32], and cognitive function was evaluated by the Telephone Interview of Cognitive Status (TICS) and word recall tests, indicating participants' executive function and short-term memory, respectively [33].

Data Collection

Control variables were collected during the recruitment and baseline health checkup prior to the intervention by investigators and clinical staff who were blinded to the intervention assignment. During the intervention, data on participants' daily physical activity level (eg, step counts) were automatically captured and uploaded by the wearable activity trackers and their paired smartphones as indicators for adoption. At the postintervention assessment, participants evaluated their satisfaction with the activity trackers by the 14-item questionnaire as an indicator for acceptability.

To further explore users' experiences with the activity trackers, participants' qualitative feedback was also collected by the group facilitators. In reference to the acceptability questionnaire, participants were encouraged to elaborate their self-monitoring experiences regarding enjoyment and comfort, motivation to use, usefulness, and ease of use. Their feedback was analyzed in a deductive manner to extract information concerning the barriers and facilitators of each acceptability domain. A formal coding process was not applied.

Statistical Methods

The main study sample size was calculated based on changes in physical activity levels. This analysis showed that 12 square dancing groups with an average of 15 participants per arm would have 85% power to detect an increase in physical activity from 1302 to 1500 metabolic equivalent of task minutes per week, assuming an intracluster correlation coefficient (ICC) of 0.05 and a 5% type I error. To further account for a 20% attrition rate, 24 dancing groups across 3 districts were needed to fulfill a total sample size of 440 individual participants.

Descriptive analyses showing participants' baseline characteristics for each arm were summarized at both the individual and cluster levels. The intention-to-treat (ITT) analysis was adopted to examine the treatment effects, minimize selection bias, and maintain the original randomization design [34]. Missing baseline covariates and missing outcomes were multiply imputed under the missing at random assumption, using individual demographic information, health status, and other outcomes and cluster identifiers, separately by randomized arms to avoid biasing treatment effects toward the null [35]. Altogether, 20 sets of complete datasets were imputed based on the chained equations. Primary analyses were then performed on each complete dataset, and combined results were obtained according to Rubin's combination rules [36]. As participants were clustered within dance groups, multilevel linear regression models were used to test for the intervention effect on continuous outcomes (ie, acceptability) at the individual participant level, while taking the cluster-level variation due to dancing groups into account. Similarly, multilevel negative binomial models were fitted to count outcomes (ie, adoption), which followed an overdispersed Poisson-like distribution. All models were adjusted for baseline covariates that were empirically suggested to be strong predictors for the adoption of wearable trackers. The length of individual follow-up days was further adjusted in the multilevel negative binomial model for daily step counts. Sensitivity analyses were conducted among participants with complete cases. Analyses were conducted using STATA 15 (StataCorp, College Station, TX, USA).

Results

Participant Profile

Figure 2 shows the participant flow. Of the 88 dancing groups initially assessed for eligibility, 26 groups did not meet the cluster inclusion criteria mainly because of dancing style and group size. Nearly half of the groups (38/88, 43%) assessed declined to participate due to lack of trust or time. The remaining 11 groups were also not eligible, as their group members were not local residents (n=2), had not been regularly practicing square dancing in the past 12 months (n=5), or less than 2 participants of the given group were willing to participate (n=4). The remaining 13 dancing groups were 1:1 randomized into the intervention arm (n=7) and the control arm (n=6). Among these eligible groups, 69 out of 82 (84%) participants of the intervention arm and 80 out of 98 (82%) participants of the control arm received the allocated treatment. During the follow up, no dancing groups withdrew; 5 participants of the intervention arm and 19 participants of the control arm were lost to follow up, 2 participants of each arm discontinued the intervention due to technical problems, and 1 participant of the intervention arm and 3 participants of the control arm withdrew from the study. The final ITT analysis sample was based on the 149 participants of 13 dancing groups, 117 (78.5%) of whom filled out the user feedback questionnaire of the wearable activity trackers.

Participants' baseline characteristics are presented in Table 1. Sociodemographic characteristics and health status were similar between the two arms. Most of the participants were married older women (mean age 62 years), retired, with an education degree of senior high school; although some had been diagnosed with chronic diseases, all participants were physically and mentally sound. The intervention arm had fewer female participants and lower mean TICS scores than the control arm. The average group size was 9.9 and 13.3 participants for the intervention arm and the control arm, respectively.

Figure 2. The flowchat of participants flow.

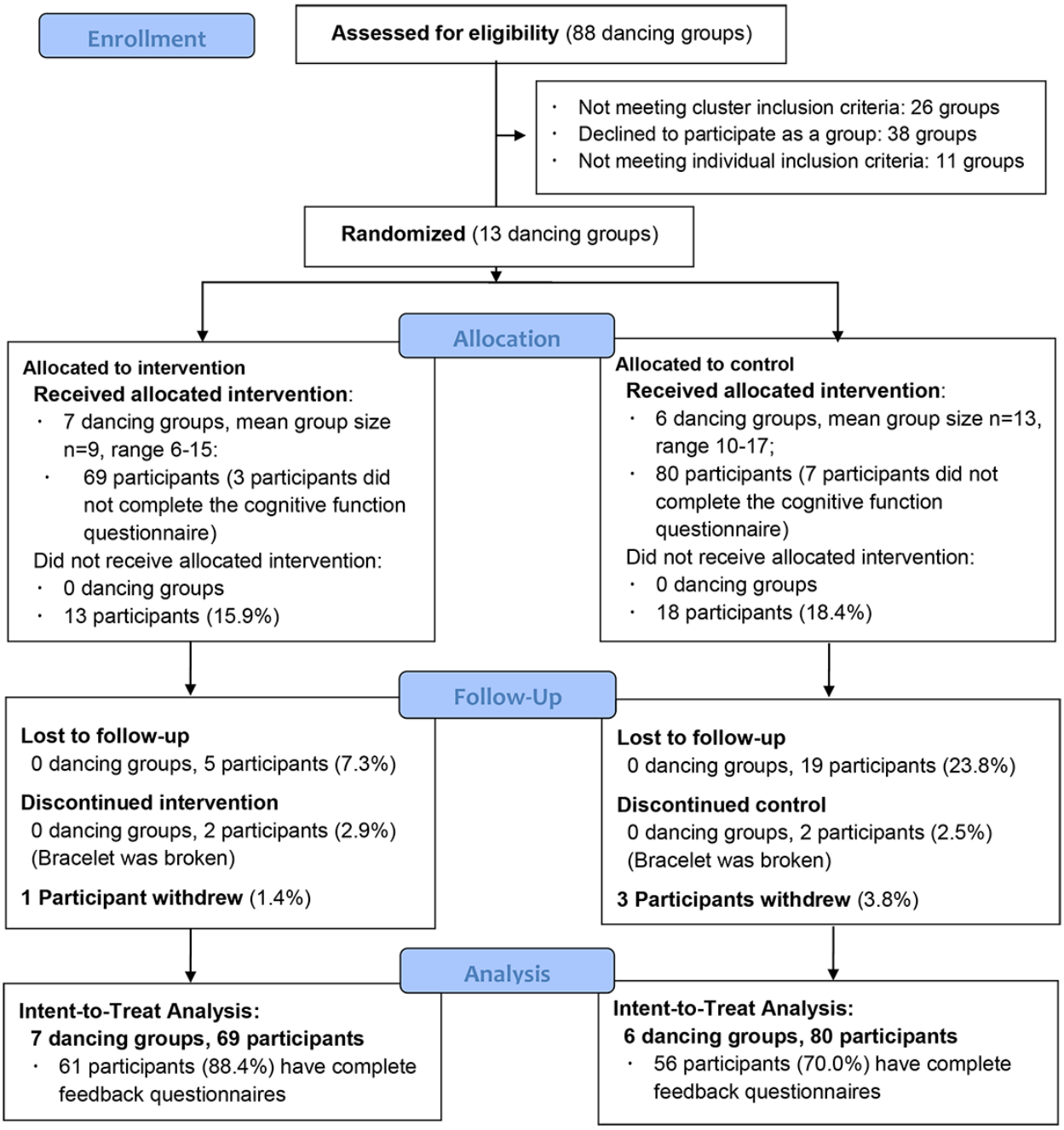


Table 1. Participant characteristics at baseline by intervention and control arms (N=149)[a].

Variables	Intervention arm (n=69)	Control arm (n=80)
Age (mean, SD)	61.8, 5.6	62.0, 5.0
Female (n, %)	63, 91	80, 100
Marital status (n, %)		
Married	60, 87	69, 86
Divorced	1, 1	3, 4
Widowed	6, 9	6, 8
Single	2, 3	2, 3
Having children (n, %)	67, 97	77, 96
Education degree (n, %)		
Primary school and below	6, 9	2, 3
Junior high school	14, 20	14, 18
Senior high school	38, 55	52. 65
University and above	11, 15	12, 15
Retired (n, %)	65, 94	80, 100
Income[b] (n, %)		
<10,000 yuan	15, 25	7, 13
10,000 to <30,000 yuan	20, 32	16, 29
30000 to <50,000 yuan	23, 38	26, 46
50,000 to <10,000 yuan	3, 5	6, 11
≥100,000 yuan	0, 0	1, 2
Self-reported health (n, %)		
Very good	4, 6	9, 11
Good	34, 49	34, 43
Average	30, 43	34, 43
Bad	1, 1	3, 4
Very bad	0, 0	0, 0
Diagnosed chronic disease (n, %)	32, 46	48, 60
SF-12[c] scores (mean, SD)		
Physical Health Score	47.1, 7.7	46.3, 7.5
Mental Health Score	53.5, 8.3	53.1, 8.6
TICS[e] score	8.2, 2.0	8.9, 1.5
Word recall test score	4.7, 1.9	4.5, 1.6
Cluster level (mean, SD)		
Number of dancing groups	7	6
Number of participants per group	9.9, 3.1	13.3, 2.5

[a]Baseline descriptions were based on unadjusted raw data without imputations.

[b]Missing values at baseline were income (n=117) and cognitive function scores (n=139).

[c]SF-12: Short Form Health Survey-12; divided into physical health and mental health scores, ranging from 0 to 100, where a score of 0 indicates the lowest level of heath, and 100 indicates the highest level of health.

[d]Cognitive function was evaluated by the TICS and word recall tests scored from 0 to 10, where higher scores indicate better function.

[e]TICS: Telephone Interview of Cognitive Status.

Outcomes Evaluation

Table 2 presents the unadjusted distributions of acceptability and adoption outcomes at the postintervention assessment in the intervention and control arms, along with the adjusted group differences for acceptability outcomes and the relative risks for adoption outcomes. A clustering effect was observed in most of the measures. The ICC ranged from 0.01 to 0.17 of the acceptability outcomes, and the between-group variance of the adoption outcomes was 0.24.

The unadjusted mean overall acceptability score (range 14–70) was higher for the intervention arm than for the control arm (Table 2). Raw ratings of individual items in each arm are shown in Figure 3. After adjusting for the clustering effect, baseline unbalanced covariates, and predictors for adoption, the absolute group difference in the overall acceptability score was estimated to be higher in the intervention arm than in the control arm. Examination on subdomains of acceptability further indicated that the difference in the overall score was mainly driven by promoted motivation, increased usefulness, and better perceived ease of use, but not due to enjoyment and comfort of using the activity trackers (Table 2).

Regarding adoption outcomes measured by step count data, the median percentage of days that participants had a valid step count record was 44.1% and 11.4% for the intervention and control arms, with 10% and 25% of each arm having invalid step counts records, respectively. As estimated by the multilevel negative binomial models, participants of the intervention arm were twice more likely to have valid daily step count data than their controlled counterparts. The average daily step counts were higher for the intervention arm than for the control arm, but the difference was not significant (Table 2). Similar findings were obtained among participants with complete cases and all covariates in the sensitivity analyses (Multimedia Appendix 2).

Table 2. Participant acceptability and adoption with activity trackers according to intervention status (N=149)[a].

Variables	Intervention Arm (n=69)	Control Arm (n=80)	ICC[b]/ Var-$_{Group}$[c,d]	Adjusted group difference (incidence relative risk and 95% CI)[e]
Acceptability[f], mean (SE)				
Overall Acceptability	45.2 (2.9)	40.4 (1.3)	0.12	6.8 (2.2-11.4)
Enjoyment and Comfort	10.4 (0.6)	9.7 (0.5)	0.09	0.9 (–0.4-2.3)
Motivation of use	11.1 (0.8)	9.7 (0.5)	0.01	2.0 (0.5-3.6)
Usefulness	13.6 (1.1)	12.1 (0.4)	0.17	2.5 (0.9-4.1)
Perceived ease-of-use	9.6 (0.5)	8.9 (0.5)	0.05	1.2 (0.1-2.4)
Adoption, median (25%-75% IQR[g])				
Percentage of days with step counts (%)	44.1 (5.4-84.4)	11.4 (0.5-36.7)	0.24	2.0 (1.2-3.3)
Average daily step count (steps/day)	7803 (5683-9724)	5653 (1052-8462)	1.09E-26	1.4 (0.7-2.5)

[a]Estimates represent the combined results of 20 sets of complete datasets imputed by the chained equations.

[b]ICC: intracluster correlation coefficient.

[c]Var-$_{Group}$: between-group variance.

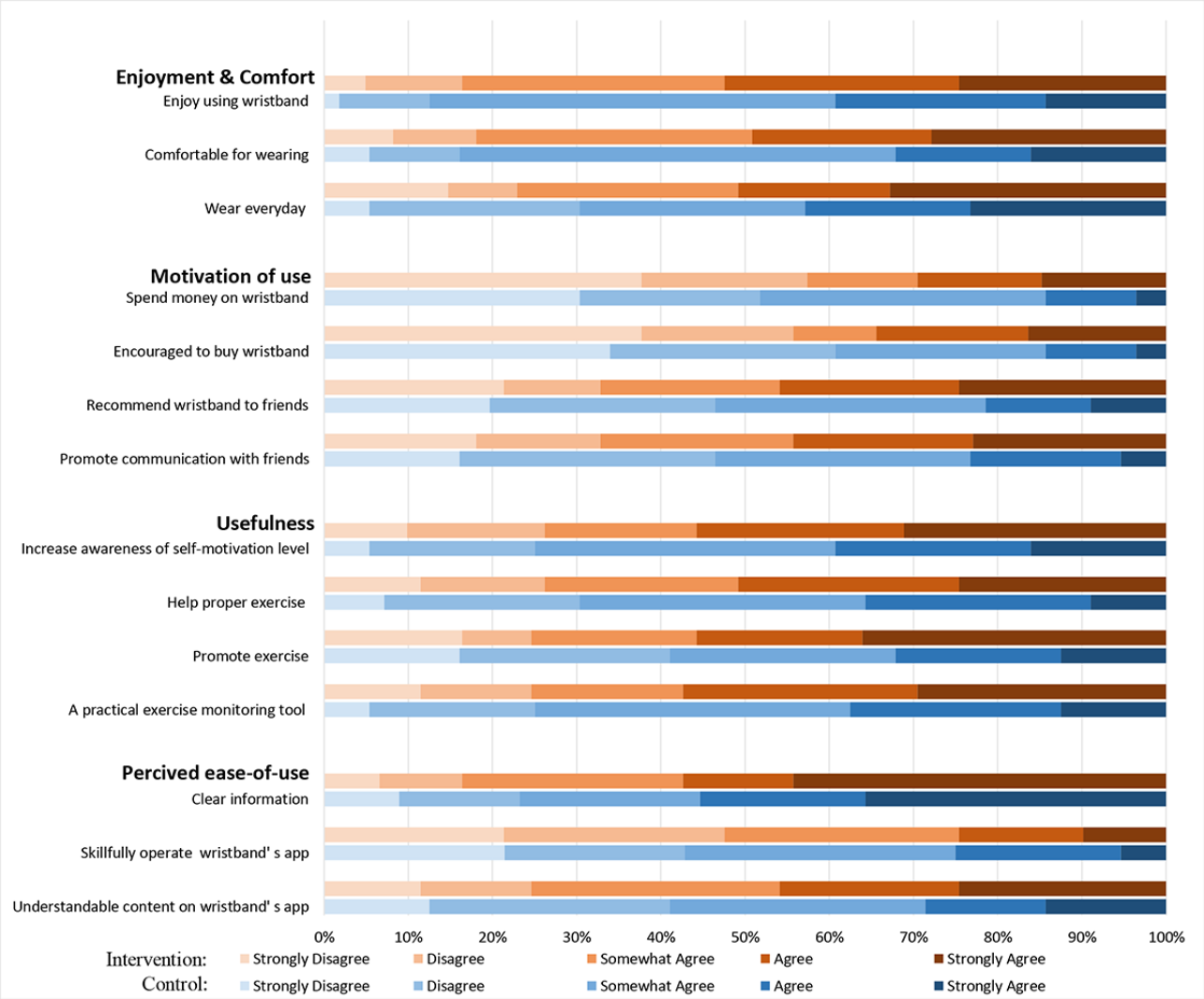
[d]Group comparison models were adjusted for age, gender, education degree, income, SF-12 physical and mental health scores, and cognitive scores measured at baseline; group differences in acceptability outcomes were estimated by multilevel linear regression models.

[e]Incidence relative rate of the adoption outcomes between groups was estimated by multilevel negative binomial models.

[f]Acceptability was evaluated by a 14-item user feedback questionnaire on a 5-point Likert scale from 1 (“strongly disagree”) to 5 (“strongly agree”). The overall acceptability scores ranged from 14 to 70, comprising 4 subdomains: users’ enjoyment and comfort (3 items, range 3-15), motivation of use (4 items, range 4-20), usefulness (4 items, range 4-20), and perceived ease of use (3 items, range 3-15).

[g]IQR: interquartile range.

Figure 3. Indiviudal item rating for acceptability questionnaire by intervetion arms.



Qualitative Feedback From Participants and Observations From Researchers

Qualitative feedback from participants and researchers' observations indicated some common problems encountered by both arms. In terms of enjoyment and comfort of wearing, participants noted that the activity trackers were extremely uncomfortable to wear when the weather was hot and humid, let alone during dancing if they were sweating heavily. Some participants reflected that the figures displayed on the trackers were too small to read, such that they had to wear glasses to read them. Regarding usefulness, a few participants thought that the function of the wearable activity trackers was quite limited, whereas other functions such as blood pressure monitoring were considered to be more useful and relevant from their point of view. Some participants also commented that they preferred the sleep monitoring function to the physical activity monitoring function. As for ease of use, our researchers found that they most frequently received inquiries about how to charge the trackers and how to synchronize data with the paired smartphone.

Discussion

Principal Findings

We assessed the effectiveness of a social group–based deployment intervention informed by the IMBS framework to promote the acceptability and adoption of wearable activity trackers among community-dwelling middle-aged and older adults. In line with our hypotheses, the results revealed that our intervention significantly increased participants' acceptability, mainly driven by improvements in the perceived motivation, usefulness, and ease of use. The intervention also effectively promoted participants' adoption, quantified as twice the amount of valid step count data of the intervention arm than the control arm.

In terms of acceptability, we found that the participants (age range 47-75 years) perceived their self-tracking experiences positively, consistent with findings among Western adults over the age of 50 years [15-18]. Prior trials identified lack of awareness about activity trackers [18] and lack of support while using them (eg, setup, charge battery, and data interpretation) [16,17] as the main use barriers for older adults. Moreover, social support and social learning from a peer [37] were also suggested as key factors to sustain long-term use [15]. In view

of previous findings, our deployment strategy specifically included a tutorial on the activity trackers' functions and their value for participants (Information Component), provided technology support (Behavioral Skill Component), and encouraged social support and comparison within/between groups. Our intervention's effectiveness was demonstrated by the promoted overall acceptability of the intervention arm over the control arm. The three subdomains of acceptability, namely usefulness, ease of use, and motivation, showed an average increase of 2 points postintervention, which may be largely attributed to the corresponding intervention components. However, our intervention did not alter the participants' perceived enjoyment and comfort in using activity trackers. It is suspected that the burden of sharing and synchronizing physical activity data, and the physical discomfort of wearing activity trackers during the humid summer (ie, our intervention period) may have negatively affected their user experiences.

The effectiveness of our deployment strategy was also supported by the adoption data, such that the intervention arm had more frequent interactions with and valid use of the activity trackers than the control arm. These findings agree with prior social connectivity–enhanced trials with device-generated outcomes among older adults. McMahon and colleagues [20] reported that physical activity education delivered to small groups facilitated by in-class discussion and experience sharing (ie, interpersonal strategy) was more effective to promote physical activity than intrapersonal strategies (eg, personal goal setting). Lyons and colleagues [21] assigned participants into premade virtual teams to allow anonymous "likes" and comments, and found small increases in daily walking time and step counts. Butry and colleagues [22] combined both group sessions and online community boards to encourage physical activity assisted by activity trackers, and found frequent tracker usage and increased physical activity that was well maintained over the 6-month follow up. Our social group–based deployment strategy was only supported by the percentage of valid step count data, but not the average daily step counts. Similar daily step counts between the intervention and control arms may reflect a ceiling effect among square dancers, who were already physically active at baseline and were less likely to be more active over a short-time period.

Strengths and Limitations

Our study contributes to the literature of the technology acceptance model (TAM) [38] owing to its rigorous RCT design, and the utilization of existing social networks to foster the acceptance and adoption of activity trackers. The TAM, originating from the field of psychology [39], posits the perceived usefulness and ease-of-use as the only drivers of usage intention and behavior. Although theoretical extensions on the basic TAM constructs have been suggested over the years, the majority of these studies are surveys or single-arm trials without appropriate comparisons, and the effect of sociocontextual modifications on technology acceptance is not well understood [38]. The extent to which older adults would like to interact with unfamiliar individuals in their age group may be quite different from their existing social network [26] where privacy would be less of a concern [17] and constant support is guaranteed. We deliberately applied our intervention among amateur Chinese square dancing groups, leveraging their group dancing routine and frequent interactions to enhance the social influence on their activity tracker-use behaviors. Despite the promising effect identified, we note that our participants were mostly females and middle-aged or slightly older, and thus extrapolating the implication to general older adults should be exercised with caution. Our empirical study nevertheless may inform further investigations on mobilizing community groups and social networks to promote voluntary technology usage among older adults.

Several limitations of our study are worth noting. First, significantly underestimating the challenges in the recruitment, we recruited only 13 dancing groups rather than the 24 groups planned. The participants recruited were thus more likely to be prone to using technology than general middle-aged and older adults. Nevertheless, this self-selection of the participants should not affect the internal validity of the effects of our program, as the intervention and control arms shared similar characteristics and were randomly assigned. However, the self-selection limited our program implication to the broader middle-aged and older population. Second, as our intervention package addressed the three main barriers of behavior changes jointly according to the IMBS framework, we were not able to distinguish the unique contribution of each intervention component to the program effect. Additional research designs may be considered in future studies, such as a factorial experiment [20] that may allow for evaluation of sole and joint effects of such an intervention. Third, our intervention relied on social interactions and required regular assistance from the research staff, particularly in the initial phase. Although social support and comparison functions have been integrated into many activity trackers recently [40], our assistance level is likely to be higher than that typically provided by commercially available activity trackers [16]. Lastly, as a phase 1 study, we have yet to capture and report long-term acceptability, adoption, and health-related outcomes, which are needed to establish the intervention's long-term behavior maintenance and effectiveness.

Conclusion

To ensure that the older population can benefit from mobile technologies, effective deployment strategies to promote technology acceptability and adoption are needed. We applied a social group–based intervention to address personal, technological, and sociocontextual usage barriers to amateur square dancing groups, and found improved acceptability and adoption of activity trackers among middle-aged and older square dancers who were mostly female. Our findings warrant future studies to investigate this social engagement strategy in other group settings of an existing social structure or meeting places, especially among less active male older adults.

Acknowledgments

The authors thank Hanzhao Fu, Sha Chen, and Qianxin Chen from the School of Public Health, Sun Yat-sen University for their assistance in data collection and literature research, and the staff of the Division of Health Management, Shayuan Primary Health Care Center for their support in recruiting the study participants. This work was supported the Guangdong Province Natural Science Foundation (2017A030310346, 2018A0303130046) and Sun Yat-sen Scientific Research Foundation for Early Career Researchers (17ykpy15). The funders played no role in the study design, data collection and analysis, or result interpretation.

Authors' Contributions

JL, SL, YY, and DX contributed to the design of the study. JL and HX wrote the initial draft and conducted the analysis. HX, XL, and SS conducted the data collection and data cleaning. SL, YY, and DX contributed to the revision and preparation of the manuscript. All authors have read and approved the final submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Acceptability questionnaire.
[DOCX File , 21 KB - mhealth_v8i4e14969_app1.docx]

Multimedia Appendix 2
Participants' acceptability and adoption with activity trackers according to intervention status with complete cases.
[DOCX File , 18 KB - mhealth_v8i4e14969_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 364 KB - mhealth_v8i4e14969_app3.pdf]

References

1. Vedel I, Akhlaghpour S, Vaghefi I, Bergman H, Lapointe L. Health information technologies in geriatrics and gerontology: a mixed systematic review. J Am Med Inform Assoc 2013;20(6):1109-1119. [doi: 10.1136/amiajnl-2013-001705] [Medline: 23666776]
2. 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;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
3. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]
4. 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]
5. 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. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
6. McCallum C, Rooksby J, Gray CM. Evaluating the Impact of Physical Activity Apps and Wearables: Interdisciplinary Review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
7. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
8. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
9. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
10. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and Barriers to mHealth Adoption in Older Adults With Heart Failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
11. Peek STM, Wouters EJM, van HJ, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]

12. Chen K, Chan AH. A review of technology acceptance by older adults. Gerontechnology 2011;10(1):1-12. [doi: 10.4017/gt.2011.10.01.006.00]
13. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
14. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to Use mHealth in Older Adults With Heart Failure. J Cardiovasc Nurs 2017 Feb 28;32(6):E1-E7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
15. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older Adults' Experiences Using a Commercially Available Monitor to Self-Track Their Physical Activity. JMIR Mhealth Uhealth 2016;4(2):e35 [FREE Full text] [doi: 10.2196/mhealth.5120] [Medline: 27076486]
16. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR Mhealth Uhealth 2016;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
17. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User Acceptance of Wrist-Worn Activity Trackers Among Community-Dwelling Older Adults: Mixed Method Study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
18. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older Adults' Acceptance of Activity Trackers. J Appl Gerontol 2017 Feb;36(2):127-155 [FREE Full text] [doi: 10.1177/0733464815624151] [Medline: 26753803]
19. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]
20. McMahon SK, Lewis B, Oakes JM, Wyman JF, Guan W, Rothman AJ. Assessing the Effects of Interpersonal and Intrapersonal Behavior Change Strategies on Physical Activity in Older Adults: a Factorial Experiment. Ann Behav Med 2017 Jun;51(3):376-390. [doi: 10.1007/s12160-016-9863-z] [Medline: 28188585]
21. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
22. Butryn ML, Arigo D, Raggio GA, Colasanti M, Forman EM. Enhancing physical activity promotion in midlife women with technology-based self-monitoring and social connectivity: A pilot study. J Health Psychol 2016 Aug;21(8):1548-1555. [doi: 10.1177/1359105314558895] [Medline: 25488937]
23. Fisher W, Fisher J, Harman J. The information-motivation-behavioral skills model: A general social psychological approach to understandingpromoting health behavior. In: Inuls J, Wallston KA, editors. Social psychological foundations of health illness. United Kingdom: Blackwell Publishing; 2003:82-106.
24. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
26. Interventions on diet and physical activity: what works. Geneva, Switzerland: World Health Organization; 2009. URL: https://www.who.int/dietphysicalactivity/whatworks/en/ [accessed 2020-01-03]
27. Liao J, Chen S, Chen S, Yang Y. Personal and Social Environmental Correlates of Square Dancing Habits in Chinese Middle-Aged and Older Adults Living in Communities. J Aging Phys Act 2019 Mar 25:1-7. [doi: 10.1123/japa.2018-0310] [Medline: 30747573]
28. Wu ML, Ke J, Zhong CH. Application of sleep quality management based on comfort theory in perioperative period of knee replacement patients. Orthopaedics 2019;10(03):221-225. [doi: 10.3969/j.issn.1674-8573.2019.03.011]
29. Global recommendations on physical activity for health. Switzerland: World Health Organization; 2010. URL: https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2019-12-30]
30. Zhao W. Chinese adult physical activity guideline. Beijing: People's Medical Publishing House; Jun 1, 2011.
31. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
32. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
33. Lei X, Smith JP, Sun X, Zhao Y. Gender Differences in Cognition in China and Reasons for Change over Time: Evidence from CHARLS. J Econ Ageing 2014 Dec 01;4:46-55 [FREE Full text] [doi: 10.1016/j.jeoa.2013.11.001] [Medline: 25530942]
34. Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. Phys Ther Rev 2013 Jul 19;14(1):36-49. [doi: 10.1179/174328809X405928]
35. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res 2018 Sep;27(9):2610-2626 [FREE Full text] [doi: 10.1177/0962280216683570] [Medline: 28034175]

36. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009 Jul 28;9:57 [FREE Full text] [doi: 10.1186/1471-2288-9-57] [Medline: 19638200]
37. Long JA, Jahnle EC, Richardson DM, Loewenstein G, Volpp KG. Peer mentoring and financial incentives to improve glucose control in African American veterans: a randomized trial. Ann Intern Med 2012 Mar 20;156(6):416-424 [FREE Full text] [doi: 10.7326/0003-4819-156-6-201203200-00004] [Medline: 22431674]
38. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
39. Davis F. A technology acceptance model for empirically testing new end-user information systems: Theory and results. Cambridge, MA: MIT Sloan School of Management; 1985. URL: https://dspace.mit.edu/bitstream/handle/1721.1/15192/14927137-MIT.pdf [accessed 2020-02-21]
40. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]

Abbreviations

ICC: intracluster correlation coefficient
IMBS: Information-Motivation-Behavioral Skills
IQR: interquartile range
IRR: incidence relative risk
ITT: intention to treat
RCT: randomized controlled trial
SF: Short Form Health Survey
TAM: technology acceptance model
TICS: Telephone Interview of Cognitive Status

Edited by G Eysenbach; submitted 08.06.19; peer-reviewed by T Mitzner, Y Wang; comments to author 02.12.19; revised version received 03.01.20; accepted 10.02.20; published 09.04.20.

Please cite as:
Liao J, Xiao HY, Li XQ, Sun SH, Liu SX, Yang YJ, Xu D(
A Social Group-Based Information-Motivation-Behavior Skill Intervention to Promote Acceptability and Adoption of Wearable Activity Trackers Among Middle-Aged and Older Adults: Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e14969
URL: https://mhealth.jmir.org/2020/4/e14969
doi:10.2196/14969
PMID:32271151

Original Paper

Objective Characterization of Activity, Sleep, and Circadian Rhythm Patterns Using a Wrist-Worn Actigraphy Sensor: Insights Into Posttraumatic Stress Disorder

Athanasios Tsanas[1,2*], BSc, BEng, MSc, DPhil (Oxon); Elizabeth Woodward[3*], BSc, DPhil (Oxon); Anke Ehlers[3,4], Dipl (Psych), PhD

[1]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[2]Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, United Kingdom

[3]Department of Experimental Psychology, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

[4]Oxford Health NHS Foundation Trust, Oxford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Athanasios Tsanas, BSc, BEng, MSc, DPhil (Oxon)
Usher Institute
University of Edinburgh
NINE Edinburgh BioQuarter
9 Little France Road
Edinburgh, EH164UX
United Kingdom
Phone: 44 131 651 7887
Email: Athanasios.Tsanas@ed.ac.uk

Abstract

Background: Wearables have been gaining increasing momentum and have enormous potential to provide insights into daily life behaviors and longitudinal health monitoring. However, to date, there is still a lack of principled algorithmic framework to facilitate the analysis of actigraphy and objectively characterize day-by-day data patterns, particularly in cohorts with sleep problems.

Objective: This study aimed to propose a principled algorithmic framework for the assessment of activity, sleep, and circadian rhythm patterns in people with posttraumatic stress disorder (PTSD), a mental disorder with long-lasting distressing symptoms such as intrusive memories, avoidance behaviors, and sleep disturbance. In clinical practice, these symptoms are typically assessed using retrospective self-reports that are prone to recall bias. The aim of this study was to develop objective measures from patients' everyday lives, which could potentially considerably enhance the understanding of symptoms, behaviors, and treatment effects.

Methods: Using a wrist-worn sensor, we recorded actigraphy, light, and temperature data over 7 consecutive days from three groups: 42 people diagnosed with PTSD, 43 traumatized controls, and 30 nontraumatized controls. The participants also completed a daily sleep diary over 7 days and the standardized Pittsburgh Sleep Quality Index questionnaire. We developed a novel approach to automatically determine sleep onset and offset, which can also capture awakenings that are crucial for assessing sleep quality. Moreover, we introduced a new intuitive methodology facilitating actigraphy exploration and characterize day-by-day data across 49 activity, sleep, and circadian rhythm patterns.

Results: We demonstrate that the new sleep detection algorithm closely matches the sleep onset and offset against the participants' sleep diaries consistently outperforming an existing open-access widely used approach. Participants with PTSD exhibited considerably more fragmented sleep patterns (as indicated by greater nocturnal activity, including awakenings) and greater intraday variability compared with traumatized and nontraumatized control groups, showing statistically significant (P<.05) and strong associations ($|R|$>0.3).

Conclusions: This study lays the foundation for objective assessment of activity, sleep, and circadian rhythm patterns using passively collected data from a wrist-worn sensor, facilitating large community studies to monitor longitudinally healthy and pathological cohorts under free-living conditions. These findings may be useful in clinical PTSD assessment and could inform therapy and monitoring of treatment effects.

(JMIR Mhealth Uhealth 2020;8(4):e14306) doi:10.2196/14306

KEYWORDS

actigraphy; sleep; Geneactiv; posttraumatic stress disorder; wearable technology

Introduction

Background

Posttraumatic stress disorder (PTSD) is a mental disorder with lifetime prevalence ranging from 1.9% [1] to 8.8% [2]. However, these figures are considerably exacerbated in conflict, torture, and rape survivors [3,4]. PTSD may develop following exposure to a *traumatic event*, which is defined as *exposure to actual or threatened death, serious injury, or sexual violence* in the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5) [5]. It is characterized by four symptom categories: (1) re-experiencing symptoms associated with the trauma, which include intrusive memories of the event, (2) avoidance of stimuli associated with the trauma, (3) negative alterations in cognition and mood, and (4) alterations in arousal.

Sleep disturbances have been termed the *hallmark* of PTSD [6] but are still poorly understood. They often have a chronic course and are among the most distressing symptoms [7]; these include difficulties falling and staying asleep [8] and nightmares [9]. Poor sleep has been associated with daytime PTSD symptom exacerbation and severity [10,11], indicating a potentially vicious maintenance cycle between sleep disturbances and PTSD symptoms [11], and highlights the importance of investigating sleep to potentially inform targeted treatment [12]. Crucially, from the perspective of this study, the monitoring environment for symptom assessment is critical: PTSD patients report sleeping better in lab settings than at home [13], where problems are more commonly detected [14,15]. This emphasizes the importance of conducting community studies under free-living conditions instead of lab-based assessments. Moreover, given that PTSD is a chronic condition, it would be desirable to develop low-cost monitoring systems that could provide insight into the patient's daily behaviors longitudinally; ideally, such systems would require minimum input from participants.

Longitudinal monitoring of mental disorders is typically achieved using the patient-reported outcome measures (PROMs), where participants are prompted to complete standardized validated questionnaires capturing generic or disease-specific symptoms [16-18]. For example, sleep disturbances are common in mental disorders such as depression and PTSD and cause considerable distress and disability. Sleep diaries are recommended as the gold standard for subjective prospective sleep monitoring [19]. Sleep diaries are typically completed over 1 week, asking participants to report the following (among others): time to bed, how long it took to fall asleep, length and number of awakenings, wake-up time, and sleep quality [19]. However, sleep diaries are inherently limited by relying on an individual's ability to estimate their own sleep times [20]; similar to PROMs, they may be subject to recall bias [21].

Therefore, although self-assessments have merit, they are by nature subjective and may not be easily comparable across individuals [22]. On the other hand, objective monitoring of daily behaviors may offer an additional dimension to understanding pathologies (here, PTSD) using passively collected data. In addition to overcoming the aforementioned problems, sensor data can be streamlined and facilitate direct comparison across participants. There have been considerable developments in the wearable sensor market over the last 5 years; in particular, wrist-worn sensors are widely available and are becoming increasingly affordable. Some have functionalities such as displaying time and communicating with smartphones to convey notifications; in short, these devices are becoming attractive beyond a research exercise, and they are generally embraced by the public. Although many companies have a business model where they employ proprietary algorithms and only offer access to processed data (eg steps, number of stairs, and sleep characteristics such as sleep onset and sleep offset), there are some devices that offer access to the raw *actigraphy* (3D acceleration) data, often complemented with additional modalities such as light, temperature, and heart rate. Overall, the raw signals from the wrist-worn sensor provide a rich source of information and researchers can develop algorithms tailored to the application at hand (often fusing different modalities). For example, actigraphy can be used to objectively quantify physical activity, which can be used as a biomarker for well-being [23]. Although it is currently not possible to obtain accurate estimates of sleep architecture and sleep stages using actigraphy, it provides a good measure of general sleep characteristics including sleep efficiency, total sleep time, sleep onset latency, and awakenings after sleep onset times [24]. Recent work emphasizes participant adherence and the potential of using wrist-worn accelerometers to collect actigraphy data in large-scale population studies [25]. Actigraphy has been used in people with PTSD [10,26] and has shown consistency with sleep diary reports in a PTSD population [27].

Unfortunately, the research literature discussing algorithmic tools to process the raw actigraphy data is limited and fragmented; there is no agreed data-processing protocol and frequently, actigraphy analysis is completed manually relying on visual inspection. This is because studies often have very different focus in terms of applications, for example, to assess general physical activity [25,28] or sleep timings [29], and it is still early days for trying to standardize the commonly extracted features to characterize actigraphy data. Most studies focus on developing and validating actigraphy algorithms using only healthy control cohorts. However, it has been shown across a range of applications that developed algorithms on healthy controls may fail to generalize sufficiently well in cohorts with pathologies [30-32].

Objectives

This study offers a principled framework summarizing many of the known algorithms and introducing new algorithmic approaches to extract potentially useful information from the actigraphy data. Although the application of this study is PTSD, the developed methodology is generic and, in principle,

applicable to settings where actigraphy data are available. To the best of our knowledge, this is the first study to investigate a large set of activity, sleep, and circadian rhythm patterns using a relatively large number of both healthy controls and a mental disorders group where participants commonly exhibit severely fragmented sleep and sleep disturbances.

The aims of this study were to (1) develop an automated approach to accurately determine sleep onset, sleep offset, and awakenings from passively collected data using a wrist-worn sensor targeting a cohort in which sleep is often disturbed, validating findings against participants' sleep diaries; (2) explore differences in these sleep characteristics under free-living conditions between PTSD participants, traumatized controls, and nontraumatized controls; and (3) methodologically contribute toward a large set of activity, sleep, and circadian rhythm patterns to objectively characterize daily behaviors and develop a simple to use package to facilitate actigraphy analysis in MATLAB (The MathWorks Inc).

Methods

Study Cohort

The study cohort comprised 115 participants: 30 nontraumatized controls, 43 trauma-exposed without PTSD, and 42 with PTSD. Inclusion criteria for all groups were as follows: aged between 18 and 65 years, could read and write in English, had no history of or current bipolar or psychosis, no current substance or alcohol dependence, and if they were taking psychotropic medication they had been on a stable dose for at least two months. Additional inclusion criteria for the PTSD group were a traumatic event experience as defined by Criterion A of the DSM-5 [5] and a current diagnosis of PTSD. Inclusion criteria for the control group were no experience of a Criterion A traumatic event and no current mental health problems. Inclusion criteria for the trauma control group were a Criterion A traumatic event experience but not meeting the criteria for a current diagnosis of PTSD. Table 1 summarizes the demographic characteristics of the study cohort. The groups were age and gender matched, and among trauma survivors (n=85), there was no difference in trauma type (interpersonal vs noninterpersonal) between PTSD and trauma controls.

The study received National Health Service (NHS) ethical approval from the South-Central Oxford C Research Ethics Committee (Ref 14/SC/0198). Recruitment began in June 2014 and ended in April 2016. Control and traumatized control participants were recruited via advertisements within the University of Oxford on the departmental website and online community forums in Oxford. The Clinical-Administered PTSD Scale [33], the gold standard for the assessment of PTSD symptom severity, was conducted by EW or JS to establish a PTSD diagnosis according to DSM-5 [5] criteria. All participants completed a sleep questionnaire at trial onset and, over the course of a week, wore a wrist-worn device and kept a detailed sleep diary (see details below).

Table 1. Demographic information for the study participants (N=115).

Demographics	Nontraumatized controls (n=30)	Trauma exposed controls (n=43)	Posttraumatic stress disorder (n=42)
Age (years), mean (SD)	31.17 (10.38)	34.02 (14.01)	32.51 (9.93)
Females, n (%)	23 (76.7)	31 (72.1)	26 (61.9)
Trauma type[a], n (%)			
Interpersonal	N/A[b]	14 (32.6)	20 (47.6)
Not interpersonal	N/A	29 (64.4)	22 (52.4)
Time since trauma (years)[c], mean (SD)	N/A	10.12 (10.18)	8.23 (9.71)

[a]Trauma type and characteristics are only for trauma survivors (n=85).

[b]N/A: not applicable.

[c]Time since trauma was calculated as the time (years) from trauma to study participation date.

Sleep Questionnaire: Pittsburgh Sleep Quality Index

Sleep quality was assessed using the standardized Pittsburgh Sleep Quality Index (PSQI) questionnaire [34]. The PSQI is a self-report assessment comprising 19 items that are mapped onto seven components (each scored in the range 0 to 3): (1) subjective sleep quality, (2) sleep latency, (3) sleep duration, (4) habitual sleep efficiency, (5) sleep disturbances, (6) use of sleeping medication, and (7) daytime dysfunction. The sum of the 7 subscale component scores generates the total score, known as *total PSQI*, which has a range of 0 to 21; scoring above 5 is used as a standard threshold to indicate poor sleep [34]. The study participants self-assessed the occurrence of sleep disturbances over the previous month on a scale from 0 (not during the last month) to 3 (3 or more times a week) for each item. PSQI was collected once in this study and serves as an overall indication of sleep quality.

Sleep Diary

A sleep diary was used to prospectively monitor participants' sleep over 7 days. Participants completed the diary in the morning answering questions about their sleep the previous night. Specifically, they recorded time they got into bed, time they started trying to sleep, sleep onset duration and wake-up time, and number of awakenings (including approximate times), known as wake after sleep onset (WASO). We clarify that by sleep onset, we used the time that participants recorded as falling asleep (ie, time they switched the lights out and started trying

to sleep plus reported sleep onset latency). The sleep diary followed recommendations for sleep research [20] and for the prospective self-monitoring of sleep [19].

Wrist-Worn Device

Participants wore a triaxial accelerometer (Geneactiv, ActivInsights Ltd) on their nondominant wrist for 7 days, coinciding with the sleep diary recording. A recent study endorsed recording at least six nights of standard actigraphy measurements for a reliable measure of self-reported sleep [35]. The device recorded 3D acceleration data at a sampling rate of 100 Hz with a dynamic range of ±8 g (g is a gravity unit, 1 g=9.81 m/second^2) and 12-bit resolution. In addition to triaxial acceleration, light (Lux) and wrist temperature (°C) were recorded. We obtained actigraphy data from 113 participants (the missing data were from 2 PTSD participants).

Methods to Process Data

The following sections describe the methodology used to process the wrist-worn data and statistical tools to compare cohorts.

Calibration and Data Preprocessing

Accelerometers convert mechanical force into electrical signals; in practice, they require calibration to ensure that their outputs are directly comparable. Electrical signals are considered a linear function of the acceleration, involving an offset and a gain factor. In some of the older devices, researchers developed algorithms to compute the calibration values [36]; for the Geneactiv used in this study, the manufacturer provides the offset and the gain factor for the 3 axes for each device. We applied the standard calibration procedure using the manufacturer provided offset and gain factors.

Subsequently, the data were resampled at 10 Hz to simplify further processing. Before extracting data patterns, we automatically detected nonwear times, so that these segments could be excluded from the analysis. The nonwear times were determined following a similar process to Zhou et al [37], marking nonwear periods lasting at least 15 min. The argument is that shorter nonwear periods are unlikely to have a marked effect on the results and attempting to determine very short nonwear times would increase the number of segments mistakenly assigned to be nonwear times.

Data Visualization and Exploration

We produced three main plot types to facilitate visualization and exploration of the data: (1) *Data summary plot*, simultaneously presenting all the raw signal modalities collected (3D acceleration, temperature, and light); (2) *Actogram*, where stacked plots depict activity over 24 or 48 consecutive hours (in the latter case, there is a 24-hour overlap between successive plots) on successive days; (3) *Colored actogram*, which is like an actogram, but here we express the average level of activity on a 10-min window using a color scale to facilitate direct comparison of activity across days.

A critical intermediate step before further visualization and processing of the data requires summarizing the raw triaxial data. Previous research suggests that there are different approaches but no unique single best way to summarize the activity [38,39]. Here, we define two simple summary measures to achieve this, the *movement* and the *xyz variation* variables, which are both expressing activity in 1-min time windows for time-efficient processing and visualization. The *movement* variable is defined as the square root of the sum of the squared successive differences of the triaxial acceleration data. This is similar to the standard Euclidean distance using the 3D data, with the additional twist that successive differences of the raw acceleration entries are used for each of the 3 axes. We have found that empirically this led to more visually appealing results as we are interested more in the *changes* of position in the 3D space. The movement variable is used to efficiently summarize the data and is the key variable we use to succinctly present activity. We also summarize the raw actigraphy data by introducing the *xyz variation*, which we defined as the rolling 10-min median of successive 1-min acceleration differences with a 90% overlapping window.

We have overlaid the nonwear duration times and sleep duration time (see the following section for details) with transparent color in the data summary plot and the actogram. The data summary plots were further annotated using the sleep diary data provided by the participants. In all cases, the aim was to identify trends in the data and develop an intuitive understanding of the continuity and stability of the emerging patterns.

Automatically Detecting Sleep

Previous research has proposed methods to automatically detect sleep using actigraphy, but these are typically evaluated only on healthy controls. Intuitively, we can consider that sleep can be fundamentally determined using actigraphy data on the basis of *sustained inactivity*. Different algorithms proposed in the literature essentially differ on how inactivity is quantified and the use of empirical thresholds, for example, see [29,40,41]. We remark that the latter two approaches were developed using *counts*, a proprietary device-specific estimate of activity. Although there has been a relatively recent attempt to provide backward compatibility with count-based schemes [42], in principle, it would be better to develop approaches using the raw actigraphy data as suggested by van Hees et al [29]. van Hees et al [29] proposed quantifying angular arm movement and assigning time segments to denote sleep when the angle is lower than 5 degrees for 5 successive minutes or more.

The sleep detection algorithm proposed in this study is somewhat more sophisticated. First, we computed the time segments that are considered *sleep candidates* on the basis of the following empirical rules that must be jointly true:

- The rolling 10-min median movement variable is lower than 0.07.
- The rolling 5-min average xyz variation is lower than 0.1.
- The rolling 5-min average light is lower than 30 Lux.

Subsequently, we used a postprocessing approach where sleep candidate segments of at least 30 min in duration were joined if they differed by up to 30 min (to form a continuous sleep candidate segment). The intervening period was recorded as an *awakening* and used to characterize sleep. Finally, all sleep candidate segments less than 2 hours were removed (the threshold might be shortened if we were interested in capturing accurately shorter sleep during the day, at the risk that sedentary

activity might be mislabeled as *sleep*). The proposed sleep detection approach requires the availability of ambient light data (some older actigraphy devices do not collect ambient light), but we have found that including this additional modality overcomes problems with sedentary activity which might otherwise be mislabeled as *sleep*.

Evaluation of Sleep Detection Accuracy

Lacking polysomnography (PSG), which is the gold standard method for thorough assessment in sleep medicine to determine sleep and sleep architecture [32], we evaluated the accuracy of the proposed sleep detection algorithm using the participants' sleep diaries. We aimed to demonstrate its competitiveness against the algorithm by van Hees et al [29] (which had been previously validated against PSG). For the algorithm of van Hees et al [29], we used their implementation in the GGIR package [43]. In all cases, we report the difference between the algorithm's estimate and the ground truth for the purpose of validation (sleep diary).

Characterizing Activity, Sleep, and Circadian Rhythm

The detailed equations for the computation of the extracted patterns are provided in Multimedia Appendix 1. The patterns along with a short description are summarized in Table 2. We remark that the categorization of the features into three groups (*activity*, *sleep*, and *circadian rhythm*) is for reporting convenience, and different categorization approaches or feature membership into these categories are possible. For example, M10, L5, RA, IS, and IV characterize diurnal rhythm behaviors and could be assigned to the circadian rhythm category, as some previous studies have suggested [44,45]. The computation of the features can be achieved using any preprocessing approach that summarizes the raw actigraphy data into a vector. Here, we used the *movement* variable. Alternative approaches using other variables that summarize the 3D actigraphy data would also be possible; we defer further elaboration regarding preprocessing for the Discussion.

Table 2. Summary of activity, sleep, and circadian rhythm patterns used in this study.

Category[a] and pattern	Description
Activity	
M10	Average activity for the 10 most active hours
M10 time	Start time of 10 most active hours
L5	Average activity for the 5 least active hours
L5 time	Start time of 5 least active hours
RA	Relative amplitude of most and least active hours
MDA	Mean diurnal activity (rise time to bed time)
MNA	Mean nocturnal activity
MA	Mean activity with diurnal and nocturnal components
Diurnal skewness	Skewness of the probability distribution of diurnal activity
Percentiles diurnal activity	5, 25, 50, 75, 95 percentiles of diurnal activity
% nocturnal activity (% NA)	Ratio of nocturnal activity over sum 24-hour activity
IS1	Interday stability using 1-hour windows
IS2	Interday stability using 1-hour windows with 30-min overlap
IV1	Intraday variability (24 hours)
IV2	Intraday variability (1440 min)
IV3	Intraday variability (24 hours with 30-min overlap)
Activity TKEO diurnal	Computing the diurnal activity variability using the Teager-Kaiser Energy Operator (TKEO)
Activity ratio TKEO	Ratio of diurnal activity variability against overall activity variability evaluated using TKEO
Activity RMSSD	Computing the diurnal activity variability using the root mean squared successive differences (RMSSD)
Activity ratio RMSSD	Ratio of diurnal activity variability against overall activity variability evaluated using RMSSD
CMSE	Composite multiscale entropy, evaluating the complexity of the time series at 5, 30, 60, 120 min
Sleep	
Sleep onset	Time starting sleep
Sleep offset	Wake up time
Sleep duration	Duration of main (nocturnal) sleep
Number wake-up	Number of wake-up periods during sleep
Wake after sleep onset (WASO) minutes	Minutes awake interrupting sleep
Sleep entropy	Entropy of the activity during sleep (variability of activity during sleep)
Percentiles sleep activity	5, 25, 50, 75, 95 percentiles of activity during sleep
Awakenings total minutes	Total number of minutes awakenings lasted for each automatically detected nocturnal sleep
Circadian rhythm	
Sleep temperature zenith	Maximum temperature during sleep
Sleep temperature zenith time	Time of maximum temperature during sleep
Sleep temperature nadir	Minimum temperature during sleep
Sleep temperature nadir time	Time of minimum temperature during sleep
Sleep temperature range	Range of temperature during sleep
Sleep onset phase	Successive differences in sleep onset timing

Category[a] and pattern	Description
Sleep offset phase	Successive differences in sleep offset timing
Cosinor: MESOR	Cosinor model: average measure of rhythm
Cosinor: Amplitude	Cosinor model: amplitude of fitted sinusoid
Cosinor: Phase	Cosinor model: phase of fitted sinusoid

[a]For algorithmic details, see Multimedia Appendix 1. Overall, we have 49 extracted patterns (counting the percentiles and the CMSE entries separately). We remark that the categorization of the patterns into the three groups (*activity*, *sleep*, and *circadian rhythm*) is for reporting convenience.

Activity

Accelerometers have been traditionally used to quantify physical activity, with an increasing number of wearables and smartphone apps capitalizing on the acceleration signals. There is a growing body of research literature on characteristics that can be extracted from raw actigraphy to quantify activity. Further details regarding the algorithmic expressions for the activity patterns extracted in this study are found in Multimedia Appendix 1.

Sleep

We developed a new approach to detect sleep, which was presented in the section *Automatically Detecting Sleep*. Subsequently, we extracted patterns to quantify nocturnal sleep, including awakenings, disturbances, and periods of excessive activity during sleep, which are of clinical interest in PTSD. The algorithmic details for the sleep patterns are presented in Multimedia Appendix 1.

Circadian Rhythm

Given the reported chronic course of sleep problems and general long-term effects in daily life, we can use the actigraphy, light, and temperature data to extract *circadian rhythm* disturbances, that is, processes with 24-hour oscillations. The term *circadian rhythm*, strictly speaking, refers to endogenous, entrainable processes; chronobiologists prefer the use of the more general term *diurnal rhythm* to describe self-sustained, repeated processes with 24-hour oscillations when their endogenous nature cannot be confirmed. Given that we are extracting both temperature and activity patterns, we will use the former expression as an umbrella term for simplicity.

A standard approach is to measure core temperature and identify minimum and maximum values over consecutive days [46]. Here, we only have access to wrist temperature, but we wanted to test whether there were any intrinsic differences observed in PTSD. Moreover, we computed differences in terms of the activity phase and sleep phases over consecutive days. Algorithmic details of the extracted circadian rhythm patterns are presented in Multimedia Appendix 1.

Statistical Analysis

This section describes the statistical tools used to visualize and compare the findings across the three cohorts.

Density Plots and Statistical Hypothesis Testing

We computed the densities of the extracted patterns using histograms for discrete random variables and using kernel density estimation with Gaussian kernels for continuous random variables. For discrete random variables, we used the chi-square test to determine whether the distributions are statistically significantly different at P=.05 level. For continuous random variables, we used the 2-sample Kolmogorov-Smirnov goodness-of-fit statistical hypothesis test to determine whether the distributions are statistically significantly different, assessing statistical significance at the P=.05 level. The null hypothesis was that the samples are drawn from the same underlying distribution. In all cases, we aimed to assess whether there were statistically significant pairwise differences between the three groups.

Statistical Association Between Patterns and Groups

We computed pairwise statistical associations between the summarized patterns and groups to quantify differences. Specifically, we computed point-biserial correlation coefficients and used the standard empirical rule of thumb approach that correlations with a magnitude larger than 0.3 are *statistically strong* [47,48].

Data Accessibility

Requests for access to the data can be made to EW, but the data cannot be placed into a publicly accessible repository.

Source Code Availability

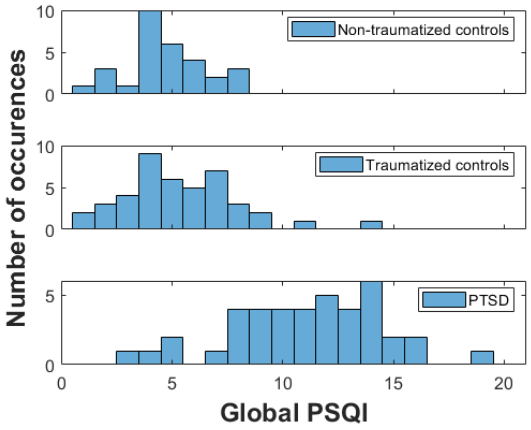
The MATLAB source code for the computation of the actigraphy patterns will be made available on the first author's website [49].

Results

Self-Reported Sleep: Sleep Questionnaire and Sleep Diary

People with PTSD reported more severe subjective sleep disturbances (higher total PSQI scores) compared with both trauma controls and controls (Figure 1). The PSQI was statistically significantly different between PTSD participants and nontraumatized controls (P<.001), and also between PTSD participants and traumatized controls (P<.001). There was no statistically significant difference in terms of PSQI between traumatized controls and nontraumatized controls (P=.73). In the PTSD group, 91% (38/42) of the participants had clinically marked sleep problems at baseline (total PSQI>5).

Figure 1. Histogram of the total Pittsburgh Sleep Quality Index scores for the three cohorts in the study. PSQI: Pittsburgh Sleep Quality Index; PTSD: posttraumatic stress disorder.



Detection of Sleep Onset and Offset With Automatic Algorithms Compared With Sleep Diary Data

Visual inspection of the sleep diaries and comparison with the actigraphy data revealed that a few participants recorded sleep times that differed considerably compared with the standard sedentary activity typically observed in actigraphy data: this may highlight inherent limitations of sleep diaries [20]. Therefore, we wanted to exclude participants where the difference was markedly clear that it seemed possible that the sleep diary entry was not reliable. This was assessed by visual inspection of the actigraphy data and superimposing self-reported sleep entries. In total, 17 out of the 113 participants were excluded from this assessment: 7 due to lack of valid sleep diary entries for at least three days, 1 due to fewer than 3 days of valid actigraphy data, and 9 due to consistent major disagreement between actigraphy and self-reported sleep entries (also see Multimedia Appendix 1). We clarify that by sleep onset, we used the time that participants recorded as falling asleep (not the time they reported as going to bed or starting trying to sleep, in most cases, there was a couple of minutes difference known as *sleep onset latency*).

The agreement of the automatic algorithms against the participants' sleep diaries in terms of sleep onset and sleep offset is summarized in Table 3. For convenience, we summarize the statistical distributions of the errors (differences in minutes between the algorithmic estimates and sleep diaries) in the form median (IQR). Furthermore, Figure 2 presents the error density plots comparing side-by-side the estimated sleep onset and sleep offset of the new proposed sleep detection algorithm and the sleep detection algorithm by van Hees et al [29], against the sleep diaries that are used as ground truth. We remark that the proposed sleep detection algorithm generalizes sufficiently well for both the nontraumatized controls and the PTSD participants. Overall, the new sleep detection algorithm appears to be very accurate and highly competitive against van Hees's algorithm, which has been used in some previous studies. As expected, the results are, in general, more accurate for the nontraumatized control group; the PTSD group appears considerably more challenging, likely due to the nature of the disorder affecting sleep, resulting in more irregular patterns, which may complicate sleep detection. Both algorithms can detect sleep offset (wake-up) more accurately than sleep onset. This indicates that people exhibit greater activity in the morning compared with before bedtime, which intuitively verifies what we would have expected.

To facilitate direct comparison of the two competing sleep detection algorithms across all samples, we also provide scatter plot results in Figure 3. The algorithm proposed in this study is compared with the algorithm suggested by van Hees et al [29], which had been developed for the programming language R (R Core Team, R Foundation for Statistical Computing). We used standard default settings in the GGIR package developed by van Hees. For both the sleep detection algorithm by van Hees and the sleep detection algorithm in this study, the sleep diary data were used only to compare findings. Subsequently, we also present Bland-Altman plots to assess the agreement between the new proposed sleep detection algorithm and the sleep detection algorithm by van Hees [29] (see Figure 4).

Finally, we provide a visual illustration in Figure 5, comparing the algorithm developed by van Hees et al [29] and the algorithm in this study for a randomly selected participant. We observe that the proposed algorithm appears to match very well the participant's self-reported onset and offset in the sleep diary. We remark that in 2 of the nights during the week this participant was monitored, the algorithm detected awakenings (illustrated with noncontinuous transparent green over the course of the night, indicating broken sleep). The participant had reported awakenings on those 2 nights, although the timings of those awakenings were not recorded. We provide further details about the subplots presented in Figure 5 in the next section.

Table 3. Comparison of actigraphy algorithms in accurately detecting sleep: difference in minutes between the algorithms' estimates and the participants' self-reports (sleep diaries).

Cohort	van Hees et al [29] sleep detection algorithm[a]		Proposed sleep detection algorithm in this study	
	Sleep onset, median (IQR)	Sleep offset, median (IQR)	Sleep onset, median (IQR)	Sleep offset, median (IQR)
Nontraumatized controls	−56 (112)	22.5 (106)	−12.5 (51)	2 (30.25)
Traumatized controls	−81 (147)	35.5 (95.5)	−18 (50)	10 (46.75)
PTSD[b] participants	−78 (131.25)	41.5 (122.5)	−34 (78.25)	10 (45.25)

[a]For the algorithm of van Hees et al [29], we used their implementation in the GGIR R package. The results indicate minutes of sleep onset difference and sleep offset difference between the actigraphy algorithm and the ground truth for the purpose of validation (sleep diary). For details on the distributions of the errors, see Figure 2.

[b]PTSD: posttraumatic stress disorder.

Figure 2. Error density plots comparing side-by-side the estimated sleep onset and sleep offset of the proposed sleep detection algorithm and the sleep detection algorithm by van Hees et al against the sleep diaries, which are used as ground truth. These findings are summarized in Table 3. PTSD: posttraumatic stress disorder.

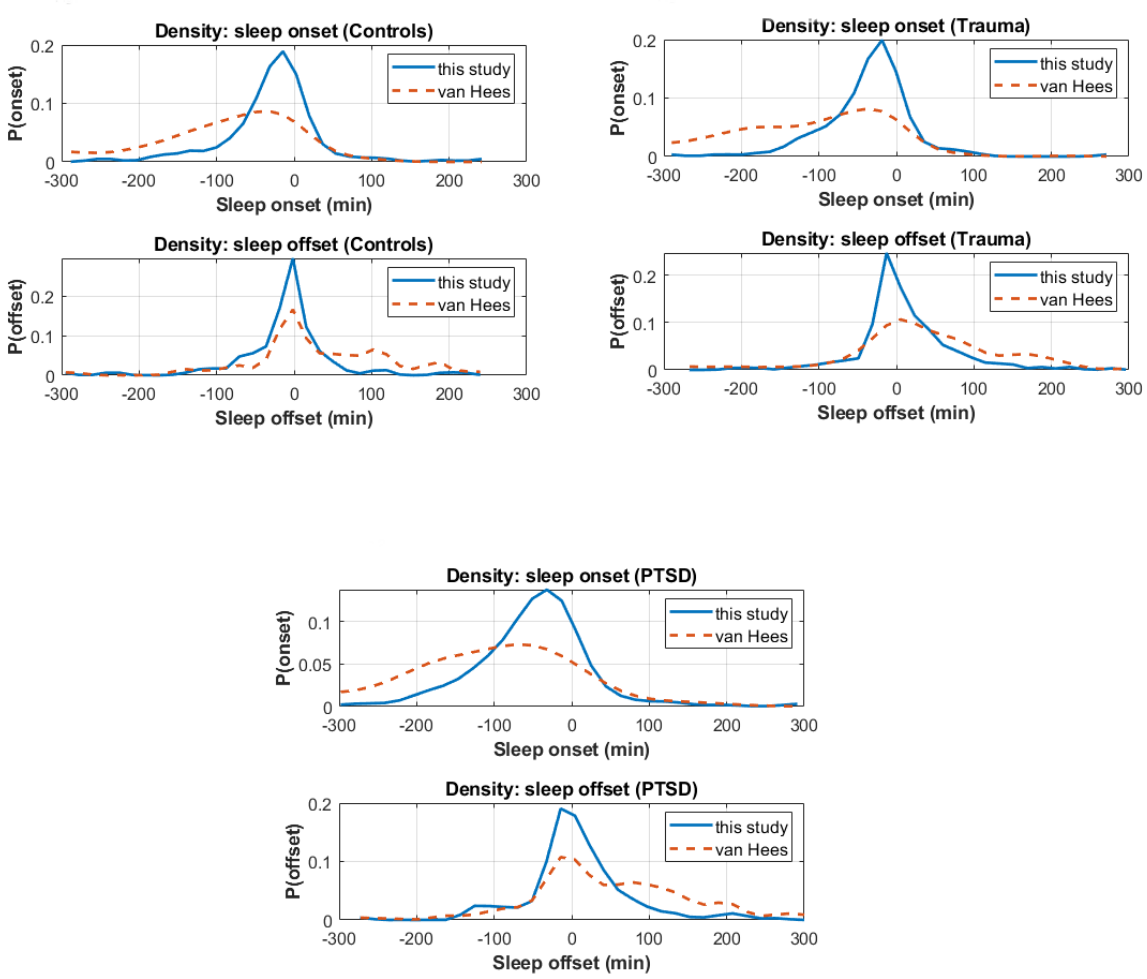


Figure 3. Scatter plots depicting the errors (in minutes) in terms of sleep detection onset and offset across the 3 cohorts for the proposed sleep algorithm against the algorithm proposed by van Hees et al. PTSD: posttraumatic stress disorder.

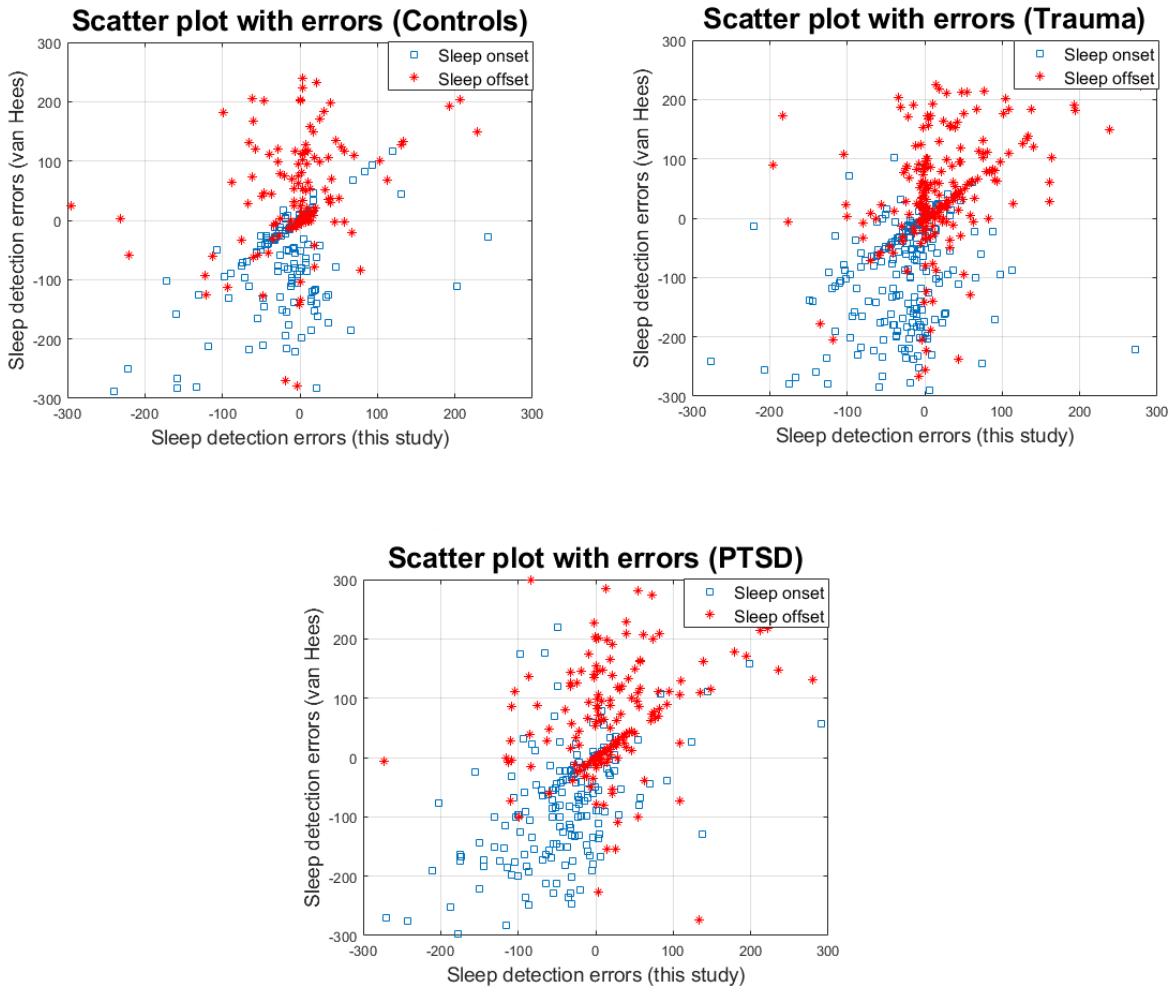


Figure 4. Bland-Altman plots to assess the agreement between the new proposed sleep detection algorithm and the sleep detection algorithm proposed by van Hees et al.

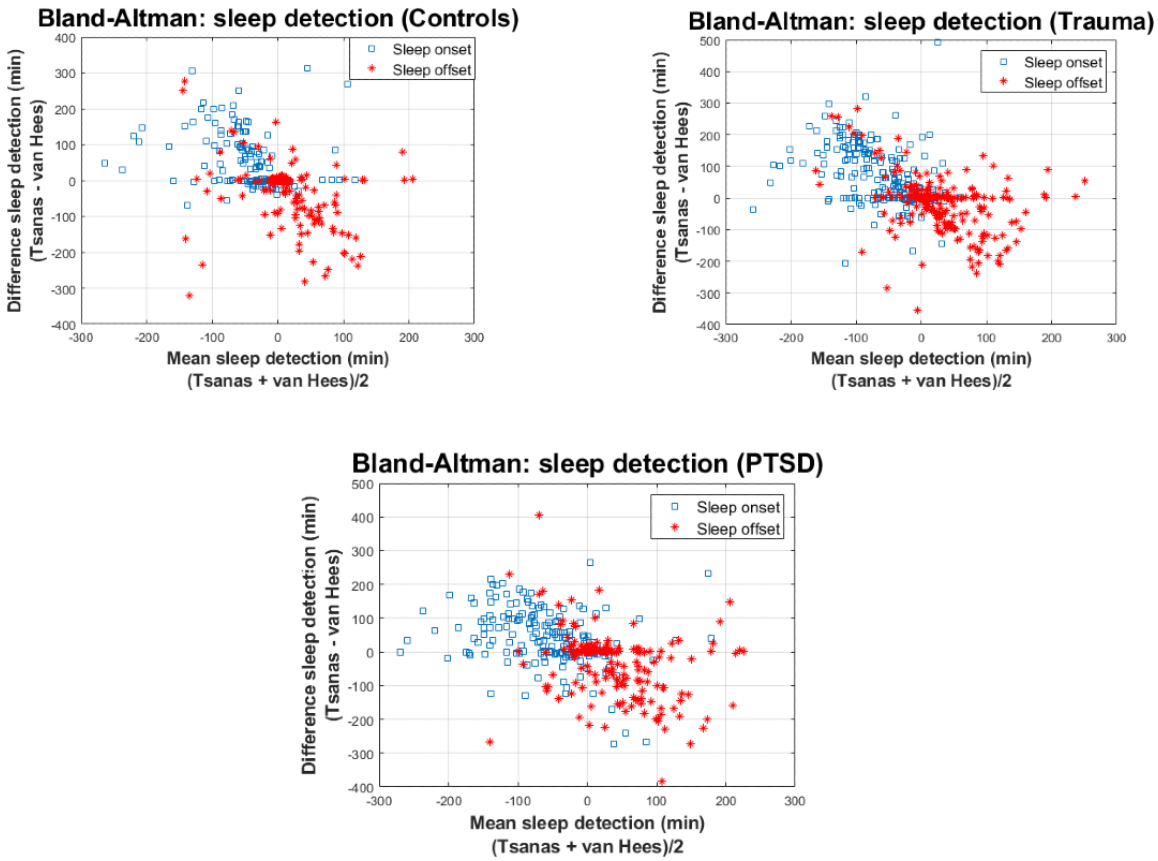
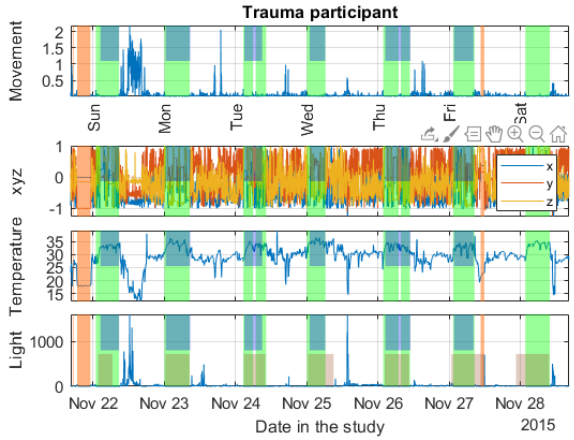


Figure 5. Illustrative indicative example comparing the new sleep detection algorithm with the algorithm proposed by van Hees et al and contrasting findings against the participants' sleep diary entries (focus on the last subplot). Transparent green indicates the detected sleep times using the proposed algorithm, transparent blue (from midway to top of the plot) indicates the ground truth from the sleep diary, and transparent sienna (from bottom to midway in the last plot) indicates the detected sleep by the algorithm of van Hees et al for comparison. Transparent light brown indicates nonwear times.



Data Visualization and Objective Outcomes

We present indicative plots for a PTSD participant to illustrate how data can be presented to visualize patterns. Figure 6 shows the data summary plot presenting the raw data, Figure 7 presents the actogram, and Figure 8 illustrates the activity using a color scale. To protect participant anonymity, we have changed the dates in these illustrations.

There are several important insights to be gained by visualizing the raw data in Figure 5: (1) during sleep, the movement is considerably reduced compared with the rest of the day and there are a few large noted changes across 2 axes usually (indicating an occasional, relatively long arm movement during sleep), (2) during sleep the temperature is elevated compared with the rest of the day, (3) we can observe the continuity of activity and sedentary periods during the day, we observe light

exposure, which is known to have a strong influence on circadian rhythms (in addition, light can be used to support monitoring wake-ups during the night, eg, to detect toilet visits). The automatically assessed sleep times coincide almost perfectly with the participant's self-reported sleep diary across all 6 days in Figure 4 (this participant did not provide an entry for the 7th day in the diary). Further examples and detailed results on the comparison of the sleep detection algorithm against sleep diaries and the sleep detection algorithm of van Hees et al [29] are provided in Multimedia Appendix 1.

The actogram in Figure 7 is useful for visualizing the long-term regularity and potential shift changes in activity. Using the sleep annotation (here automatically deduced), we determined the regularity of sleep timings. We also observed movements during sleep, which can be used to quantify sleep quality.

The colored actogram in Figure 8 complements the standard actogram: visualizing the active times of subsequent days once again provides insight into pattern regularity, which can be useful for extracting circadian rhythm patterns. The extracted patterns in the study were motivated in part by visualizing multiple plots of the form presented in Figures 6-8 from the three cohorts.

Figure 6. Indicative summary of the collected data for one of the posttraumatic stress disorder participants in the study: 3D acceleration (x, y, z axes), temperature, and light. The first row, movement, is a summary metric of the triaxial acceleration (see text for details). The vertical transparent light green color indicates the automatically assessed sleep times; the transparent light brown color indicates nonwear times. The top midway transparent blue indicates sleep diary entries (which can be used as proxy ground truth). PTSD: posttraumatic stress disorder.

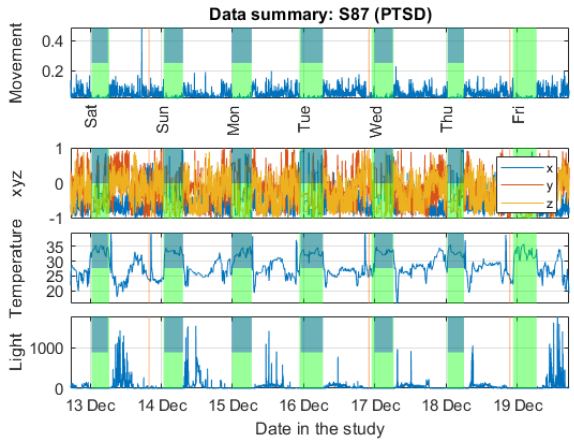


Figure 7. Indicative actogram for one of the posttraumatic stress disorder participants in the study (same participant as in Figure 6). The data on the second half (24:00 to 48:00 hours) of each horizontal plot are repeated as the first (00:00 to 24:00 hours) data on each subsequent horizontal plot; the aim was to have a continuity beyond midnight for the participant. The vertical transparent light green color indicates the automatically assessed sleep times; the transparent light brown color indicates nonwear times. PTSD: posttraumatic stress disorder.

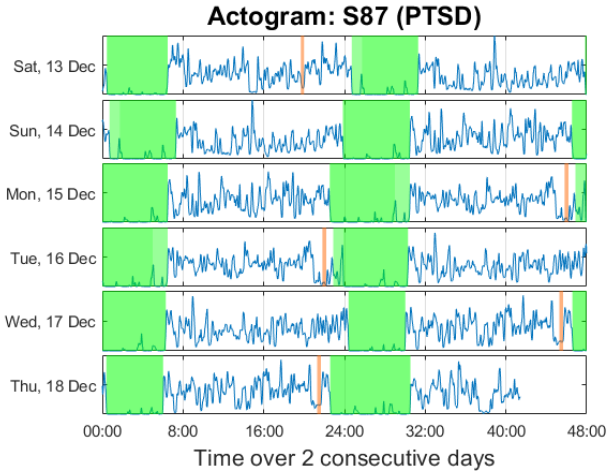
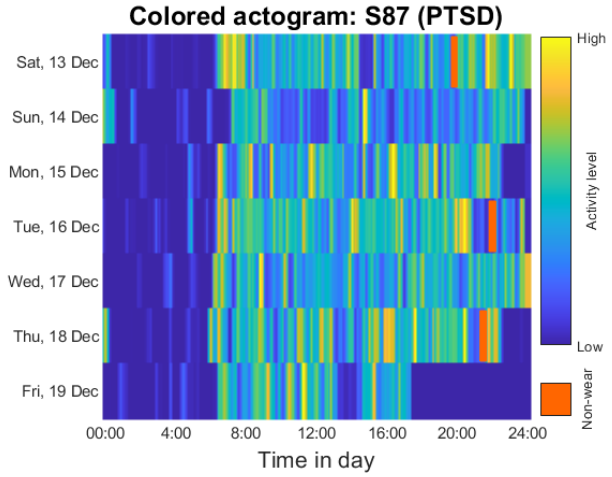


Figure 8. Indicative colored actogram for one of the PTSD participants in the study (same participant as in Figures 6 and 7) to represent activity over 10-min windows. The vertical transparent brown color indicates automatically assessed nonwear times. PTSD: posttraumatic stress disorder.



Group Differences

Ultimately, each of the extracted patterns characterizes each day, that is, for each participant, we typically have a 49×6 (extracted 49 patterns for 6 days). We summarize these patterns in terms of the mean (SD) values for each pattern to characterize each participant. Table 4 summarizes the statistical distributions of the summarized patterns for the three cohorts, along with pairwise comparisons between cohorts. The direction of the effects can be inferred from the signs of the correlation coefficients: negative signs in the correlations indicate that the summarized pattern generally exhibits lower scores for the first group in the comparison. Overall, we observed that the PTSD cohort exhibited some statistically strong ($|R|>0.3$) differences compared with the nontraumatized control group. For example, the sleep entropy appears to be a good differentiator between PTSD and the nontraumatized control group. Descriptive statistics of the patterns shown in Table 4 for each of the three groups are presented in Multimedia Appendix 1. The direction of the effect can be inferred from the sign of the correlation coefficient.

In Multimedia Appendix 1, we present additional plots to convey a visual impression on indicative differences between a PTSD participant, a traumatized control participant, and a nontraumatized control participant.

Table 4. Indicative pairwise statistical comparisons and correlations of the summarized patterns (features) across the three cohorts.

Pattern	Statistical comparisons (*P* values)			Correlations (point biserial correlation coefficient)		
	Control vs trauma	Control vs PTSD[a]	Trauma vs PTSD	Control vs trauma	Control vs PTSD	Trauma vs PTSD
IV2[b]	.06	*.005* [c]	.06	0.22	*0.33*	0.21
Sleep entropy	*.002*	*.005*	.88	*−0.34*	*−0.32*	0.016
Awakenings total minutes	*.25*	*.01*	.16	0.14	*0.30*	0.16
Number wake ups	.33	*.03*	.06	0.12	0.27	0.21
WASO[d] total minutes	.41	*.03*	.05	0.10	0.26	0.21

[a]PTSD: posttraumatic stress disorder.

[b]IV2: intraday variability (1440 min).

[c]Statistically significant associations (at the *P*=.05 level) are italicized. We present five indicative summarized patterns that exhibit the largest correlation magnitudes for the binary comparisons between groups. The negative signs in the correlations indicate that the summarized pattern generally exhibits lower scores for the first group in the comparison. Correlations with a magnitude over 0.3 are considered statistically strong. Further details are presented in Multimedia Appendices 1 and 2 (Multimedia Appendix 2 presents all the investigated variables).

[d]WASO: wake after sleep onset.

Discussion

Principal Findings

Sensor-based at-home monitoring is rapidly emerging with the proliferation of wearables. Wearables provide a convenient platform for chronic condition management, facilitating detailed longitudinal assessment across diverse metrics, including activity, sleep, and circadian rhythm patterns. This study demonstrated the potential of using raw triaxial acceleration, light, and temperature data to gain insight into weekly patterns of activity and sleep in the participants' everyday life. We illustrated compact approaches to visualize annotated data modalities and proposed a novel algorithmic approach to

estimate sleep onset and offset (including awakenings) that appears to be very accurate against sleep diaries.

We emphasize that the extent of our claim regarding the accuracy of the new sleep detection algorithm is that it replicates reasonably accurately the sleep diaries, which is a proxy for the true sleep onset and sleep offset. Any automated sleep detection algorithm validated on sleep diaries implicitly relies on having accurate labels to report performance and would ideally need to be benchmarked against PSG. We decided to exclude 9 participants from further analysis due to consistent major disagreement between actigraphy and self-reported sleep diary entries following visual inspection of the data summary plots (see Multimedia Appendix 1).

Importantly, some of the actigraphy algorithms were sensitive in showing distinguishable patterns for participants with PTSD compared with nontraumatized controls, which were in line with the self-reported sleep disturbances. These results are summarized in Table 4 and Multimedia Appendix 1. For example, sleep-based features (awakenings, WASO minutes) may provide a useful objective indicator of the degree of restlessness experienced by PTSD patients during sleep. Similarly, the intraday variability (IV2) differentiates well PTSD from nontraumatized controls, possibly indicating bursts of activity in PTSD throughout the day, whereas nontraumatized controls exhibited less activity variability within each day on average. Collectively, these findings are well reflected in the PTSD literature, where PTSD patients are known to have difficulty with sleep, nightmares, and variable sleep [7-10], attributes which are quantified here by the different sleep-related actigraphy metrics. The results provide further important insight into the quality of objective sleep in PTSD. The objective sleep differences are consistent with actigraphy studies that have found differences (such as increased wake after sleep) between people with PTSD and healthy controls [8,50] and between trauma survivors with and without PTSD [51]. However, objective sleep duration did not differ between groups, consistent with some [27], but not all previous actigraphy studies [50]. Therefore, this study adds to the tentative understanding of the nature of sleep disturbances in PTSD by means of the quantified disturbances in the reported patterns and highlights possible targets for potential intervention to further improve comorbid problems during therapy.

We remark that the results reported in Table 4 have not been corrected for multiple comparisons (eg, some researchers use the Bonferroni correction). These corrections are used to reduce type I errors (rejecting the null hypothesis when it is true) but have the important side effect of introducing type II errors (accepting the null hypothesis when it is false). This has urged researchers to suggest that effect sizes, correlation coefficients, and other metrics should be used to support research findings [51]. Here, we report the point biserial correlations for the extracted patterns, some of which appear to denote statistically strong associations ($|R|>0.3$).

The empirical rules and thresholds for the sleep detection algorithm were originally developed by the first author using his own Geneactiv data (n=1, collected for over 4 years) to correctly match onset, offset, and awakenings. From an algorithmic developer's perspective, using one's own data has the distinct advantage of cross-referencing activities, knowing the exact underlying ground truth across multiple days (including naps, awakenings, days of bad sleep, etc) and observing the recorded actigraphy signals. The thresholds of the algorithm were adjusted using a small subset (n<10) of the dataset described in this study, but we purposefully did not formally explore in detail optimizing thresholds to correctly match sleep diary entries to avoid overfitting. Ultimately, the sleep detection algorithm proposed in this study will need to be further validated in new cohorts to demonstrate how well it generalizes. The proposed sleep detection algorithm has functionality to detect awakenings, as illustrated in Figure 5. However, we do not have the self-reported onset and offset timings of awakenings; therefore, unfortunately, this is something we could not properly test and validate. Further work is required to demonstrate the potential of the proposed sleep detection algorithm in correctly detecting awakenings and comparing findings against self-reports or preferably PSG. Typically, algorithms developed to process actigraphy signals rely on hard thresholds (including the algorithm by van Hees et al [29]). It is possible that the thresholds chosen for the sleep detection algorithm proposed here might need to be refined in a different cohort (or with a different sensor recording actigraphy signals).

We believe that the colored actogram is a novel, convenient representation to understand longitudinal patterns of behavior. It serves as an intuitive summary of activity levels, potentially highlighting breaks in patterns. In principle, it could also be used as a composite for multiple participants, for example, to visually compare cohorts and identify whether at an aggregate level, there are some consistent patterns of behavior.

Comparison With Prior Work

The topic of automatically detecting sleep using wrist-worn devices has recently attracted attention [29]. However, the algorithms developed by van Hees et al [29] have not been validated in a group that exhibits considerable sleep problems [29], which is of particular importance in clinical assessment of mental disorders [12]. The new algorithmic approach proposed here to detect sleep has been developed with the aim of being sensitive to cohorts exhibiting perturbed sleep like PTSD and for extracting sleep patterns. In Multimedia Appendix 1, we provided indicative examples of this new approach, and in Table 3 and Figure 2, we demonstrate its competitiveness compared with the algorithm developed by van Hees et al [29]. Nevertheless, we remark that the sleep diaries used to assess sleep onset and wake-up times in this study are inherently limited and subject to participant self-report bias [20]. Therefore, although the current results are extremely promising, more rigorous validation is required to establish the validity of the new sleep detection algorithm against PSG and ideally in a larger sample, including diverse sleep pathologies. PSG provides the objective ground truth for sleep onset, sleep offset, and awakenings: the actigraphy-based patterns extracted computed by GGIR or any algorithmic package cannot be considered as PSG validated. Hence, the focus of using PSG is primarily to assess and compare the accuracy of competing sleep detection algorithms (ideally tested on both healthy controls and people

with different sleep disorders). The subsequent computation of sleep-based patterns is heavily dependent on accurate sleep estimates.

High-end equipment has been previously used to provide objective assessment of activity and sleep in mental disorders assessment, including PTSD [52]. PSG is the gold standard for sleep assessment and remains the only accurate approach to gain insight into the actual sleep architecture [32]. However, PSG relies on the use of expensive specialized equipment, and is logistically costly as assessment requires multiple hours of dedicated time by a sleep-certified expert [24]. Standard PSG also requires the participant's physical presence in the clinic; however, some recent studies have demonstrated the potential of home-based PSG recordings [53]. Similarly, the ActivPAL monitor is considered the gold standard in free-living conditions for step-counting, time spent in sitting/lying, standing, and walking postures [54]. However, it is designed to be worn at the thigh, fixed with a special adhesive pad, and hence is not very practical for long-term use. These difficulties motivate the use of alternative, easy-to-use, more affordable sensors, and wrist-worn watches are prime candidates in that respect. The data in this study were collected using the Geneactiv watch, which has attracted considerable interest across different research groups worldwide [36,42]. A similar popular wrist-worn device that was used, for example, in the UK Biobank study [25], Axivity AX3, has been shown to provide equivalent signal vector magnitude output with Geneactiv [55]; hence, the developed algorithms should, in principle, be directly applicable in studies where that device was used, crucially in the UK Biobank. Similar devices that provide access to the raw triaxial accelerometer data might require some cross-device calibration; otherwise, the developed framework should generalize well. Some older actigraphy devices used *counts* instead of raw acceleration signals and hence tools used to be device specific; although a relatively recent study provides for backward compatibility [42], the trend is moving toward tools that capitalize on the raw actigraphy data.

Summarizing the 3D accelerometry signals in a vector is crucial and is a required preprocessing step in advance of computing the actigraphy patterns (eg, IS, IV, etc). There are many different approaches reported in the research literature but no unique single best way to summarize the activity [38,39]. For example, the Euclidean Norm Minus One [38] is sensitive to calibration errors. Other approaches often rely on short-term windows [39] aiming to smoothen accelerometry fluctuations owing to internal accelerometer noise and hence might not effectively capture transient movements. The proposed approach in this study for the computation of the *movement* as an accelerometry summary, aims to address inherent accelerometer noise fluctuations by effectively operating on successive differences in the raw 3D accelerometer data before computing the Euclidean distance. We tentatively argue that this instantaneous-based approach rather than using local windows might have some advantages in terms of mitigating inherent accelerometer noise. Further work is required to assess whether there is any superior approach toward summarizing 3D accelerometry signals, for example, against a gold standard.

Alternatively, someone might use the subsequent computation of patterns (such as IS, IV, etc) working on each of the different accelerometry summaries (which are used as a preprocessing step) and assess how those patterns might be associated with a clinical outcome. This effectively draws parallels with the feature selection problem in data analytics, lacking ground truth of which are the *true features* and which are predictive of an outcome; researchers apply feature selection algorithms and feed different classifiers. On the basis of the classification performance, they can assess which feature selection algorithms perform best in a given problem [56,57]. We emphasize that for this approach to be valid and generalizable, researchers would need to perform comparisons across different datasets, ideally associating the extracted features with different outcomes. It is also possible that there are different combinations of accelerometry summaries and computation of patterns that work best; future work would be needed to investigate this in more detail.

Limitations

The primary limitations of this study are (1) the lack of PSG data to validate findings and (2) the study duration of 7 days. With long-term data, more detailed markers of weekly and monthly activity, sleep, and circadian rhythm variability could be developed and further explored. Participants adhered well and wearing the watch was not reported as disconcerting by any participant, which suggests that longer term monitoring may be viable in accordance with a recent study in the UK BioBank [25]. The sample size was sufficiently large for the exploratory aims of this study; nevertheless, larger cohorts might provide better insight into the nature of PTSD. Verification of the developed algorithm for sleep estimation (including awakenings) against PSG would be important in future studies to confirm sleep and wake detection, and further clarify the results. Similarly, we do not have detailed daily self-reported outcome measures (eg, daily mood self-reports as we had used longitudinally in related previous research [18,58,59]) other than the sleep diaries, which could have been associated with actigraphy-extracted patterns and hence further validate the developed algorithms.

The sleep detection algorithm proposed in this study capitalizes on the accelerometry and ambient light modalities; the latter is useful for differentiating sedentary activity and sleep. However, this suggests that the current version of the sleep detection algorithm would not be backwards compatible with devices that only record accelerometry signals. With the sophistication of wearables, additional modalities are becoming available (such as heart rate) and could be harnessed to potentially further improve sleep detection.

More generally, the light sensor modality should be used carefully in the analysis when developing algorithmic tools: lack of detected light does not necessarily indicate that someone is in a dark room. The wrist sensor may hide under a long sleeve, for example, in long-sleeved clothes or pajamas. There may also be abrupt changes in the detected light signal, if the sensor is temporarily visible (or vice-versa blocked).

The study participants wore the standard Geneactiv (Geneactiv Original). The temperature sensor is encased within the

waterproof housing of the watch; hence, the recorded temperature depends on how tightly the participants wear the watch: the sensor actually measures a mixture of body and room temperatures. In addition, during the night, temperature changes may (at least partly) indicate that a person has moved their hand above or under the blanket. The wrist temperature recorded has been primarily useful to detect nonwear times and to assess within-person changes, but not for between-person comparison. Researchers wishing to compare recorded temperatures between participants may want to explore a device that provides direct skin temperature recordings such as the Geneactiv Sleep variant [60]. Therefore, the use of the raw temperature measurements in this study should be interpreted very tentatively. For example, the increase in the recorded temperature during sleep seen in Figure 5 is likely a reflection of environmental temperature increase rather than an increase in the wrist temperature of the participant.

Conclusions

We envisage the developed algorithmic framework laying the foundation for using actigraphy analysis in different settings where raw wrist-worn triaxial accelerometer data are available, aiming to monitor healthy and pathological cohorts longitudinally. We encourage research colleagues to use and expand on the user-friendly MATLAB source code provided in this study to facilitate actigraphy data visualization and analysis. Among mental health conditions, applications to depression are of interest given that low levels of activity and poor sleep are characteristics of the disorder. Future research could investigate extracting additional patterns from the raw signals, potentially complementing it with additional modalities such as heart rate and geolocation, which are embedded in some recent devices. We are currently exploring the potential of using and extending the objective measures provided in this study to monitor longitudinal PTSD behavior, therapy effects, and long-term recovery.

Acknowledgments

The authors are grateful to all study participants for their meticulous adherence to the research protocol. The authors thank Juliane Sachschal for assisting with recruitment, interviews, and testing of the participants. They also thank Jennifer Wild, Emma-Warnock Parkes, and the staff of the Centre for Anxiety Disorders and Trauma, South London, and Maudsley NHS Foundation Trust for their help with recruitment. The study was funded by a Medical Research Council DTA award to EW and by Wellcome Trust grant 069777/200796 to AE. This study was supported by the National Institute for Health Research Oxford Health Biomedical Research Centre. AT was supported at the early stages of this work by the Wellcome Trust through a Centre grant no. 098461/Z/12/Z, *The University of Oxford Sleep and Circadian Neuroscience Institute (SCNi)* during the early stages of this work. The study was also supported by the Health Data Research UK, which receives funding from HDR UK Ltd (HDR-5012), funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), the British Heart Foundation, and the Wellcome Trust. The funders had no role in the study or the decision to submit this work to be considered for publication.

Authors' Contributions

AT developed the algorithms, analyzed the data, and wrote the first draft of the manuscript. EW designed the study with AE, wrote the protocol and ethics application, collected and entered the data, analyzed self-report data, and provided clinical insight for the interpretation of the activity data. AE codesigned the study and protocol, supervised recruitment, diagnostic interviews, and the running of the study. All authors critically reviewed and commented on the text and gave final approval for submission of this manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further details on algorithmic background and additional results.
[PDF File (Adobe PDF File), 1551 KB - mhealth_v8i4e14306_app1.pdf]

Multimedia Appendix 2
Statistical comparisons of features across cohorts.
[XLSX File (Microsoft Excel File), 23 KB - mhealth_v8i4e14306_app2.xlsx]

References

1. Alonso J, Angermeyer M, Lépine JP, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. The European study of the epidemiology of mental disorders (ESEMeD) project: an epidemiological basis for informing mental

health policies in Europe. Acta Psychiatr Scand Suppl 2004;109(420):5-7. [doi: 10.1111/j.1600-0047.2004.00325.x] [Medline: 15128382]
2. Bunting BP, Ferry FR, Murphy SD, O'Neill SM, Bolton D. Trauma associated with civil conflict and posttraumatic stress disorder: evidence from the Northern Ireland study of health and stress. J Trauma Stress 2013 Feb;26(1):134-141. [doi: 10.1002/jts.21766] [Medline: 23417880]
3. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the national comorbidity survey. Arch Gen Psychiatry 1995 Dec;52(12):1048-1060. [doi: 10.1001/archpsyc.1995.03950240066012] [Medline: 7492257]
4. Steel Z, Chey T, Silove D, Marnane C, Bryant RA, van Ommeren M. Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: a systematic review and meta-analysis. J Am Med Assoc 2009 Aug 5;302(5):537-549. [doi: 10.1001/jama.2009.1132] [Medline: 19654388]
5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Fifth Edition. Washington, DC: American Psychiatric Association Publication; 2013.
6. Ross RJ, Ball WA, Sullivan KA, Caroff SN. Sleep disturbance as the hallmark of posttraumatic stress disorder. Am J Psychiatry 1989 Jun;146(6):697-707. [doi: 10.1176/ajp.146.6.697] [Medline: 2658624]
7. Nappi C, Drummond SP, Hall JM. Treating nightmares and insomnia in posttraumatic stress disorder: a review of current evidence. Neuropharmacology 2012 Feb;62(2):576-585 [FREE Full text] [doi: 10.1016/j.neuropharm.2011.02.029] [Medline: 21396945]
8. Straus L, Drummond SP, Nappi CM, Jenkins MM, Norman SB. Sleep variability in military-related PTSD: a comparison to primary insomnia and healthy controls. J Trauma Stress 2015 Feb;28(1):8-16 [FREE Full text] [doi: 10.1002/jts.21982] [Medline: 25630526]
9. Leskin G, Woodward SH, Young HE, Sheikh JI. Effects of comorbid diagnoses on sleep disturbance in PTSD. J Psychiatr Res 2002;36(6):449-452. [doi: 10.1016/s0022-3956(02)00025-0] [Medline: 12393315]
10. Cox R, Olatunji BO. A systematic review of sleep disturbance in anxiety and related disorders. J Anxiety Disord 2016 Jan;37:104-129. [doi: 10.1016/j.janxdis.2015.12.001] [Medline: 26745517]
11. van Liempt S. Sleep disturbances and PTSD: a perpetual circle? Eur J Psychotraumatol 2012;3 [FREE Full text] [doi: 10.3402/ejpt.v3i0.19142] [Medline: 23050070]
12. Sheaves B, Porcheret K, Tsanas A, Espie CA, Foster RG, Freeman D, et al. Insomnia, nightmares, and chronotype as markers of risk for severe mental illness: results from a student population. Sleep 2016 Jan 1;39(1):173-181 [FREE Full text] [doi: 10.5665/sleep.5342] [Medline: 26350467]
13. Woodward S, Bliwise DL, Friedman MJ, Gusman DF. Subjective versus objective sleep in Vietnam combat veterans hospitalized for PTSD. J Trauma Stress 1996 Jan;9(1):137-143. [doi: 10.1007/bf02116839] [Medline: 8750457]
14. Germain A, Hall M, Shear MK, Nofzinger EA, Buysse DJ. Sleep disruption in PTSD: A pilot study with home-based polysomnography. Sleep Biol Rhythms 2006 Oct;4(3):286-289 [FREE Full text] [doi: 10.1111/j.1479-8425.2006.00230.x]
15. Herbst E, Metzler TJ, Lenoci M, McCaslin SE, Inslicht S, Marmar CR, et al. Adaptation effects to sleep studies in participants with and without chronic posttraumatic stress disorder. Psychophysiology 2010 Nov;47(6):1127-1133 [FREE Full text] [doi: 10.1111/j.1469-8986.2010.01030.x] [Medline: 20456661]
16. 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]
17. Schwartz S, Schultz S, Reider A, Saunders E. Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. J Affect Disord 2016 Feb;191:88-93 [FREE Full text] [doi: 10.1016/j.jad.2015.11.013] [Medline: 26655117]
18. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Osipov M, Clifford G, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. J Affect Disord 2016 Nov 15;205:225-233 [FREE Full text] [doi: 10.1016/j.jad.2016.06.065] [Medline: 27449555]
19. Carney C, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 1;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
20. Buysse D, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
21. Proudfoot J, Parker G, Pavlovic DH, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
22. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Goodwin G, de Vos M. Clinical insight into latent variables of psychiatric questionnaires for mood symptom self-assessment. JMIR Ment Health 2017 May 25;4(2):e15 [FREE Full text] [doi: 10.2196/mental.6917] [Medline: 28546141]

23. 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]
24. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 1;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
25. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
26. Kobayashi I, Lavela J, Mellman TA. Nocturnal autonomic balance and sleep in PTSD and resilience. J Trauma Stress 2014 Dec;27(6):712-716. [doi: 10.1002/jts.21973] [Medline: 25403523]
27. Kobayashi I, Huntley E, Lavela J, Mellman TA. Subjectively and objectively measured sleep with and without posttraumatic stress disorder and trauma exposure. Sleep 2012 Jul 1;35(7):957-965 [FREE Full text] [doi: 10.5665/sleep.1960] [Medline: 22754042]
28. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790 [FREE Full text] [doi: 10.1093/aje/kwt330] [Medline: 24500862]
29. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One 2015;10(11):e0142533 [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
30. Tsanas A. Oxford University Research Archive. University of Oxford, UK; 2012. Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning, PhD thesis URL: https://ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74 [accessed 2020-03-31]
31. Tsanas A, Clifford GD. Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing. Front Hum Neurosci 2015;9:181 [FREE Full text] [doi: 10.3389/fnhum.2015.00181] [Medline: 25926784]
32. Kryger T, Roth T, Dement WC. Principles and Practice of Sleep Medicine. Fifth Edition. Canada: Elsevier; 2010.
33. Weathers FW, Bovin MJ, Lee DJ, Sloan DM, Schnurr PP, Kaloupek DG, et al. The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): development and initial psychometric evaluation in military veterans. Psychol Assess 2018 Mar;30(3):383-395 [FREE Full text] [doi: 10.1037/pas0000486] [Medline: 28493729]
34. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
35. Aili K, Åström-Paulsson S, Stoetzer U, Svartengren M, Hillert L. Reliability of actigraphy and subjective sleep measurements in adults: the design of sleep assessments. J Clin Sleep Med 2017 Jan 15;13(1):39-47 [FREE Full text] [doi: 10.5664/jcsm.6384] [Medline: 27707448]
36. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol (1985) 2014 Oct 1;117(7):738-744 [FREE Full text] [doi: 10.1152/japplphysiol.00421.2014] [Medline: 25103964]
37. Zhou S, Hill RA, Morgan K, Stratton G, Gravenor MB, Bijlsma G, et al. Classification of accelerometer wear and non-wear events in seconds for monitoring free-living physical activity. BMJ Open 2015 May 11;5(5):e007447 [FREE Full text] [doi: 10.1136/bmjopen-2014-007447] [Medline: 25968000]
38. van Hees VT, Gorzelniak L, León ED, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
39. Karas M, Bai J, Strączkiewicz M, Harezlak J, Glynn NW, Harris T, et al. Accelerometry data in health research: challenges and opportunities. Stat Biosci 2019 Jul;11(2):210-237. [doi: 10.1007/s12561-018-9227-2] [Medline: 31762829]
40. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001 Sep;2(5):389-396. [doi: 10.1016/s1389-9457(00)00098-8] [Medline: 14592388]
41. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994 Apr;17(3):201-207. [doi: 10.1093/sleep/17.3.201] [Medline: 7939118]
42. te Lindert BH, van Someren EJ. Sleep estimates using microelectromechanical systems (MEMS). Sleep 2013 May 1;36(5):781-789 [FREE Full text] [doi: 10.5665/sleep.2648] [Medline: 23633761]
43. van Hees VT, Fang Z, Zhao JH, Heywood J, Mirkes E, Sabia S, et al. The Comprehensive R Archive Network. 2019 Dec 4. GGIR: Raw Accelerometer Data Analysis URL: https://cran.r-project.org/web/packages/GGIR/index.html [accessed 2020-04-03]
44. van Someren EJ, Kessler A, Mirmiran M, Swaab DF. Indirect bright light improves circadian rest-activity rhythm disturbances in demented patients. Biol Psychiatry 1997 May 1;41(9):955-963. [doi: 10.1016/S0006-3223(97)89928-3] [Medline: 9110101]

45. Di J, Spira A, Bai J, Urbanek J, Leroux A, Wu M, et al. Joint and individual representation of domains of physical activity, sleep, and circadian rhythmicity. Stat Biosci 2019 Apr 15;11(2):371-402. [doi: 10.1007/s12561-019-09236-4]
46. Benloucif S, Guico MJ, Reid KJ, Wolfe LF, L'hermite-Balériaux M, Zee PC. Stability of melatonin and temperature as circadian phase markers and their relation to sleep times in humans. J Biol Rhythms 2005 Apr;20(2):178-188. [doi: 10.1177/0748730404273983] [Medline: 15834114]
47. Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, et al. Psychological testing and psychological assessment. A review of evidence and issues. Am Psychol 2001 Feb;56(2):128-165. [doi: 10.1037/0003-066x.56.2.128] [Medline: 11279806]
48. Tsanas A, Little MA, McSharry PE. A methodology for the analysis of medical data. In: Sturmberg JP, Martin C, editors. Handbook of Systems and Complexity in Health. New York, USA: Springer; 2013:113-125.
49. The Data Analytics Research and Technology in Healthcare (DARTH group). 2019. Software URL: https://www.darth-group.com/software [accessed 2020-04-03]
50. Calhoun P, Wiley M, Dennis MF, Means MK, Edinger JD, Beckham JC. Objective evidence of sleep disturbance in women with posttraumatic stress disorder. J Trauma Stress 2007 Dec;20(6):1009-1018. [doi: 10.1002/jts.20255] [Medline: 18157880]
51. Nakagawa S. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav Ecol 2004;15(6):1044-1045. [doi: 10.1093/beheco/arh107]
52. Klein E, Koren D, Arnon I, Lavie P. Sleep complaints are not corroborated by objective sleep measures in post-traumatic stress disorder: a 1-year prospective study in survivors of motor vehicle crashes. J Sleep Res 2003 Mar;12(1):35-41 [FREE Full text] [doi: 10.1046/j.1365-2869.2003.00334.x] [Medline: 12603785]
53. Heinzer R, Vat S, Marques-Vidal P, Marti-Soler H, Andries D, Tobback N, et al. Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study. Lancet Respir Med 2015 Apr;3(4):310-318 [FREE Full text] [doi: 10.1016/S2213-2600(15)00043-0] [Medline: 25682233]
54. 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;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
55. Ladha C, Ladha K, Jackson D, Olivier P. Shaker Table Validation of OpenMovement AX3 Accelerometer. In: Proceedings of the Third International Conference on Ambulatory Monitoring of Physical Activity and Movement. 2013 Presented at: ISMPB'13; June 17-19, 2013; Amherst, MA URL: https://www.researchgate.net/publication/312500148_Shaker_table_validation_of_OpenMovement_accelerometer [doi: 10.13140/RG.2.2.31312.51208]
56. Guyon I, Eliseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157-1182 [FREE Full text]
57. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]
58. Palmius N, Tsanas A, Saunders KE, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017 Aug;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
59. Carr O, Saunders KE, Bilderbeck AC, Tsanas A, Palmius N, Geddes JR, et al. Desynchronization of diurnal rhythms in bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Apr 12;8(1):79 [FREE Full text] [doi: 10.1038/s41398-018-0125-7] [Medline: 29643339]
60. Activinsights. GENEActiv Sleep URL: https://www.activinsights.com/actigraphy/geneactiv-sleep/ [accessed 2020-04-03]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders 5th edition
NHS: National Health Service
PROM: patient-reported outcome measure
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
PTSD: posttraumatic stress disorder
WASO: wake after sleep onset

Edited by G Eysenbach; submitted 07.04.19; peer-reviewed by J Di, F Lamers, M Duncan, J Pärkkä; comments to author 19.05.19; revised version received 15.10.19; accepted 02.03.20; published 20.04.20.

Please cite as:
Tsanas A, Woodward E, Ehlers A
Objective Characterization of Activity, Sleep, and Circadian Rhythm Patterns Using a Wrist-Worn Actigraphy Sensor: Insights Into Posttraumatic Stress Disorder
JMIR Mhealth Uhealth 2020;8(4):e14306
URL: http://mhealth.jmir.org/2020/4/e14306/
doi:10.2196/14306
PMID:32310142

Original Paper

Comparing the Usability and Acceptability of Wearable Sensors Among Older Irish Adults in a Real-World Context: Observational Study

Alison Keogh[1], BSc, MSc, PhD; Jonas F Dorn[2], BSc, PhD; Lorcan Walsh[3], BSc, PhD; Francesc Calvo[2], BSc, MSc; Brian Caulfield[1], BSc, MSc, PhD

[1]Insight Centre for Data Analytics, UCD School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
[2]Data and Digital, Novartis Pharma AG, Basel, Switzerland
[3]Data Sciences, Novartis Business Services, Dublin, Ireland

Corresponding Author:
Alison Keogh, BSc, MSc, PhD
Insight Centre for Data Analytics
UCD School of Public Health, Physiotherapy and Sports Science
University College Dublin
O'Brien Centre for Science, Science Centre East
Belfield
Dublin, 4
Ireland
Phone: 353 863762075
Email: Alison.keogh@ucd.ie

Abstract

Background: Wearable devices are valuable assessment tools for patient outcomes in contexts such as clinical trials. To be successfully deployed, however, participants must be willing to wear them. Another concern is that usability studies are rarely published, often fail to test devices beyond 24 hours, and need to be repeated frequently to ensure that contemporary devices are assessed.

Objective: This study aimed to compare multiple wearable sensors in a real-world context to establish their usability within an older adult (>50 years) population.

Methods: Eight older adults wore seven devices for a minimum of 1 week each: Actigraph GT9x, Actibelt, Actiwatch, Biovotion, Hexoskin, Mc10 Biostamp_RC, and Wavelet. Usability was established through mixed methods using semistructured interviews and three questionnaires, namely, the Intrinsic Motivation Inventory (IMI), the System Usability Scale (SUS), and an acceptability questionnaire. Quantitative data were reported descriptively and qualitative data were analyzed using deductive content analysis. Data were then integrated using triangulation.

Results: Results demonstrated that no device was considered optimal as all scored below average in the SUS (median, IQR; min-max=57.5, 12.5; 47.5-63.8). Hexoskin was the lowest scored device based on the IMI (3.6; 3.4-4.5), while Biovotion, Actibelt, and Mc10 Biostamp_RC achieved the highest median results on the acceptability questionnaire (3.6 on a 6-point Likert scale). Qualitatively, participants were willing to accept less comfort, less device discretion, and high charging burdens if the devices were perceived as useful, namely through the provision of feedback for the user. Participants agreed that the *purpose of use* is a key enabler for long-term compliance. These views were particularly noted by those not currently wearing an activity-tracking device. Participants believed that wrist-worn sensors were the most versatile and easy to use, and therefore, the most suitable for long-term use. In particular, Actiwatch and Wavelet stood out for their comfort. The convergence of quantitative and qualitative data was demonstrated in the study.

Conclusions: Based on the results, the following context-specific recommendations can be made: (1) researchers should consider their device selection in relation to both individual and environmental factors, and not simply the primary outcome of the research study; (2) if researchers do not wish their participants to have access to feedback from the devices, then a simple, wrist-worn device that acts as a watch is preferable; (3) if feedback is allowed, then it should be made available to help participants remain engaged; this is likely to apply only to people without cognitive impairments; (4) battery life of 1 week should be considered as

a necessary feature to enhance data capture; (5) researchers should consider providing additional information about the purpose of devices to participants to support their continued use.

(JMIR Mhealth Uhealth 2020;8(4):e15704) doi:10.2196/15704

KEYWORDS

wearable technology; usability; mixed methods; user satisfaction

Introduction

Background

The technological advancements of recent years are challenging the traditional methods of data capture within clinical trials. In particular, the use of wearable technology offers unprecedented access to a variety of accurate, objective health care data that can be captured remotely, thus providing real-time access to large amounts of patient data [1,2]. Wearable devices are considered more convenient for participants by enabling them to collect data themselves, potentially resulting in improved protocol compliance and retention [3].

Given the relatively recent development of wearable devices, research has primarily focused on evaluating their clinical validity [4]. However, in order for these devices to be successfully incorporated into clinical trials, not only must they reliably capture accurate data, but critically, participants must be willing to wear and engage with them over a sustained period. The International Organization for Standardization defines usability as the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments [5]. To evaluate these components, researchers need to understand the barriers and facilitators to the participant's adherence with devices, to ensure that researchers do not inadvertently select clinically useful yet inappropriate devices, thus risking trial outcomes [6]. However, limited empirical evidence exists evaluating participant-centered usability of wearable devices within clinical trials [1], with wear-time and adherence rates used as proxy usability assessments. Furthermore, existing evaluations are limited by a focus on consumer-based products [7-10], short testing periods (ie, 24 hours or less) [11], the evaluation of a single wearable device only [12,13], and by the use of either qualitative or quantitative methods of data collection (but not both); thus, limiting the researchers full understanding of the participant's experiences [14].

Given the increasing prevalence of chronic conditions, clinical trials that focus on cohorts of older adults will be a key focus of future research. Older adults often report of requiring assistance with technology [10,15,16], making it important to investigate the experiences of older adults with various wearable devices, particularly in those which are intended for medical and research environments, to understand which devices participants prefer wearing, and whether any barriers to their use exist. In particular, it is important that industry partners and research groups, who plan to run clinical trials, test a variety of devices in real-life remote monitoring situations that mirror the contexts and environments in which trials may take place.

Objectives

Therefore, the primary aim of this study was to investigate the usability of a variety of wearable sensors in a real-world context by asking older adults to wear them in their home environment for a minimum of one week. Specifically, this was completed to establish the sensors' utility and usability, beyond data quality, from the participant's perspective and understand how these perceptions may affect their use in clinical trials.

Methods

Study Design and Participants

This was a six-week observational study that adopted mixed methods. No detailed inclusion or exclusion criteria existed; however, participants were required to be above 50 years of age, healthy, and fully independent in their daily lives. As this was an exploratory study, a power analysis was not undertaken. Eight participants from Dublin and the wider Wicklow and Kildare area, Ireland were recruited using purposive, convenience sampling through local flyers and existing connections between December 2017 and February 2018 to allow for comparisons of user experience, both between and within participants. Recruitment ceased once data saturation was reached in the qualitative analysis.

Included Devices

Seven, small, noninvasive wearable sensor devices, designed to track activity and sleep data were selected: Actigraph GT9X Link (Actigraph LLC), Actibelt (Trium), Actiwatch Spectrum Plus (Philips), Biovotion Everion (Biovotion), Hexoskin (Carre Technology), Mc10 Biostamp_RC (MC10 Inc), and Wavelet (Wavelet Health; Table 1). These specific devices were selected by the industry partners of this study who wished to assess the usability of devices that may be used to track physical activity in future clinical trials. Devices were selected to compare the range of locations and level of user interaction that are available on the market for this purpose.

Table 1. Basic functional and usability information regarding the devices included within the study.

Device (manufacturer)	Tethered to	Intended use	User app required	User interface	Medical grade[a]	Battery life[b]	Memory capacity
Actigraph GT9X Link (Actigraph LLC) [17]	Wrist	Sleep, actigraphy, and energy expenditure	Yes (optional)	Watch screen	Yes	1 week	4 GB
Actibelt (Trium) [18]	Waist (flex-belt or leather belt	Actigraphy	No	None	No	3 months	1800 GB
Actiwatch Spectrum Plus (Philips) [19]	Wrist	Sleep and actigraphy	No	Watch screen	Yes	1 week	1 MB
Biovotion Everion (Biovotion) [20]	Upper arm	Heart rate, respiratory rate, actigraphy, skin temperature, heart rate variability, and oxygen saturation	Yes	None	Yes	24 hours	Server-based memory, 3 days of data capture on device
Hexoskin (Carre Technology) [21]	Torso	Heart rate and actigraphy	No	None	No	>24 hours	600 hours
Mc10 Biostamp_RC (Mc10 Inc) [22]	Upper thorax[c]	Heart rate and actigraphy	No	None	Yes	2-5 days	Server-based memory, 3 days of data capture on device
Wavelet (Wavelet Health) [23]	Wrist	Sleep and actigraphy	Yes	None	No	24-36 hours	Not reported

[a]Defined by manufacturers according to the Food and Drug Administration and European guidelines.

[b]As reported by the device manufacturer.

[c]In this study only, other attachment points exist.

Study Procedure

At the entry point to the study, participants provided written informed consent, after which an opening interview was undertaken to establish their views on wearable technology in health and their previous experiences with wearable devices. Participants were then provided with a device and instructed to wear the device at all times (if possible, during their normal activities, except showering, for the duration of the week). Devices were worn for a full seven days each. The order of the devices was randomized to minimize bias. Depending on the device, participants were not required to interact with the device other than to charge them, if the device required. A week after the first testing session, participants returned their device and were provided with a new sensor. Participants were asked to complete three validated outcome measures (as described below); while semistructured interviews were completed at the end of each deployment week, so that feedback was provided specifically for each device independently. Upon completion of the study, participants completed a final semistructured interview, wherein they were asked about their overall perceptions of the included sensors within the study and which devices they preferred and why. Device deployment was randomized to limit the risk of bias.

Data Collection and Outcome Measures

Quantitative Data Collection

Brief demographics of the participants were collected (ie, sex, age, height, weight, and any previous experience with sensors). In total, three questionnaires were given to each participant regarding each of the sensors.

- *The Systems Usability Scale (SUS)*: It measures the usability of a device/system/technology [24-26]. It consists of a 10-item questionnaire with five response options for respondents from *1: strongly disagree* to *5: strongly agree*, resulting in a potential minimum score of 0 and a maximum of 100.
- *Intrinsic Motivation Inventory (IMI)*: IMI is a multidimensional questionnaire intended to assess the participant's experiences related to a target activity [27], in this case, wearing the wearable device. The instrument contains 22 items on a 7-point Likert scale, ranging from *1: not at all true* to *7: very true*. The measure assesses six subscales: *interest/enjoyment*, *perceived competence*, *effort/importance*, *pressure/tension*, *value/usefulness*, and *perceived choice*.
- *Acceptability questionnaire by Jacucci et al* [28]: Jacucci et al [28] aimed to assess users' acceptance of wearable devices across dimensions including comfort, fear of technology, and privacy. Participants were asked to rate the extent to which they agreed or disagreed with each of the 26-item statements on a 6-point Likert scale ranging from *1: completely disagree* to *6: completely agree*, on 10 individual subsections.

Qualitative Data Collection

The aim of the qualitative phase was to explore the participant's opinions of the devices and the factors they felt influenced their

use of the same (interview guide provided in Multimedia Appendix 1). A female research physiotherapist (AK) with a PhD in behavior change (including two years of experience and training in qualitative research) and currently working in the area of digital health completed the semistructured interviews to extract more information from participants about certain aspects of the design or usability of the device. Interviews were completed in either participants' homes or place of work, depending on their preference. Scratch notes were taken by AK during the interviews, which were also audio-recorded and transcribed verbatim by AK. As the sample was purposively gathered, some participants were known to the researcher and thus, a rapport was already established. Participants were aware of the purpose of the research through the participant information leaflet and consent form they signed before participating. Before completing the research, AK had pilot tested each device to ensure they were set up correctly; thus, she witnessed experiences of some of the potential barriers and facilitators to their use.

Data Analysis

Quantitative Data Analysis

The SUS score was computed for each participant following standard scoring methodology [24]. Descriptive statistics were calculated to find out the median (IQR; min-max) result per device. To score the IMI, all negatively worded statements were inversely translated by subtracting the participant's score from eight. Following this, the average score for each of the six categories was calculated for each participant and group median (IQR; min-max) scores were calculated for each category for each device independently. A median result for the acceptability questionnaire was calculated per device, alongside a median result for each of its 10 subsections independently. In the absence of reference interpretations of the IMI and acceptability questionnaire, the midpoint of Likert scale was selected as the minimum level of acceptability of a device [29-32].

Qualitative Data Analysis

Deductive content analysis was undertaken for each of the transcribed texts using a realist approach, whereby the researcher assumed that the opinions of the participants reflected their true perceptions and should be taken as real [33]. A deductive content analysis was undertaken to categorize the participant's responses based on previous knowledge [34]. Specifically, literature has suggested that perceived usefulness, comfort, and ease of use are critical factors of usability [10,35-37], thus, these were selected as the categories for which the content of the transcribed audio recordings would be assessed. In addition, because the research question focused on understanding whether participants would accept using these devices within a clinical trial, this was pragmatically selected as an additional category. Following the steps outlined in previous research [35], the researcher (AK) familiarized herself with the texts and then identified the content which corresponded with each of the preidentified categories [34,35]. Data saturation was deemed to have occurred when no additional learnings regarding the devices and their features were identified under the selected categories. This analysis was then discussed with another member of the research team (BR), who was experienced in qualitative research, to ensure accuracy in coding. Specific quotations, which were deemed to represent the most important aspects of participants' experiences were selected for inclusion by AK and BR. Participant checking did not take place as part of this study, and transcripts were not provided to the participants.

Data Integration

A triangulation design was completed at the interpretation level of data analysis to provide a more complete picture of each device, to enhance the reliability of the study, and to support data saturation [38]. Specifically, a meta-matrix was created to facilitate comparisons of the results by presenting the quantitative data in tabular format alongside the summarized qualitative themes. For each sensor independently, all results were displayed on the same page, to determine whether there was convergence, partial convergence, discrepancy, or silence [39-42].

Ethics Approval and Consent to Participate

This study received ethical approval from the University College Dublin Human Ethics Committee (ref: LS-17-92-Caulfield). All participants provided written informed consent.

Results

Demographic Information

Participant demographic information can be found in Table 2. Six participants reported feeling comfortable or very comfortable using technology. Three were wearing an activity tracker, while the remaining three had worn them in the past. The final two participants rated their technology comfort levels as medium, with no previous experiences of using wearable devices. All participants wore each of the seven devices, with the exception of Hexoskin. The reasons for which are outlined within the results. In addition, all participants reported wearing the devices at all times during the week, with the exception of Hexoskin. However, no formal assessment of adherence was completed.

Table 2. Participant demographic information.

Characteristic	Value
Gender (n)	
Male	5
Female	3
Age (years), mean (range)	62 (53-72)
Level of education (n)	
Third level	3
Secondary level	4
Primary level	1
Employment status (n)	
Retired	4
Employed	4
Experience with wearable devices (n)	
Yes (current or past)	6
No	2

Quantitative Results

System Usability Scale

The median score for all devices on the SUS was 57.5 (IQR 12.5; min-max=47.5-63.8) out of a possible score of 100. None of the tested devices were deemed to be *good* by participants, as all seven achieved scores of less than 68 (30). Actibelt achieved the highest median result of 63.8 (IQR 12.5; min-max=47.5-67.5), while Hexoskin achieved the lowest median result of 47.5; min-max=37.5- 57.5 (Table 3). The results for all of the devices fall between the 10th and the 30th percentile, meaning that all were considered below average [24].

Intrinsic Motivation Inventory

The median score for all devices on the IMI was 4.6 (1.0; 3.6-5.2) on the 7-point Likert scale. No device achieved very high results (Table 3). Hexoskin was the only device to score below the midpoint of Likert scale (3.6; 3.4-4.5), suggesting that participants would not be autonomously motivated to wear this device.

Acceptability Questionnaire

The median score for all devices on the acceptability questionnaire was 3.5 (0.5; 3.2-3.6). The highest median results were achieved by Biovotion, Actibelt, and Mc10 Biostamp_RC, with each achieving results of 3.6 on the 6-point Likert scale (Table 3).

Table 3. Participants' self-reported usability of each device according to (1) Intrinsic Motivation Inventory, (2) System Usability Scale, and (3) Acceptability questionnaire.

Questionnaire domains	Actigraph, median (IQR); min-max	Actibelt, median (IQR); min-max	Actiwatch, median (IQR); min-max	Biovotion, median (IQR); min-max	Hexoskin[a], median; min-max	Mc10, median (IQR); min-max	Wavelet, median (IQR); min-max
Intrinsic Motivation Inventory (n=22 questions; 7-point Likert scale)							
Median	4.3 (0.8); 3.9-5.4	4.1 (0.9); 3.3-5.1	4.7 (1.1); 3.0-5.4	5.2 (0.3); 4.0-5.5	3.6; 3.4-4.5	4.5 (1.1); 1.7-5.6	4.7 (0.8); 4.3-5.1
Interest	3.5 (1.4); 2.3-5.3	3.4 (1.3); 2.8-5.0	4.5 (1.8); 1.5-5.5	6.0 (1.0); 2.5-7.0	3.5; 3.5-4.3	3.5 (1.0); 1.0-4.3	5.3 (0.6); 4.7-7.0
Competence	6.7 (3.2); 2.7-7.0	6.2 (1.4); 5.3-7.0	6.3 (2.0); 4.7-7.0	6.5 (1.9); 3.4-7.0	4.3; 4.0-4.3	5.0 (1.7); 3.6-7.0	6.7 (0.8); 3.0-7.0
Effort	3.3 (2.9); 2.0-5.8	3.8 (2.5); 2.3-5.5	3.5 (3.0); 2.5-6.3	3.9 (1.6); 2.0-5.8	3.5; 2.0-3.8	4.3 (2.8); 1.8-6.8	4.0 (1.3); 1.0-5.5
Pressure	1.3 (2.0); 1.0-3.3	1.0 (0.3); 1.0-1.3	1.0 (2.0); 1.0-3.7	1.8 (1.7); 1.0-5.0	3.3; 2.0-3.7	3.0 (3.0); 1.0-4.0	2.0 (2.8); 1.0-4.0
Choice	6.9 (0.9); 6.0-7.0	6.9 (1.4); 3.0-7.0	7.0 (1.0); 5.5-7.0	6.8 (1.5); 5.3-7.0	4.3; 4.0-7.0	6.8 (1.8); 1.5-7.0	4.0 (0.0); 3.3-4.0
Usefulness	4.9 (2.5); 3.0-5.5	3.9 (2.1); 1.8-5.5	5.5 (3.0); 1.0-7.0	6.1 (1.7); 4.0-7.0	3.0; 1.8-3.3	5.0 (2.3); 1.0-6.8	6.8 (0.9); 5.0-7.0
System Usability Scale (n=10 questions; 5-point Likert scale, score out of 100)							
Total score	60.0 (15.6); 50.0-67.5	63.8 (12.5); 47.5-67.5	57.5 (15.0); 50.0-65.0	56.6 (13.1); 45.0- 70.0	47.5; 37.5- 57.5	55.0 (12.5); 45.0-65.0	56.3 (9.4); 50.0-62.5
Acceptability questionnaire (n=26 questions; 6-point Likert scale)							
Median score	3.6 (0.9); 2.8-5.2	3.4 (1.0); 2.8-4.7	3.2 (0.8); 3.0-4.0	3.6 (0.6); 3.0-4.8	3.2; 3.0-3.5	3.6 (0.4); 3.0-3.9	3.5 (0.4); 3.2-4.0
Attitude	5.3 (1.6); 3.7-6.0	5.2 (1.1); 4.7-6.0	4.3 (1.0); 4.0-6.0	4.7 (1.8); 3.3-6.0	4.0; 3.3-4.7	4.3 (1.7); 4.0-5.7	4.3 (1.6); 3.7-6.0
Anxiety	1.8 (2.5); 1.0-5.3	1.8 (2.6); 1.0-5.3	2.7 (1.7); 1.0-3.0	2.5 (2.9); 1.0-4.3	2.7; 2.3-2.7	3.0 (1.0); 2.3-3.7	2.3 (1.8); 1.0-5.0
Facilitating conditions	2.5 (4.8); 1.0-6.0	2.5 (2.6); 1.0-4.0	1.5 (1.0); 1.0-3.0	2.5 (2.3); 1.0-3.5	5.5; 3.5-6.0	3.0 (4.5); 1.0-6.0	1.5 (1.0); 1.0-2.5
Perceived usefulness	4.5 (2.7); 3.3-6.0	3.5 (2.3); 1.0-6.0	4.3 (2.3); 1.0-6.0	4.8 (1.5); 4.0-6.0	2.3; 1.3-3.0	3.3 (2.7); 1.0-6.0	5.2 (1.0); 4.0-6.0
Perceived effort	3.8 (3.0); 3.0-6.0	3.8 (1.5); 3.5-5.0	3.5 (0.0); 3.5-4.0	4.5 (2.3); 3.0-6.0	5.0; 3.5-5.0	3.5 (1.0); 3.0-5.5	3.5 (0.5); 3.0-6.0
Behavioral intentions	3.5 (1.4); 1.0-6.0	3.0 (1.2); 1.7-4.3	3.7 (1.0); 2.3-6.0	3.8 (0.8); 3.3-4.3	2.7; 2.7-3.0	3.0 (1.7); 2.7-4.7	3.8 (1.2); 3.0-4.3
Psychological attachments	3.8 (2.1); 1.5-6.0	3.8 (2.4); 1.0-6.0	4.5 (3.0); 3.0-6.0	4.5 (1.8); 2.5-6.0	2.5; 1.5-3.5	3.0 (2.0); 1.0-6.0	4.0 (1.5); 1.5-5.0
Privacy	2.5 (1.4); 1.0-5.0	3.0 (2.4); 1.0-6.0	2.5 (1.0); 1.0-5.0	3.3 (1.9); 1.0-4.5	2.5; 2.5-3.0	3.0 (3.0); 1.0-6.0	2.8 (1.6); 1.0-4.0
Enjoyment	3.7 (1.5); 2.7-4.7	4.0 (1.1); 2.0-4.3	3.0 (1.0); 2.7-3.7	2.7 (1.1); 2.0-4.3	3.7; 3.0-4.3	3.7 (1.7); 1.7-5.0	2.7 (0.3); 1.0-3.0
Comfort	3.3 (1.3); 2.0-5.3	4.3 (1.8); 2.3-4.3	4.0 (1.3); 2.7-4.3	4.0 (0.5); 3.7-4.7	3.0; 2.3-4.0	2.7 (1.3); 2.0-4.3	4.3 (0.7); 2.7-5.0

[a]n=3 participants. Hexoskin was removed from the study after receiving the feedback from the first three participants to use it. The burden they reported was considered too high to ask any remaining participants to use it. Therefore, no IQR exists.

Qualitative Results

Interviews per device ranged from 10-21 min in length. Exit interviews at the end of the study ranged from 18 to 38 min in length. The findings for each device under the headings of *comfort of device*, *perceived usefulness of device*, *ease of use of device*, and *likelihood of wearing a device* are provided throughout the results with supporting quotations (participant numbers listed in parentheses).

Comfort of Devices

Participants believed that wrist-worn sensors were the most versatile and easy to use, and therefore, the most suitable for long-term use. In particular, Actiwatch and Wavelet stood out for their comfort. Wavelet, in particular, was remarked to be similar in design to *Fitbit*, resulting in its acceptability. However, the clasp method of closing the watch was not secure unless carefully completed, resulting in one participant losing a device. Actigraph was the only watch-based device that received negative feedback under the heading of comfort. The bulkiness of the device, perceived outdated design, and the frequency with which it *snagged* in participants' clothes were the reasons for negative feedback.

Actibelt was perceived as surprisingly comfortable by all participants who expected it to be more cumbersome than it was. In contrast, Mc10 Biostamp_RC was notable for its lack of comfort. It was considered itchy. Participants noted that they were aware of Mc10 Biostamp_RC's potential to fall off, while female participants were aware that the device was visible underneath certain clothing:

> *I just thought the most convenient and simplest one was the Wavelet. Well it was small, it was unobtrusive, it was a good design, it wasn't as bulky as the Actigraph and it just looked like a normal kind of Fitbit.* [101, male, age 64 years, employed]

> *The ideal device is in a watch form because they are the easiest thing to wear, the ones that don't interfere with day to day activities as much and they don't interfere with what clothing you're wearing, unless they're very bulky.* [401, female, age 56 years, employed]

Ease of Use of Devices

The devices that required little to no interaction from participants were considered the easiest to use (ie, Actibelt, Actiwatch, and Actigraph GT9X Link). Although Mc10 Biostamp_RC did not require participants to engage with it, once it was on, participants were required to change the adhesive stickers every 1-3 days, resulting in the uncertainty and concern about the accuracy of their replacements; thus, the accuracy of the data provided by the device. In response, participants used the red marks on their skin left by the devices as guides to help them:

> *Oh I didn't like the stamps [Mc10]...Well they were a bit fiddly to put on in the first place. They had the gel and it was hard to quite know the exact place to put them on, and then they can come off quite easily and then you have to put them back on...and then you have to take them on and off when you are having your shower, so they were almost completely impractical, certainly from a long-term point of view, you couldn't do that for more than a couple of days.* [101, male, age 64 years, employed]

Wavelet and Biovotion provided participants with feedback through a mobile phone app, which was also the method required to monitor the battery level of the devices. For most participants this was not problematic, as the feedback provided by the device was interesting; therefore, engaging with the app was not a burden. However, the majority of participants agreed that long battery life was essential for long-term use of wearable devices, with a minimum of one week considered ideal. The need to charge a device daily was deemed unacceptable. Thus, this was a barrier to the sustained use of both Wavelet and Biovotion. One participant forgot to check the battery levels and as a result, missed the data collection of a number of days. In addition, Wavelet required users to select within the app, when they would go to sleep, resulting in an additional task, which was again, often forgotten:

> *Now perhaps if you have it for a long time you just purely get into the habit of doing it but it was very easy to forget because you know there's I suppose, bed time you should get into procedures because I do, I remember to charge things to do stuff. If there was something that was on the device itself even if it was a little button that says sleep.* [601, male, age 52 years, employed]

The devices most difficult to use were also those that were the least favored. Specifically, Hexoskin was considered as an excessive burden on participants, as it required users to moisten the chest sensors within the vest frequently (every 15-20 min) to capture the heart and breathing rate data accurately. This was deemed impractical and disruptive to activities of daily living; therefore, a decision was made to cease the testing of the device, following the feedback from first three participants:

> *If you look at something like the vest [Hexoskin], which was very irritating that you had to keep wetting the sensors…I'd wear it for 24 hours but it's not something that I would wear for a week and I certainly wouldn't wear it for six weeks...No matter the feedback…because it's just too limiting in your day to day activity…having to reach around under your breasts to find this piece of cloth that's a sensor and then wet it is not something you can do easily in a public place.* [401, female, age 56 years, employed]

Perceived Usefulness of Devices

For the majority of participants, the best devices were those they felt they received the most feedback from (ie, Wavelet, Biovotion). Indeed, participants seemed willing to compromise on *small annoyances* if they were personally getting something from the device. The devices with little to no feedback were not perceived as useful, with some participants appearing indifferent to the devices owing to this reason (ie, Actibelt, Actiwatch GT9X Link, and Mc10 Biostamp_RC). Nonetheless, participants were able to understand how these devices may still be valuable to others, including clinicians and researchers, and thus, were prepared to wear these devices *in the name of science*:

> *Well, because there was no feedback, it [Actiwatch] was pointless to me but in fairness to it was absolutely no trouble at all, you just forget it's there, its design is better [than the Actigraph]. As you can see, I'm wearing it on a wrist with another watch and it just wasn't an issue at all …it played no part in my life at all…first of all it's just one piece, it's got, even though it not much of a beveled edge, it's got enough that things won't snag on it as much. I do find it just sits better on the wrist the strap seems to be softer, more malleable.* [601, male, age 52 years, employed]

> *It's there and it has no function [Actibelt]. There's no feedback, there's no information, there's no feedback telling you what's happening.* [301, female, age 62 years, employed]

Wavelet was reported to be the most useful device by participants who valued the simple graphs provided within the app (ie, sleep and heart rate). Actigraph GT9X Link was initially considered very basic, as the only information it provided was step count. Although, the participants did become accustomed to being able to easily check their step count throughout the day. Finally, even though Biovotion provided participants with innovative feedback (data were presented in an integrated spiral depicting a full day of information within a clock), the potential usefulness of future iterations of the device was greater than its current version. In particular, participants desired numerical data in addition to the spiral graph, to help them understand normal reference values. The suggestion by one participant that the device was *ahead of its time* is important, as it suggests that Biovotion is a promising product (dependent on future iterations) that may have a strong role to play in the monitoring of patient health:

> *I think it [Biovotion] was meant to measure things like your peripheral circulation or something, but again it gives you a number, it doesn't tell you whether that means that your peripheral circulation is good, bad or indifferent…otherwise it's just like a gimmick, it's there you've got this little spiral that's colourful, bit entertaining to look at…but you don't get a chart to show what it was at various times during the day unless you just interpret what the spiral is showing,…all you get is real-time readings…it seems to be like the ultra-high definition televisions when they came out, they were fantastic, they looked wonderful but you couldn't get ultra-high definition programs, so basically the televisions were head of its time. In a sense I think then maybe that this device is ahead of its time.* [401, female, age 56 years, employed]

Likelihood of Wearing a Device During a Trial

Participants agreed that the *purpose of use* is a key enabler for long-term compliance. These views were particularly noted by those not currently wearing an activity-tracking device. Although these participants explained that they did not personally feel the need to track their own activities, they suggested that they would not object to wearing a device for longer periods (ie, 8-12 weeks). For instance, in situations if they had to (ie, in the context of a clinical trial or by a clinician) and if the device was reasonably comfortable and easy to use. For most devices, participants reported that they would only wear them only if it was necessary, suggesting that their use of these devices would be born out of compliance rather than a specific, intrinsically motivated intention:

> *I would find it bothersome [having to wear the Mc10 within a trial]…I would be willing to do it you know because I think it's good, but I was actually glad that today was the last day of these.* [701, female, age 63 years, retired]

> *I would do it for the sake of science, and for this, but I certainly wouldn't, under no circumstances would I purchase it or use it kind of on an ongoing basis.* [601, male, age 52 years, employed]

Integrated Results

Convergence was predominantly seen across each of the devices independently across the four headings: comfort, ease of use, usefulness, and likelihood of wearing the device. Specifically, an agreement could be observed between the qualitative and quantitative results overall; thus, providing support for each of the results. Table 4 provides a sample of this matrix, specifically for Actigraph GT9X Link. A full list of results for each individual sensor is available within Multimedia Appendix 2.

Table 4. Matrix of integrated qualitative and quantitative data for Actigraph GT9X Link (this device was used as an example).

Outcome of interest	Quantitative result, median (IQR); min-max	Qualitative result	Convergence; discrepancy; silence
Comfort	• Midpoint of the Likert scale for perceived comfort (acceptability questionnaire): 3.3 (1.3); 2.0-5.3	• Somewhat comfortable • Unanimously agreed that the device was too big • For some, along with excessive strap length, the device irritated them to the point of being uncomfortable • Others felt that despite the size, the device was nonetheless comfortable	Convergence
Perceived usefulness	• Midpoint for interest (IMI[a]): 3.5 (1.4); 2.3-5.3 • Midpoint for usefulness (IMI): 4.9 (2.5); 3.0-5.5 • Midpoint for effort/importance (IMI): 3.3 (2.9); 2.0-5.8 • OK usability (SUS[b]): 60.0 (15.6); 50.0-67.5 • High perceived usefulness (acceptability questionnaire): 4.5 (2.7); 3.3-6.0 • Midpoint enjoyment (acceptability questionnaire): 3.7 (1.5); 2.7-4.7	• Step count was both interesting and useful • Further feedback was desired • Device was considered *boring* due to its limited functionality • Dual function as a watch appreciated	Convergence
Ease of use	• High perceptions of competence (IMI): 6.7 (3.2); 2.7-7.0 • Midpoint for perceived effort (acceptability questionnaire): 3.8 (3.0); 3.0-6.0 • Midpoint for effort/importance (IMI): 3.3 (2.9); 2.0-5.8	• Participants felt that the device was simple to use, as there was little to no interaction required with it • Limited difficulties reported	Partial convergence
Likelihood of wearing a device	• Low pressure to wear (IMI): 1.3 (2.0); 1.0-3.3 • High perceived choice (IMI): 6.9 (0.9); 6.0-7.0 • Midpoint behavioral intentions (acceptability questionnaire): 3.5 (1.4); 1.0-6.0 • Midpoint psychological attachments (acceptability questionnaire): 3.8 (2.1); 1.5-6.0 • Low facilitating conditions (acceptability questionnaire): 2.5 (4.8); 1.0-6.0	• Participants were unclear whether this was a device suitable for long-term use • The limited functionality is a plus for some and a barrier to others • Almost everyone willing to wear it *for science* or if instructed by a health care professional • Outside of a trial, the device was considered too bulky for long-term use • Participants became used to it as the trial progressed; with many preferring it to other tested devices	Partial convergence

[a]IMI: Intrinsic Motivation Inventory.

[b]SUS: System Usability Scale.

Discussion

Principal Findings

This study aimed to investigate the usability of multiple wearables sensors within a real-world context by focusing on the human factors associated with their use in a group of older adults. This aim was achieved using mixed methods to determine participants' likeliness to use and compliance with each device during a clinical trial; as judged through a week's worth of constant wear. The results of this study further demonstrate the complexity involved in selecting a wearable device, as none of the tested sensors were considered optimal due to the influence of a variety of factors, including the feedback provided by the devices, their comfort, and their battery life.

Comparison With Prior Work

A key strength of this study was the comparison of multiple devices within the same cohort of participants, thus offering an opportunity to accurately compare one device to another in the context of participants' daily lives. The benefit of this multi-sensor approach, compared with other studies [11,43,44] was that within and between participant assessment of numerous devices, all with varying features and locations, our study allowed participants to note barriers that otherwise may not have been remarked without this easy and swift comparison. For example, Biovotion and Actibelt were noted for how little they interfered with activities of daily living, despite the initial expectation that they would be a burden. Furthermore, findings were strengthened by the use of mixed methods as the integrated findings typically converged; thus, demonstrating the robustness of the results. Although quantitative comparisons alone failed to provide a detailed understanding of why devices may differ, qualitative research does not always allow for generalizability. Integrating the two approaches provided a deeper understanding and comparison of what participants prioritized and favored within devices.

All devices in this study achieved SUS scores below average [45], suggesting they are only marginally usable. However, due

to the small sample size in this study, these results should be interpreted with caution, as they cannot be generalizable to the wider population. In addition, the participants in this study were familiar with technology, which may limit direct comparisons with other research. Nonetheless, the quantitative results may provide some useful insights regarding the potential for these devices to be used in clinical trials. Specifically, low scores in the SUS are common, even among popular consumer devices including Fitbit [46]. A trade-off between comfort and functionality appears to exist, whereby participants are willing to accept a slightly less comfortable device, provided it serves a purpose that they value [47]. This is evidenced by participants consistently repeating that they would accept small annoyances for a device they perceived as beneficial. Indeed, it has been suggested that the "function of any wearable tool must outweigh any physical or social discomfort felt in wearing it, and less desirable devices may meet with higher standards for comfort and fit." This finding echoes recent studies wherein participants were most likely to purchase and recommend devices based on their features, battery life, ease of use, and reliability [46,48-51]. Specifically, in relation to older adults, this study repeated the findings of previous research in that devices, which were deemed to be comfortable, fit seamlessly into daily routines, and demonstrated a clear perceived benefit to the participants were the devices that were favored [12,52]. Participants in this study consistently listed Wavelet and Biovotion as their preferred devices owing to the combination of useful feedback, comfort, and seamless interaction with their daily lives. However, the ability of participants to easily check the battery level of devices is a necessity, especially within a clinical trial wherein consistent data collection is paramount. Even though perceived usefulness and perceived ease of use are critical components for participants' intention to use a wearable device [10,37], both Wavelet and Biovotion may be limited in the sense that their battery level needs to be regularly monitored by users.

Interestingly, participants have been shown to consistently select a favorite device, irrespective of the evidence they gather to refute this. This was mirrored in this study as participants overwhelmingly agreed that Actibelt was one of the most comfortable, least obtrusive devices, had the longest battery life, and yet consistently failed to list it as a favorite. The perceived importance of feedback is likely to be the sole reason for this discrepancy, therefore, highlighting one of the most important findings of this research: for participants to be motivated to wear a device, they must see a purpose for it. For example, Actibelt and Actiwatch were very comfortable to all participants; however, neither device provided feedback. As participants were not confident whether they understood what data were being collected, the devices were not considered useful by the participants. In contrast, Actigraph GT9X Link was cumbersome and bulky, yet its simple feedback made it a device that participants appreciated.

When the results of this study are combined with previous research [10,12,46], it is clear that participants in multiple cohorts, both healthy and clinical, are broadly accepting of wearable technology, and once they can see the use of a relatively comfortable devices, they will be willing to wear them. However, one important insight that needs to be considered by both researchers and device manufacturers alike; participants are often able to see the future capability of wearable technology beyond its current function and are often left disappointed by the realities of a device when compared with the potential (eg, the measurement of blood pressure with Biovotion). Thus, research investigating the usability of wearable devices is consistently strengthening the argument that user-centered design is critical for compliance, and that users must gain some sort of advantage from wearing these devices. For most users, this is gained through the provision of feedback. Although, it remains unclear as to what level of feedback is considered necessary by participants, especially within cohorts with cognitive impairments. Given that many medical devices are not routinely designed to provide feedback, the result of this is a clash between health and consumer attributes in cohorts that desire and can cognitively interpret it [51]. Indeed, a common research hypothesis is that wearable devices may alter clinical trial outcomes because of real-time metrics and the ability of users to self-monitor their behavior [53]. However, sustained and meaningful behavior change has yet to be consistently demonstrated through consumer-based wearables alone [36,37,46,53]. Therefore, it should be considered whether feedback is a tangible risk to clinical trial outcomes. If it is not a risk, the provision of feedback may be one of the most important variables to consider when selecting a device for users without a cognitive impairment, as its presence provides participants with a perceived value for the device, which may support enhanced compliance. In response to this, researchers need to consider whether they can select a device that provides participants with some form of feedback (eg, heart rate), while remaining blind to the primary outcome measure of the trial (eg, physical activity). This is in regard to the acknowledgement that the future device development needs to incorporate desired participant functions to enhance compliance.

Limitations

The results of this study should be considered alongside its limitations. Firstly, the findings cannot be generalized to the wider population due to the small number of participants, specifically older adults, many of whom were comfortable with technology. Thus, the findings of this study cannot be widely generalizable. However, as technology becomes more pervasive, older participants will become accustomed to its use, and thus, understanding the experiences of those who are comfortable with technology is nonetheless useful. Indeed almost 80% of older adults in one study reported using some form of technology in their lives [13]; however, it must be acknowledged that the experiences of people in their mid-60s cannot be compared with those in their 70s or above [13]. Additionally, although eight participants is a small number, participants acted as their own controls by comparing the use of multiple devices, thus, providing valuable within-study comparisons. Furthermore, the clinical utility and accuracy of these devices was not evaluated as part of this study. However, since this study commenced, some manufacturers have, or are about to release new versions of these devices on the market (eg, Actigraph). In addition, no formal measure of wear-time was collected within this study. Therefore, the results rely on participants' self-report of whether they used the device or not. However, given that the

focus of this study was on the usability of the device, compliance was not considered an important quantitative variable. For instance, in the case of Hexoskin, participants made it clear that they would not comply, and did not continue to wear the device due to its lack of usability. Given the aim of this study, this qualitative finding was more valuable than a quantitative measure of compliance as they highlighted the reasons why compliance was poor rather than simply whether it was or not. Finally, the result for Mc10 Biostamp_RC are likely to have been negatively influenced by the placement of the sensors on the pectoral muscles of participants, while Hexoskin is not intended for long-term monitoring. Future research should deploy the Mc10 Biostamp_RC device on alternative locations to determine whether the findings seen here are replicated. Since completing this study, the Biostamp_RC has been discontinued by Mc10 and has been replaced by Biostamp nPoint. Despite these limitations, the recommendations within this study may be of practical support for researchers considering which device to use within their trials.

Conclusions

By using mixed methods and testing each device for a week, this study gained a robust understanding of the complexities of selecting a device for use within a clinical trial. The results indicate that no single sensor was considered optimal by participants due to a variety of factors, including the feedback provided by the device, its comfort, and battery life. Participants favored devices that they perceived they gained value from and were willing to overlook annoyances to receive feedback. Based on these results, the following context-specific recommendations can be made:

1. Researchers should consider their device selection in relation to both individual and environmental factors and not simply the primary outcome of the research study.
2. If researchers do not wish their participants to have access to the feedback from the devices, then a simple, wrist-worn device that acts as a watch is preferable.
3. If feedback is allowed, then it should be made available to help keep participants engaged. This is likely to apply only to people without cognitive impairments.
4. Battery life of 1 week should be considered as a necessary feature to enhance data capture.
5. Researchers should consider providing additional information about the purpose of devices to participants to support their continued use.

Acknowledgments

The authors would like to thank Craig Marronn for his help with the data collection of this study, and Brenda Reginatto for her assistance with the qualitative analysis of this study. This study was funded by Novartis Pharma. Novartis funded the purchase of the tested devices within this study.

Authors' Contributions

JD, LW, FC, and BC outlined the study design and aim. AK collected and analyzed the data. All authors contributed to the final version of the manuscript. AK is the guarantor of the study.

Conflicts of Interest

JD, FC, and LW work for Novartis.

Multimedia Appendix 1
Interview guide.
[DOCX File , 19 KB - mhealth_v8i4e15704_app1.docx]

Multimedia Appendix 2
Full list of tables of triangulated data per device.
[DOCX File , 29 KB - mhealth_v8i4e15704_app2.docx]

References

1. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54 [FREE Full text] [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
2. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
3. Bates M. IEEE Pulse Magazine. 2016 Mar 14. New Trends in Clinical Trials URL: https://pulse.embs.org/march-2016/new-trends-in-clinical-trials/ [accessed 2019-03-20]
4. Rosenberger M, Buman M, Haskell W, McConnell M, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
5. International Organization for Standardization. ISO 13407:1999(en) Human-Centred Design Processes for Interactive Systems URL: https://www.iso.org/obp/ui/#iso:std:iso:13407:ed-1:v1:en [accessed 2018-03-20]

6. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(Suppl 1):S52-S63. [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21656954]
7. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability study of mainstream wearable fitness devices: feature analysis and System Usability Scale evaluation. JMIR Mhealth Uhealth 2018 Nov 8;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
8. Jia Y, Wang W, Wen D, Liang L, Gao L, Lei J. Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. PeerJ 2018;6:e5350 [FREE Full text] [doi: 10.7717/peerj.5350] [Medline: 30065893]
9. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
10. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: A mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
11. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Ind Manag Data Syst 2015;115(9):1704-1723. [doi: 10.1108/imds-03-2015-0087]
12. Farina N, Sherlock G, Thomas S, Lowry RG, Banerjee S. Acceptability and feasibility of wearing activity monitors in community-dwelling older adults with dementia. Int J Geriatr Psychiatry 2019 Apr;34(4):617-624. [doi: 10.1002/gps.5064] [Medline: 30701592]
13. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth 2016 Apr 13;4(2):e35 [FREE Full text] [doi: 10.2196/mhealth.5120] [Medline: 27076486]
14. Toomey E, Matthews J, Hurley DA. Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention for people with osteoarthritis and low back pain. BMJ Open 2017 Aug 4;7(8):e015452 [FREE Full text] [doi: 10.1136/bmjopen-2016-015452] [Medline: 28780544]
15. Denton F, Spencer B. Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. Can J Aging 2010 Mar;29(1):11-21. [doi: 10.1017/S0714980809990390] [Medline: 20202262]
16. Schenkenfelder R, Selinger S. FFH Open Access Repository. 2016. A Comparison of Multiple Wearable Devices Regarding their User Experience During Running URL: http://ffhoarep.fh-ooe.at/bitstream/123456789/686/1/125_218_Schenkenfelder_FullPaper_en_Final.pdf [accessed 2020-02-18]
17. ActiGraph. ActiGraph GT9X Link URL: https://www.actigraphcorp.com/actigraph-link/ [accessed 2018-03-01]
18. Trium Analysis Online. URL: https://www.trium.de/ [accessed 2018-03-01]
19. Philips Respironics. Actiwatch Spectrum Plus URL: http://www.actigraphy.com/solutions/actiwatch/actiwatch-plus.html [accessed 2018-03-01]
20. Biovotion. URL: https://www.biovotion.com [accessed 2018-03-01]
21. Hexoskin Smart Shirts. URL: https://www.hexoskin.com [accessed 2018-03-01]
22. MC10: Wearable Healthcare Technology & Devices. Our Products URL: https://www.mc10inc.com/our-products [accessed 2018-03-01]
23. Wavelet HI. California, USA. 2018 Mar. Wavelet Health URL: http://www.wavelethealth.com/ [accessed 2018-03-01]
24. Brooke J. Hell - Jens Oliver Meiert. SUS - A Quick and Dirty Usability Scale URL: http://hell.meiert.org/core/pdf/sus.pdf [accessed 2018-03-20]
25. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123 [FREE Full text]
26. O'Reilly M, Slevin P, Ward T, Caulfield B. A wearable sensor-based exercise biofeedback system: mixed methods evaluation of Formulift. JMIR Mhealth Uhealth 2018 Jan 31;6(1):e33 [FREE Full text] [doi: 10.2196/mhealth.8115] [Medline: 29386171]
27. Centre for Self-Determination Theory. Intrinsic Motivation Inventory (IMI) URL: http://www.selfdeterminationtheory.org/questionnaires/10-questionnaires/50 [accessed 2018-03-20]
28. Orso V, Gamberini L, Guardigli E, Spagnolli A, Varotto A. Measuring User Acceptance of Wearable Symbiotic Devices: Validation Study across Application Scenarios. In: Proceedings of the International Workshop on Symbiotic Interaction. 2014 Presented at: Symbiotic'14; October 30-31 2014; Helsinki, Finland p. 87-98. [doi: 10.1007/978-3-319-13500-7]
29. Tessier D, Sarrazin P, Ntoumanis N. The effect of an intervention to improve newly qualified teachers' interpersonal style, students motivation and psychological need satisfaction in sport-based physical education. Contemp Educ Psychol 2010;35(4):242-253. [doi: 10.1016/j.cedpsych.2010.05.005]
30. Bryant C, Lewis P, Bennell K, Ahamed Y, Crough D, Jull G, et al. Can physical therapists deliver a pain coping skills program? An examination of training processes and outcomes. Phys Ther 2014 Oct;94(10):1443-1454. [doi: 10.2522/ptj.20130444] [Medline: 24903113]
31. Nadler J, Weston R, Voyles E. Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. J Gen Psychol 2015;142(2):71-89. [doi: 10.1080/00221309.2014.994590] [Medline: 25832738]

32. Keogh A, Matthews J, Segurado R, Hurley D. Feasibility of training physical therapists to deliver the theory-based self-management of osteoarthritis and low back pain through activity and skills (SOLAS) intervention within a trial. Phys Ther 2018 Feb 1;98(2):95-107. [doi: 10.1093/ptj/pzx105] [Medline: 29088437]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
35. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee T. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
36. Lee J, Kim D, Ryoo HY, Shin BS. Sustainable wearables: wearable technology for enhancing the quality of human life. Sustainability 2016;8(5):466 [FREE Full text] [doi: 10.3390/su8050466]
37. Lunney A, Cunningham N, Eastin M. Wearable fitness technology: a structural investigation into acceptance and perceived fitness outcomes. Comput Human Behav 2016;65(5):114-120. [doi: 10.1016/j.chb.2016.08.007]
38. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9):1408-1416 [FREE Full text]
39. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. Br Med J 2010 Sep 17;341:c4587. [doi: 10.1136/bmj.c4587] [Medline: 20851841]
40. Hanson WE, Creswell JW, Clark VL, Petska KS, Creswell JD. Mixed methods research designs in counseling psychology. J Couns Psychol 2005;52(2):224-235 [FREE Full text] [doi: 10.1037/0022-0167.52.2.224]
41. Ivankova NV, Creswell JW, Stick SL. Using mixed-methods sequential explanatory design: from theory to practice. Field Methods 2006;18(1):3-20 [FREE Full text] [doi: 10.1177/1525822x05282260]
42. Driscoll D, Appiah-Yeboah A, Salib P, Rupert DJ. Merging qualitative and quantitative data in mixed methods research: how to and why not. Ecol Environ Anthropol 2007;3(1):19-28 [FREE Full text]
43. Rupp MA, Michaelis JR, McConnell DS, Smither JA. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. Appl Ergon 2018 Jul;70:77-87. [doi: 10.1016/j.apergo.2018.02.005] [Medline: 29866329]
44. Karamanoğlu A, Erbug C. Perceived Qualities of Smart Wearables: Determinants of User Acceptance. In: Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces. 2011 Presented at: DPPI'11; June 22 - 25, 2011; Milan, Italy p. 1-8 URL: http://dl.acm.org/citation.cfm?doid=2347504.2347533 [doi: 10.1145/2347504.2347533]
45. Sauro J. MeasuringU. 2011 Feb 2. Measuring Usability With the System Usability Scale (SUS) URL: https://measuringu.com/sus/ [accessed 2018-03-20]
46. Steinert A, Haesner M, Steinhagen-Thiessen E. Activity-tracking devices for older adults: comparison and preferences. Univ Access Inf Soc 2018;17(2):411-419. [doi: 10.1007/s10209-017-0539-7]
47. Bodine K, Gemperle F. Effects of Functionality on Perceived Comfort of Wearables. In: Proceedings of the Seventh IEEE International Symposium on Wearable Computers. 2003 Presented at: ISWC'03; October 21-23, 2003; White Plains, NY, USA URL: http://ieeexplore.ieee.org/document/1241394/ [doi: 10.1109/iswc.2003.1241394]
48. Williamson B, Aplin T, de Jonge D, Goyne M. Tracking down a solution: exploring the acceptability and value of wearable GPS devices for older persons, individuals with a disability and their support persons. Disabil Rehabil Assist Technol 2017 Nov;12(8):822-831. [doi: 10.1080/17483107.2016.1272140] [Medline: 28100086]
49. Nascimento B, Oliveira T, Tam C. Wearable technology: What explains continuance intention in smartwatches? J Retail Consum Serv 2018;43:157-169 [FREE Full text] [doi: 10.1016/j.jretconser.2018.03.017]
50. Sultan N. Reflective thoughts on the potential and challenges of wearable technology for healthcare provision and medical education. International Journal of Information Management 2015 Oct;35(5):521-526. [doi: 10.1016/j.ijinfomgt.2015.04.010]
51. Zhang M, Luo M, Nie R, Zhang Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. Int J Med Inform 2017 Dec;108:97-109. [doi: 10.1016/j.ijmedinf.2017.09.016] [Medline: 29132639]
52. Farina N, Lowry R. Older adults' satisfaction of wearing consumer-level activity monitors. J Rehabil Assist Technol Eng 2017;4:2055668317733258 [FREE Full text] [doi: 10.1177/2055668317733258] [Medline: 31186939]
53. Patel M, Asch D, Volpp K. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]

Abbreviations

IMI: Intrinsic Motivation Inventory
SUS: System Usability Scale

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by N Farina, L Cilar; comments to author 21.10.19; revised version received 15.11.19; accepted 28.01.20; published 20.04.20.

Please cite as:
Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B
Comparing the Usability and Acceptability of Wearable Sensors Among Older Irish Adults in a Real-World Context: Observational Study
JMIR Mhealth Uhealth 2020;8(4):e15704
URL: http://mhealth.jmir.org/2020/4/e15704/
doi:10.2196/15704
PMID:32310149

Original Paper

Consumer Perceptions of Wearable Technology Devices: Retrospective Review and Analysis

Kimberly P L Chong[1*], BS; Julia Z Guo[1*], BS; Xiaomeng Deng[1*], BS; Benjamin K P Woo[1*], MD

University of California, Los Angeles, Sylmar, CA, United States

*all authors contributed equally

Corresponding Author:
Julia Z Guo, BS
University of California, Los Angeles
14445 Olive View Drive
Sylmar, CA, 91342
United States
Phone: 1 747 210 3830
Email: juliaguo@mednet.ucla.edu

Abstract

Background: Individuals of all ages are becoming more health conscious, and wearable technology devices (eg, Fitbit and Apple Watch) are becoming increasingly popular in encouraging healthy lifestyles.

Objective: The aim of this paper was to explore how consumers use wearable devices.

Methods: A retrospective review was done on the top-rated verified purchase reviews of the Fitbit One posted on Amazon.com between January 2014 and August 2018. Relevant themes were identified by qualitatively analyzing open-ended reviews.

Results: On retrieval, there were 9369 reviews with 7706 positive reviews and 1663 critical reviews. The top 100 positive and top 100 critical comments were subsequently analyzed. Four major themes were identified: sleep hygiene (“charts when you actually fall asleep, when you wake up during the night, when you're restless--and gives you a cumulative time of “actual sleep” as well as weekly averages.”), motivation (“25 lbs lost after 8 months – best motivator ever!”), accountability (“platform to connect with people you know and set little competitions or group…fun accountability if you set a goal with a friend/family.”), and discretion (“able to be clipped to my bra without being seen.”). Alternatively, negative reviewers felt that the wearable device’s various tracking functions, specifically steps and sleep, were inaccurate.

Conclusions: Wearable technology devices are an affordable, user-friendly application that can support all individuals throughout their everyday lives and potentially be implemented into medical surveillance, noninvasive medical care, and mobile health and wellness monitoring. This study is the first to explore wearable technology device use among consumers, and further studies are needed to examine the limitless possibilities of wearable devices in health care.

(JMIR Mhealth Uhealth 2020;8(4):e17544) doi:10.2196/17544

KEYWORDS

wearable technology devices; Fitbit; Amazon; sleep

Introduction

Wearable technology devices are applications for monitoring and tracking fitness-related metrics such as steps taken, distance walked or ran, and calories consumed. The World Health Organization recommends that adults aged 18 to 64 years should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week, at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week, or an equivalent combination of both to improve cardiorespiratory and muscular fitness and bone health, and reduce the risk of chronic diseases and depression [1]. Wearable devices provide an application for individuals to meet those guidelines. These devices are becoming increasingly popular due to advancements in technology and the public’s increased health consciousness. Wearable technology was determined to be the top fitness trend worldwide in 2016 and 2017, and continues to be number 1 in the top 20 worldwide fitness trends for 2019 [2-4]. Because of this growing trend, it becomes important to determine wearable devices’ potential to be used, not only in fitness, but in various aspects of health care. With continual improvements, wearable technology devices offer a wide array of functions, including the ability to track steps,

calories consumed and burned, floors climbed, sleep, and heart rate, as well as providing silent alarms. By using wearable devices, individuals are becoming more health conscious, thus, enabling them to take control of their own health. Therefore, we conducted an initial review of wearable technology devices, based on reviews posted on Amazon.com, to obtain an initial understanding of how this type of technology might be further implemented in health care.

Methods

A retrospective review of the top-rated verified purchase reviews of the Fitbit One posted on Amazon.com between March 2013 and August 2018 was conducted. The text of the top 100 most helpful positive and the top 100 most critical comments or reviews as rated by consumers were analyzed. To analyze the responses gathered on Amazon.com, the author used qualitative analysis to identify relevant themes. This research was deemed exempt from the University of California, Los Angeles Institutional Review Board.

Results

Findings

On initial retrieval, 9369 reviews of the Fitbit One on Amazon.com were submitted over the 5-year period previously mentioned. A total of 200 reviews were subsequently analyzed, 100 positive and 100 critical comments. Four major themes were identified: sleep hygiene, motivation, accountability, and discretion. Textbox 1 provides themes along with representative quotes. Demographic data of those who posted reviews were not available, unless the reviewer provided this information within their review. For example, one reviewer wrote “at the age of 75+…”.

Textbox 1. Themes of wearable technology device uses and representative quotes.

Sleep hygiene

- “I wondered why I was so tired when I got up in the morning. The fitbit really does track my sleep patterns. I found that I was awake numerous times (it even tell you exactly what times you are awake) and it shows when you are restless (it also shows those exact times). I found that about 1/2 of the time I am asleep I am not getting actual sleep time.”
- “Another feature is the sleep tracking; having sleep apnea, I can check my sleep quality from both my CPAP machine and the One. Some nights I wake up quite a bit, and the graphs reflect this.”

Motivation

- “…this works! It's easy to check during the day to keep you on target. In fact I actually WANT to check it, to see my progress. My goal, of course is 10,000 steps a day. Thanks to my Fitbit One, I know I'm going to get there on a regular basis. It gets me out and walking and keeps me moving. I am constantly challenging myself. I finally found something that motivates me to exercise.”
- “If you’re highly competitive getting one of these then competing with your friends on it might work…help keep you motivated.”

Accountability

- “…held me accountable and reminded me to get moving. After sitting at a desk the majority of the day, I would notice my steps were low and I would go for a walk/run or take an exercise class to reach my goal of 10,000 steps.”
- “there's a platform to connect with people you know and set little competitions or group goals…fun accountability if you set a goal with a friend/family.”

Discretion

- “…because I didn't want to wear a wristband all day since I am constantly typing I knew it would drive me nuts… I really like this little guy I clip it to my bra… don't even notice it is there throughout the day.”
- “The wristbands are all the rage, I know, but this little guy has a key advantage for me -- it clips onto my bra. Honestly, that is my #1 favorite feature. I don't have to worry about what it looks like because no one can see it. Ta-da! Out of sight, out of mind; no need to worry about whether a silicone bracelet goes with your outfit.”

Theme 1: Sleep Hygiene

Sleep was one of the more commonly cited uses. Fitbit One users reported using the sleep tracking function, which records the duration of sleep and when users are “asleep”, “restless”, and “awake”. One positive review stated, “I wondered why I was so tired when I got up in the morning…found that about ½ of the time I am asleep I am not getting actual sleep time.” Another positive review stated, “The sleep graph…encourages…to improve their sleep habits.” Some parents even found the Fitbit One useful in alerting them on when to take their child to see a physician due to the recordings of the child’s sleep patterns.

> *When my son tracked his sleep patterns, he found that he was averaging only about 2 hours of sleep a night. He has been diagnosed with sleep apnea and is receiving treatment.*

Users also enjoyed the silent alarm function, which allows the tracker to wake users up with a quiet vibration. One positive review stated, “ love the silent alarms – I was skeptical that a little thing buzzing on my wrist would wake me up – but it works great and my husband appreciates how quiet it is.”

Theme 2: Motivation

Several users discussed how they had been able to use the Fitbit One to continue to motivate themselves to stay active and continue to set goals as they strive to live healthier lifestyles. They have been able to successfully integrate the tracking functions into their daily routines, including steps and activities, calories burned, floors climbed, weight, water consumption, food intake, and exercise. One person wrote, "Since you become aware of what you are doing, it is self-motivating to adjust and add more activity." Another person stated:

> *We all know we have to be more active, but this little tracker reminds you, with facts and real data, how much real activity we are doing every day, and that is an eye-opener and a great motivator. I want to beat the tracker so I walk to get to the 10,000 steps.*

Not only is it a good tool for self-motivation, but the ability to form a personal network on the app with friends and family allows users to help motivate each other. One user wrote:

> *…the community is fun. I have a number of friends, family, and co-workers who have devices and we issue challenges a lot…definitely a nice way to get support.*

By using the tracking functions, the Fitbit One also helps users along their weight loss journey. Many reviews boasted about how using a wearable device helped them lose weight.

> *I've integrated my One into my everyday life, and I'm the happiest because of it. It's been just 2 ½ months since my first weight at the doctor's office. As of yesterday, I've dropped 24 pounds. And I did so naturally.*

Theme 3: Accountability

Several reviewers reported that they had been using the Fitbit One to keep them accountable. One user stated, "I do see myself improving my activity level from this as well as my eating habits, it makes one accountable for one's actions." Another wrote:

> *All I want right now is a reminder-helper-assistant-cool gadget, that helps me realize how much activity I'm doing or not doing, pats me in the back when I do something good and holds me accountable when I get lazy, all in a nice, cool, fun and easy way… I'm feeling accomplished, energetic and in control, and that to me is absolutely worth it.*

Theme 4: Discretion

Interestingly, discretion was a feature of the Fitbit One that many reviewers commented about. One positive review stated, "…didn't want everyone I came across to see that I was tracking my steps." Also, because this specific device allows the options of wearing it on one's wrist, clipped to clothing, or even tucked into a pocket, many users wrote positive reviews about this flexibility. One user wrote:

> *this little guy has a key advantage for me – it clips onto my bra. Honestly, that is my #1 favorite feature. I don't have to worry about what it looks like because no one can see it…no need to worry about whether a silicone bracelet goes with your outfit.*

Others did not want to be constantly preoccupied with checking their trackers: "I chose the clip-on style vs. the wristband because I didn't want to be hyper-aware of my steps and then start obsessing over them."

As expected, not all customer reviews were positive. Although the step tracker "seems pretty accurate" according to multiple users, some do complain of inaccuracies especially with the floors climbed tracker. One critical review stated:

> *The reliability of the count is not great but is acceptable when it comes to counting steps. When it comes to counting the floors climbed…found it ridiculously inaccurate. For example, a few times it indicated that I had climbed 10 or 15 floors when I had not left (my single floored) apartment.*

Some users also found the sleep tracker to be inaccurate. One user commented, "When it came to tracking my sleep, it suggested that I had minutes of sleep after I slept throughout the night." Other users also critiqued the possible inaccuracies of the sleep tracker due to the wrist Velcro band falling off while they were sleeping. Some critical reviews also pointed out that the Fitbit One did not contain a heart rate monitor and was not waterproof, but newer models have added these features and more.

Discussion

Principal Results

Our study suggests that wearable technology devices could be helpful tools to provide functional and social stimulation to individuals of all ages and support healthy lifestyles. Four major themes were discovered, including sleep hygiene, motivation, accountability, and discretion. This research looked at current real-world use of wearable technology devices as it examined "top rated, verified purchase" reviews of the Fitbit One. Compared with research tools like Actigraph, these devices are considered less accurate for some measurements [5,6]. However, they are generally less invasive, cheaper, more user-friendly, and more fashionable, as well as offering more functionality. The relative accessibility and affordability of wearable technology devices provides an opportunity to expand nationally and even worldwide. Multiple users commented on how Fitbit was a "lifesaver", from alerting a mother to bring her son to a physician to individuals who are obese who finally found a way to persevere with their weight loss journey. Currently, studies are being done with wearable devices and different disease states like sleep and arrhythmia. Although these studies show that the devices still have insufficient accuracy for clinical settings, solving technical issues and continuing to optimize clinically oriented features could make them available for use in clinical practice in a nondistant future [7-10]. Although Fitbit may not currently replace diagnostic tools like polysomnography for obstructive sleep apnea or Holter monitoring for atrial fibrillation, wearable devices may provide a cost-effective opportunity to alert individuals to see a physician earlier and prevent life-threatening complications.

Limitations

There were some limitations to this study. First, this was a retrospective review of comments that were only posted on Amazon.com. The purpose of these reviews was to provide opinions to other shoppers interested in purchasing the Fitbit One. There is a potential bias when only using posted reviews from selected customers since many more devices were sold than there were reviews. The strength of the study was that the comments were obtained from real-world users, but further research is needed on the actual use of wearable devices and potential applications in medical surveillance, noninvasive medical care, and mobile health and wellness monitoring [11-14]. Second, these reviews were limited to one specific company and technology, but studies have shown no significant difference in tracking when compared to other wearable devices like the Apple Watch, Jawbone, and Mi Band [15]. Third, we were unable to obtain specific demographic information on the users, so further research would be helpful in determining the different benefits within various age groups with the use of wearable technology devices. Fourth, there are benefits and risks to using wearable devices and further research should be done to identify these risks [16]. Nonetheless, valuable information can be obtained from this exploratory analysis to help guide in future research.

Future Research

Future research could investigate the barriers to the use of wearable technology devices in individuals of all ages and additional features that would encourage more use of these products. Possible improvements include making modifications for more precise tracking of steps taken and floors climbed, and sleep, as well as the use of a heart rate monitor and a longer lasting and stable material that is water resistant. This information could be valuable to the companies that develop the wearable devices for future upgrades to better serve customers. Thus, they could provide the mutually beneficial opportunity for increased sales and further enhancement of healthy living worldwide. In addition, further research is needed to determine how wearable technology devices can be implemented in health care. Research that is disease specific (eg, obstructive sleep apnea, atrial fibrillation, depression, obesity) would also be helpful in examining how best to expand the use of wearable devices. For example, sleep tracking can be used to diagnose obstructive sleep apnea; the heart rate monitor and possible advancing of monitoring heart activity can be used for arrhythmias; step and activity tracking can help physicians track individuals at risk for depression; food log and calorie trackers can track individuals at risk for conditions such as obesity, diabetes, hypercholesterolemia, or anorexia nervosa; stress level monitoring can track anxiety and panic attacks; and blood pressure monitoring can help those at risk for hypertension.

Conclusions

In conclusion, wearable technology devices show great promise in many aspects of health care, from fitness to health and wellness monitoring to possible future diagnostic tools. The wide range of reported quality of life improvement that the wearable technology device already provides (sleep hygiene, motivation, accountability, and discretion) are just a few of the various possibilities for companies to further develop this technology to impact lives worldwide.

Conflicts of Interest

None declared.

References

1. World HO. Global Recommendations On Physical Activity For Health. Geneva, Switzerland: World Health Organization,; 2010.
2. Thompson WR. Worldwide survey of fitness trends for 2016. ACSM s Health & Fitness Journal 2015;19(6):9-18. [doi: 10.1249/fit.0000000000000164]
3. Thompson WR. Worldwide survey of fitness trends for 2017. ACSM s Health & Fitness Journal 2016;20(6):8-17. [doi: 10.1249/fit.0000000000000252]
4. Thompson WR. Worldwide survey of fitness trends for 2019. ACSM s Health & Fitness Journal 2018;22(6):10-17. [doi: 10.1249/fit.0000000000000438]
5. Reid RER, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
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 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
7. Moreno-Pino F, Porras-Segovia A, López-Esteban P, Artés A, Baca-García E. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. J Clin Sleep Med 2019 Nov 15;15(11):1645-1653. [doi: 10.5664/jcsm.8032] [Medline: 31739855]
8. Al-Kaisey AM, Koshy AN, Ha FJ, Spencer R, Toner L, Sajeev JK, et al. Accuracy of wrist-worn heart rate monitors for rate control assessment in atrial fibrillation. Int J Cardiol 2020 Feb 01;300:161-164. [doi: 10.1016/j.ijcard.2019.11.120] [Medline: 31787389]

9. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
10. Svensson T, Chung U, Tokuno S, Nakamura M, Svensson AK. A validation study of a consumer wearable sleep tracker compared to a portable EEG system in naturalistic conditions. J Psychosom Res 2019 Nov;126:109822. [doi: 10.1016/j.jpsychores.2019.109822] [Medline: 31499232]
11. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
12. Gao W, Emaminejad S, Nyein HYY, Challa S, Chen K, Peck A, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 2016 Jan 28;529(7587):509-514 [FREE Full text] [doi: 10.1038/nature16521] [Medline: 26819044]
13. Jin H, Huynh T, Haick H. Self-healable sensors based nanoparticles for detecting physiological markers via skin and breath: toward disease prevention via wearable devices. Nano Lett 2016 Jul 13;16(7):4194-4202. [doi: 10.1021/acs.nanolett.6b01066] [Medline: 27328179]
14. Chia GLC, Anderson A, McLean LA. Behavior change techniques incorporated in fitness trackers: content analysis. JMIR Mhealth Uhealth 2019 Jul 23;7(7):e12768 [FREE Full text] [doi: 10.2196/12768] [Medline: 31339101]
15. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
16. Owens J, Cribb A. 'My Fitbit Thinks I Can Do Better!' Do health promoting wearable technologies support personal autonomy? Philos Technol 2017 Jun 6;32(1):23-38. [doi: 10.1007/s13347-017-0266-2]

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by L Liu, C Chee, S Shu, T Kawashita, T Muto; comments to author 05.02.20; revised version received 05.02.20; accepted 29.02.20; published 20.04.20.

Please cite as:
Chong KPL, Guo JZ, Deng X, Woo BKP
Consumer Perceptions of Wearable Technology Devices: Retrospective Review and Analysis
JMIR Mhealth Uhealth 2020;8(4):e17544
URL: http://mhealth.jmir.org/2020/4/e17544/
doi:10.2196/17544
PMID:32310148

Original Paper

Accuracy of Optical Heart Rate Sensing Technology in Wearable Fitness Trackers for Young and Older Adults: Validation and Comparison Study

Hsueh-Wen Chow[1], PhD; Chao-Ching Yang[1], MSc

Graduate Institute of Physical Education, Health & Leisure Studies, National Cheng Kung University, Tainan City, Taiwan

Corresponding Author:
Hsueh-Wen Chow, PhD
Graduate Institute of Physical Education, Health & Leisure Studies
National Cheng Kung University
No. 1 University Rd
East District
Tainan City
Taiwan
Phone: 886 62757575 ext 81806
Email: hwchow@mail.ncku.edu.tw

Abstract

Background: Wearable fitness trackers are devices that can record and enhance physical activity among users. Recently, photoplethysmography (PPG) devices that use optical heart rate sensors to detect heart rate in real time have become popular and help in monitoring and controlling exercise intensity. Although the benefits of using optical heart rate monitors have been highlighted through studies, the accuracy of the readouts these commercial devices generate has not been widely assessed for different age groups, especially for the East Asian population with Fitzpatrick skin type III or IV.

Objective: This study aimed to examine the accuracy of 2 wearable fitness trackers with PPG to monitor heart rate in real time during moderate exercise in young and older adults.

Methods: A total of 20 young adults and 20 older adults were recruited for this study. All participants were asked to undergo a series of sedentary and moderate physical activities using indoor aerobic exercise equipment. In this study, the Polar H7 chest-strapped heart rate monitor was used as the criterion measure in 2 fitness trackers, namely Xiaomi Mi Band 2 and Garmin Vivosmart HR+. The real-time, second-by-second heart rate data obtained from both devices were recorded using the broadcast heart rate mode. To critically analyze the results, multiple statistical parameters including the mean absolute percentage error (MAPE), Lin concordance correlation coefficient (CCC), intraclass correlation coefficient, the Pearson product moment correlation coefficient, and the Bland-Altman coefficient were determined to examine the performances of the devices.

Results: Both test devices exhibited acceptable overall accuracy as heart rate sensors based on several statistical tests. Notably, the MAPE values were below 10% (the designated threshold) in both devices ($Garmin_{Young}$=3.77%; $Garmin_{Senior}$=4.73%; $Xiaomi_{Young}$=7.69%; and $Xiaomi_{Senior}$=6.04%). The scores for reliability test of CCC for Garmin were 0.92 (Young) and 0.80 (Senior), whereas those for Xiaomi were 0.76 (Young) and 0.73 (Senior). However, the results obtained using the Bland-Altman analysis indicated that both test optical devices underestimated the average heart rate. More importantly, the study documented some unexpected outlier readings reported by these devices when used on certain participants.

Conclusions: The study reveals that commonly used optical heart rate sensors, such as the ones used herein, generally produce accurate heart rate readings irrespective of the age of the user. However, users should avoid relying entirely on these readings to indicate exercise intensities, as these devices have a tendency to produce erroneous, extreme readings, which might misinterpret the real-time exercise intensity. Future studies should therefore emphasize the occurrence rate of such errors, as this will likely benefit the development of improved models of heart rate sensors.

(JMIR Mhealth Uhealth 2020;8(4):e14707) doi:10.2196/14707

KEYWORDS

pulse; photoplethysmography; wearable device; aerobic exercise

Introduction

Growing Popularity and Functions of Wearable Fitness Trackers

Wearable fitness trackers have gained popularity worldwide, and their annual sales continue to grow [1,2]. These trackers were listed as the No. 1 fitness trends in the years 2016, 2017, 2019, and 2020 in a worldwide survey conducted by the American College of Sports and Medicine [3-6]. The advantages of these wearable devices are that they are convenient to use and measure various parameters noninvasively. In addition, they allow the users to monitor their daily physical activities in a free-living environment instead of controlled laboratory settings.

Earlier versions of fitness trackers, equipped with triaxial accelerometers and a gyroscope, could sense motions made by the users, monitor their activity metrics, and provide estimated information such as walking and running in terms of steps or distance, energy expenditure, sedentary time, sleep patterns, and activity routes (with GPS function). Most of these fitness trackers were placed on the wrist. The users obtained the real-time information from the display on the trackers or received feedback through connected mobile phone apps.

The recent application of photoplethysmography (PPG) in wrist-based wearable fitness trackers has enabled newer versions of fitness trackers to detect heart rates. This breakthrough provides several benefits. First, heart rate is a vital component in cardiovascular fitness assessments and an important parameter in exercise training programs [7]. Second, resting heart rate is also a widely used parameter for general health assessments to detect cardiovascular diseases [8]. Thus, the development of fitness trackers that have heart rate detection technologies has brought about several additional benefits that were absent in older models.

PPG measures heart rates based on the changes in vascular blood flow during the cardiac cycle [9]. It has previously been applied in medical devices such as oximeters [10]. This technology has since been integrated and commercialized as optical heart rate monitors by companies such as Mio and Omron. The number of commercial companies producing such devices has gradually grown in the last 5 years (ie, Apple Watch, Fitbit, and Garmin), along with the design and development of such products and research [1,2,10-16].

Validation of Fitness Trackers

Despite the growing popularity and functions of these fitness trackers and substantial investments in commercial advertisements, many users have expressed concerns regarding the data accuracy of these trackers [17]. Inaccurate and inconsistent readings are major reasons for negative user experiences, which discourage the continued use of these devices [17-20]. The concerns regarding the data accuracy of these trackers influence the users in terms of their perceptions of personal health and program interventions or research evaluations that adopt these devices.

Most commercially available fitness trackers use step counts as a parameter to indicate the level of physical activity. The step-count function of these devices has been widely scrutinized in studies examining their accuracy [21-23]. Importantly, while generally producing accurate results, these devices did not report reliable step-count readings in certain conditions, such as slow walking or while performing unnatural hand movements [21-24]. A systematic review investigated the validity and reliability of Fitbit and Jawbone trackers. The results revealed that most studies validated the tracker accuracy and indicated that it had a higher accuracy for step counts, followed by that for distance and physical activity and finally for energy consumption and sleep [23]. Nevertheless, most studies recommend caution when deriving energy expenditure estimations directly using these readings [11,13,25,26]. In addition, studies have started to examine the validity and reliability of the fitness trackers among older adults instead of young adults because they might present different movements such as gait patterns or speeds [27,28].

Accuracy of Optical Heart Rate Monitoring

The accuracy of heart rate displayed on the fitness trackers with optical heart rate monitors has also been investigated [11,14,16,29-31]. Common research methods for the development of these optical heart rate monitors involve fitness assessments using basic indoor training equipment such as treadmills, stationary cycles, and sometimes elliptical machines. This type of study allows researchers to evaluate the feasibility of implementing optical heart rate monitors in aerobic training for the general population [1,2,10-16].

Previous studies have reported that, generally, optical sensing fitness trackers have acceptable accuracy. However, the accuracy might vary across brands [16,31] in terms of activity patterns or speed, exercise intensities [10,14,31], skin tone [10], room temperature [32], placement of sensors [29], or compression-induced and motion-induced artifacts [13,32-34]. For example, in a study conducted by Boudreaux et al [13], participants wore 8 different fitness trackers, and an increase in exercise intensity reduced the accuracy of heart rate measurement. In another validation study, the measured heart rate showed a minor deviation compared with the actual heart rate in participants with a dark skin tone [30].

Although the adoption of heart rate fitness trackers with optical heart rate sensors in the medical field is still debatable [12,35], there have been several lawsuits regarding the accuracy of heart rate information [36,37]. Assessing the reliability and validity of the heart rate readings provided by these trackers is essential because they are vital in clinical settings, and these trackers have been increasingly accepted by consumers as a tool for self-monitoring or in many intervention programs for health management [11,14].

Research Gaps

Owing to the limitations on raw data acquisition in commercial fitness trackers, previous studies have only used average heart rate data [14] or manually recorded the heart rate at certain intervals [11]. However, averaging the heart rate or recording it at a certain time point is problematic because both fail to represent any change or variability [38]. Studies that have

compared continuous heart rate in more detail revealed that evaluating the accuracy of these test devices at a second-by-second level is difficult [2]. One study used video recording to manually determine the second-by-second heart rate, which was a labor-intensive and time-consuming method [12]. Moreover, potential variables such as age, ethnicity, and gender were not considered in earlier studies [2,14]. For example, a majority of the participants of several studies that have been conducted in the US-European regions were white (Fitzpatrick skin type I or II) [2,12,16]. PPG technology uses an optical sensor that illuminates light and measures the change in light absorption by the skin, which varies with change in blood volume; thus, the accuracy of heart rate monitoring using PPG is subject to skin structures [39]. Typically, the skin changes with age, that is, "fine wrinkles, roughness, mottled hyperpigmentation, dilated blood vessels, and loss of skin tone" are observed [40]. In addition, age-related changes such as arterial stiffness can influence the pulse shape in PPG [32]. Therefore, appropriate validation of these devices for different age groups among non-white participants is imperative.

Aim of the Study

This study evaluated the heart rate reading performances of 2 commercially available fitness trackers in various settings using a second-by-second data acquisition approach. Moreover, to determine whether age would generate discrepancies in the readouts, young and senior participants were characterized separately. This study was conducted in Taiwan to validate 2 trackers used by the yellow skin tone population (Fitzpatrick skin type III or IV) [41,42].

Methods

Participants

To determine a credible sample size for achieving statistical power in the intraclass coefficient correlation (ICC) test, this study used R package (ICC.Sample. Size, GPL-3; 2015, R core team, R Foundation for Statistical Computing). Based on the formula proposed by Zou [43], the number of participants (n) required for achieving a target power of 0.90 was 8. Therefore, this study involved 20 adults aged 65 years and above (Senior) and 20 adults aged between 20 years and 26 years (Young). All participants had no clinical history of cardiovascular diseases, neurological disorders, lower limb injuries, or any other factors that would render them unfit to perform the exercise. To ensure consistency, individuals with tattoos or birthmarks on the position where the device was to be worn were not included in the study. To minimize possible sex-driven discrepancies, the sex ratio in both the Senior and Young groups was kept identical (20:20).

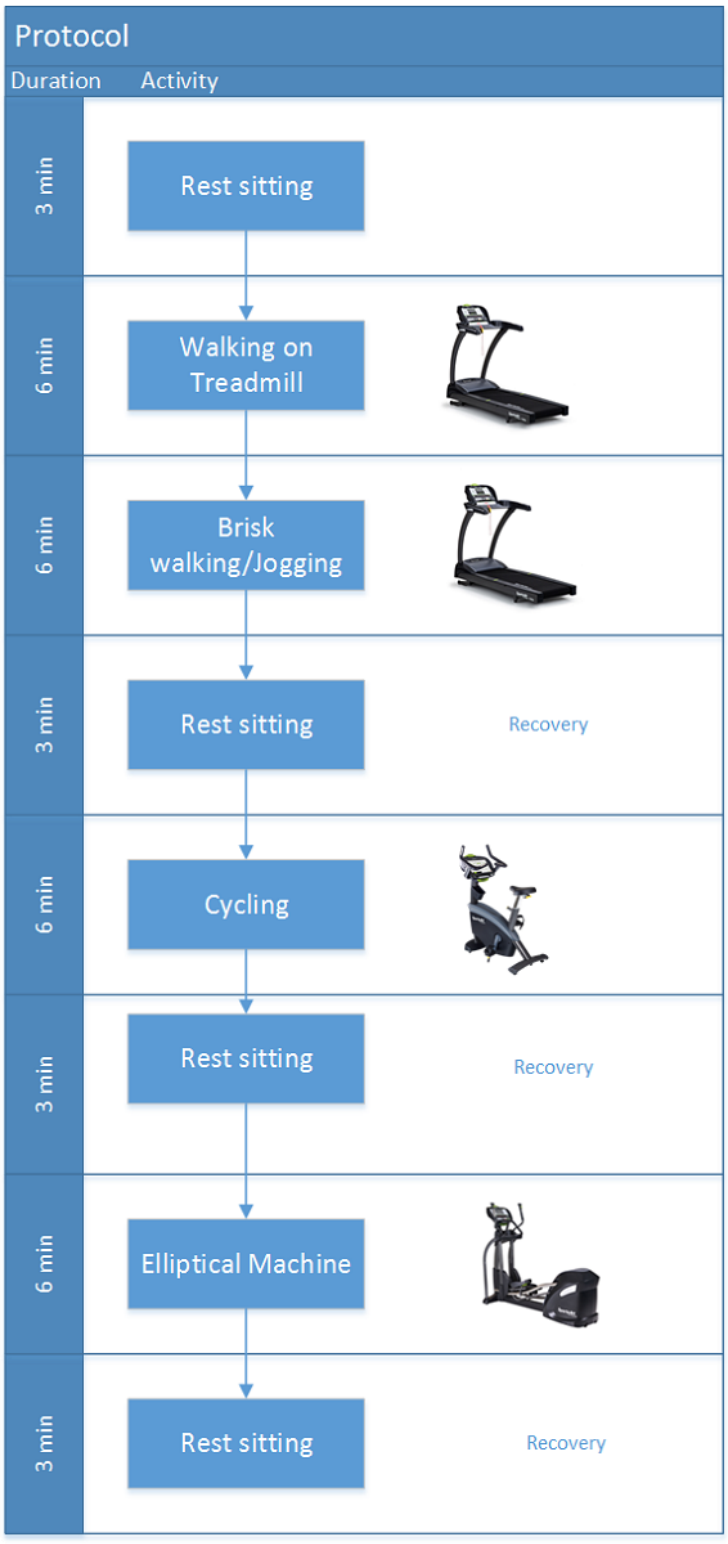
Research Device

This study used the Polar heart rate strap (H7, Polar Electro Oy), widely used as the criterion for measuring heart rate in sports science studies [2,44]. The optical fitness trackers selected for this study were Xiaomi Mi Band 2 (Xiaomi Cooperation) and Garmin Vivosmart HR+ (Garmin International Inc) because these 2 fitness trackers share a significant market share in the Asia Pacific region, which is expected to grow. Mi Band 2 was equipped with a PPG module (with 2 LED lights) and an accelerometer to detect heart rate and sense motion. Vivosmart HR+ was also equipped with a PPG module (with 3 LED lights) and an accelerometer. In addition, GPS chips are embedded in the Vivosmart HR+ for measuring the travel distance during outdoor exercises.

Both the devices provided information regarding step counts, energy expenditure, notification for breaking up the prolonged sedentary time, and smart notifications, and both claimed accurate heart rate detection. In addition, the 2 devices had the broadcast heart rate mode, a feature that enables the transmission of second-by-second heart rate data through Bluetooth or ANT+ to the paired receiving device, and served a similar function of the conventional heart rate strap. Moreover, wrist-based fitness trackers were easy to wear and remove and, thus, eased the discomfort of wearing chest straps for monitoring the real-time heart rate during traditional exercise and fitness training programs or interventions [10,45]. Specifically, PPG fitness trackers provide pulse rate data that are obtained with an increase or decrease in blood pressure in the arteries because of the contraction and relaxation of the heart, thus leading to a noticeable pulse. Although the signals of pulse waveforms are different from those of heartbeat waveforms, the pulse rate can be analyzed to represent the heart rate [32]. The term heart rate has been used in this study in line with many studies on heart rate fitness trackers [2,10-12,15,29,30,38,46]. Hence, in this study, the heart rate will be used in its broadest sense to refer to the readings from the optical fitness trackers.

The second-by-second heart rate data-receiving app Cardio Training (Angelfmarcos) used in this study was acquired from the Android platform. The equipment adopted in this study included 3 indoor aerobic fitness equipment: treadmill, upright stationary bike, and elliptical machine (Figure 1). These types of equipment were widely demonstrated in the previous exercise protocols and proved to be ideal and safe for aerobic training [2,10,47-49].

Figure 1. Exercise protocol.



Procedure

Before the Trial

The study was approved by the Institutional Review Board of the National Cheng Kung University Hospital (IRB number: B-ER-106-134). All participants gave written consent to participate in the trial and were provided a detailed explanation of the complete research protocol before the commencement of the study. All participants were given the option to voluntarily withdraw from the trial at any time during the study.

Polar H7 chest-strapped heart rate monitors and wrist-strapped optical fitness trackers were fixed onto the participants by the researcher according to the manufacturer instructions. Next, the broadcast heart rate mode of the optical fitness trackers was activated by the researcher simultaneously. Data transmission to the tablets or mobile phones was then checked.

Exercise Protocol

Initially, participants were asked to be seated quietly for 15 min to record their resting heart rates (HR_{rest}) using the Polar H7 heart rate monitors. The general formula (220−age in years) was used for calculating the maximal heart rate (HR_{max}) of each individual. Based on the HR_{rest} and HR_{max}, a personalized *moderate exercise intensity* was determined for each participant. This was defined by 40% to 60% of heart rate reserve, which is the difference between HR_{max} and HR_{rest} [50]. Finally, participants were led to the exercise area and shown the proper usage and adjustment of the specific fitness equipment.

To evaluate the heart rate detection accuracy of the test devices during different activities, participants were instructed to perform a sequence of sedentary and aerobic exercises [2,10]. The sequence was divided into phases, and heart rates were recorded using the Cardio Training app at each phase. The participants were initially guided to adjust the workout level of equipment accordingly to prevent exhaustion before the end of the trial. Specifically, the measurement began with the participants seated (rest sitting), which represented a typical sedentary behavior. Next, participants were asked to walk on the treadmill for 6 min (the warm-up phase) before engaging in more vigorous exercises. Every period of the exercise phase lasted for 6 min. The step-by-step protocol is presented in Figure 1. Rest sitting time was given to the participants between each phase, during which the heart rate measurement would continue.

During the exercise phases, participants were encouraged to maintain moderate exercise intensity. Real-time feedback and instructions were given by the researcher verbally as guided by the heart rate data acquired from the Polar H7 heart rate monitor. Except in circumstances where the participant deviated from moderate exercise intensity, in which the resistance level was adjusted accordingly, no further intervention by the researcher was made during the entire trial.

Statistical Analyses

Using the Cardio Training app, the second-by-second heart rate data generated from the trials were exported as CSV files. A total of 2161 readings, corresponding to 2161 seconds (including the first reading at the beginning of the protocol), were obtained and recorded for each participant. Compared with previous studies, in which heart rate measurements were less frequent (ie, every 15 seconds/every minute or only at the end of each exercise phase) [11,15,16], the statistical results produced from the current dataset are likely to be more representative because they enabled the researchers to discern some potential outlier readings. To compare the accuracy of test devices, various statistical methods were chosen based on recommendations from relevant studies [2,10,26,38,51]. All statistical tests were performed using SPSS 18.0 (IBM) and MedCalc statistical software (MedCalc).

Reliability

To compare the reliability between the criterion measurement device (Polar H7) and the 2 test optical fitness trackers, 3 reliability tests were used, namely the Lin concordance correlation coefficient (CCC), Pearson product moment correlation coefficient (PPMCC), and ICC tests (two-way mixed, single measures, and absolute agreement). Discrepant standards were used for interpreting the results of the reliability correlation tests. For instance, Gillinov et al [2] set the CCC value greater than 0.80 to represent acceptable reliability, whereas Boudreaux et al [13] set ICC values from 0.60 to 0.75 to represent moderate reliability and from 0.75 to 0.90 to indicate superior reliability. Moreover, other studies on applied sports science have proposed a slightly different version of interpreting ICC values: values between 0.50 and 0.75 indicated moderate reliability, whereas other thresholds were the same [52]. This study used all 3 of the aforementioned reliability tests.

Analysis of Paired Difference

Paired absolute differences from mean absolute error (MAE) and mean absolute percentage error (MAPE) were determined to reveal the differences between the criterion measurement and measurements generated by the test devices among respective age groups and during different phases of the exercise (MAPE is calculated by subtracting the HR readings from the Mi or Garmin from the Polar H7 and then dividing by the Polar H7). Results with error values below 10% were considered reliable [13].

Bland-Altman Analysis

To determine the agreement of the criterion measurement and measurements generated by the optical fitness trackers, Bland-Altman analysis was applied to explore the mean bias and 95% CI limits of agreement. The results from different age groups and during different phases of the exercise were analyzed and represented graphically.

Results

Reliability of Examined Devices

The results of MAE, MAPE, and correlation tests from both the Young and Senior groups are shown in Tables 1 and 2. In the Young group, the Garmin device achieved MAPE values of less than 10% in all the conditions tested (Table 1), indicating that overall, the heart rate readings produced by the Garmin device were reliable [2,13]. By contrast, whereas the Xiaomi device generally achieved MAPE values of less than 10%, it did not do so during cycling and elliptical phases (Table 1), suggesting that the reliability of the Xiaomi device was likely influenced by the types of activities performed.

In the Senior group, the performances of both test devices during different activities were reliable (MAPE values below 10%, Table 1). Notably, the MAPE values achieved by the Xiaomi device were, on average, higher than those produced by the Garmin device, indicating that the Xiaomi product was overall less reliable than the Garmin one. However, the standard deviation of MAPE achieved by the Garmin device was higher in the Senior group (SD_{Senior}=10.49%) than in the Young group (SD_{Young}=6.9%; Table 1), suggesting that the reliability of the Garmin device was likely affected by age differences and that it became less reliable in the older population.

Table 1. Mean absolute percentage error (MAPE) and Bland-Altman analyses of heart rate readings of the Young and Senior groups during different activity phases.

Group, activity, number of readings, and device			MAPE analysis, mean (SD)		Bland-Altman analysis
			Mean absolute error (bpm)	Mean absolute percentage error	Mean difference (lower to upper limits of agreement)
Young					
Rest					
	3620				
		Ga[a]	2.98 (3.14)	3.96 (4.17)	−1.4 (−9.4 to 6.6)
		Mi[b]	3.27 (4.48)	4.46 (6.05)	0 (−10.9 to 10.8)
Walking					
	7200				
		Ga	3.35 (4.73)	3.77 (5.29)	0.2 (−11.5 to 11.2)
		Mi	6.39 (7.93)	7.46 (9.93)	3.7 (−14.8 to 22.3)
Running					
	7200				
		Ga	3.48 (7.66)	2.85 (6.29)	−2.6 (−18.3 to 13.1)
		Mi	10.41 (12.99)	8.32 (10.54)	6.7 (−23.1 to 36.6)
Cycling					
	7200				
		Ga	6.19 (14.41)	4.92 (10.79)	−5.7 (−34.3 to 23.0)
		Mi	14.05 (20.56)	10.93 (15.36)	−13.4 (−54.5 to 27.8)
Elliptical					
	7200				
		Ga	3.06 (5.11)	2.52 (4.32)	−2.0 (−13.0 to 9.0)
		Mi	14.06 (19.73)	10.77 (14.88)	−13.3 (−53.0 to 26.4)
Recovery					
	10,800				
		Ga	4.40 (7.22)	4.38 (6.85)	1.0 (−15.4 to 17.5)
		Mi	4.86 (7.90)	4.73 (7.60)	0.5 (−17.7 to 18.7)
Total					
	43,220				
		Ga	4.03 (8.21)	3.77 (6.90)	−1.6 (−19.3 to 16.1)
		Mi	8.85 (14.46)	7.69 (11.66)	−2.6 (−35.5 to 30.3)
Senior					
Rest					
	3620				
		Ga	1.96 (3.53)	2.45 (4.11)	−1.0 (−8.7 to 6.6)
		Mi	4.03 (6.54)	5.59 (9.67)	1.2 (−13.7 to 16.1)
Walking					
	7200				
		Ga	6.72 (10.56)	7.06 (10.96)	4.3 (−18.8 to 27.3)
		Mi	8.09 (12.77)	8.69 (13.85)	2.8 (−26.3 to 31.9)
Running					

Group, activity, number of readings, and device			MAPE analysis, mean (SD)		Bland-Altman analysis
			Mean absolute error (bpm)	Mean absolute percentage error	Mean difference (lower to upper limits of agreement)
	7200				
		Ga	2.7 (4.36)	2.54 (4.08)	−1.4 (−11.0 to 8.3)
		Mi	7.46 (16.73)	7.02 (16.14)	3.6 (−31.6 to 38.8)
Cycling					
	7200				
		Ga	3.85 (11)	3.65 (9.69)	−3.2 (−25.2 to 18.7)
		Mi	3.91 (7.5)	3.78 (6.92)	−2.4 (−18.3 to 13.6)
Elliptical					
	7200				
		Ga	5.19 (10.94)	5.04 (11.51)	0.6 (−23.1 to 24.3)
		Mi	7.31 (11.55)	6.38 (9.36)	−5.2 (−30.0 to 19.6)
Recovery					
	10,800				
		Ga	5.43 (11.58)	5.92 (13.48)	2.2 (−22.4 to 26.9)
		Mi	4.85 (8.46)	5.05 (8.97)	0.1 (−19.0 to 19.3)
Total					
	43,220				
		Ga	4.6 (9.93)	4.73 (10.49)	0.5 (−20.9 to 21.9)
		Mi	6.02 (11.39)	6.04 (11.33)	−0.1 (−25.3 to 25.2)

[a]Ga: Garmin Vivosmart HR^{+}.

[b]Mi: Xiaomi Mi Band 2.

The data revealed that the Garmin device achieved CCC values above the designated threshold (0.80) in both age groups (Table 2), suggesting that it was generally accurate. By contrast, the Xiaomi device failed to achieve overall CCC values above the designated threshold in both age groups (Table 2), indicating that it exhibited suboptimal accuracy in heart rate sensing. Notably, similar to the MAPE values described earlier, whereas the Xiaomi device achieved identical CCC values in both age groups (CCC_{Young}=0.73; CCC_{Senior}=0.73), the Garmin device's CCC values fluctuated between the 2 age groups (CCC_{Young}=0.93; CCC_{Senior}=0.80; Table 2), indicating that its accuracy was also likely influenced by age differences. Taken together, these data suggest that the Garmin device, in general, produced more reliable and accurate heart rate readings than the Xiaomi one.

Table 2. Correlation analyses of heart rate readings of Young and Senior groups during different activity phases.

Group, activity, number of readings, and device			Correlation		
			CCC[a]	ICC[b]	PPMCC[c]
Young					
Rest					
	3620				
		Ga[d]	0.9037	0.9038	0.914
		Mi[e]	0.8475	0.8475	0.8475
Walking					
	7200				
		Ga	0.8577	0.8577	0.8598
		Mi	0.6074	0.6074	0.6461
Running					
	7200				
		Ga	0.8552	0.8552	0.8858
		Mi	0.5428	0.5428	0.6185
Cycling					
	7200				
		Ga	0.5888	0.5889	0.6569
		Mi	0.2874	0.2874	0.4037
Elliptical					
	7200				
		Ga	0.9104	0.9104	0.9261
		Mi	0.3267	0.3267	0.4734
Recovery					
	10,800				
		Ga	0.8972	0.8972	0.8993
		Mi	0.8863	0.8863	0.8931
Total					
	43,220				
		Ga	0.9254	0.9254	0.9277
		Mi	0.7603	0.7603	0.767
Senior					
Rest					
	3620				
		Ga	0.9262	0.9262	0.9306
		Mi	0.7320	0.7321	0.7369
Walking					
	7200				
		Ga	0.5925	0.5925	0.722
		Mi	0.4464	0.4464	0.5469
Running					
	7200				

Group, activity, number of readings, and device		Correlation		
		CCC[a]	ICC[b]	PPMCC[c]
	Ga	0.9246	0.9246	0.9311
	Mi	0.4592	0.4593	0.5288
Cycling				
7200				
	Ga	0.4799	0.4799	0.5129
	Mi	0.6856	0.6856	0.7081
Elliptical				
7200				
	Ga	0.7516	0.7516	0.7684
	Mi	0.6612	0.6612	0.6992
Recovery				
10,800				
	Ga	0.7055	0.7055	0.7253
	Mi	0.7929	0.793	0.7934
Total				
43,220				
	Ga	0.8000	0.8000	0.8084
	Mi	0.7258	0.7258	0.7341

[a]CCC: concordance correlation coefficient.

[b]ICC: intraclass coefficient correlation.

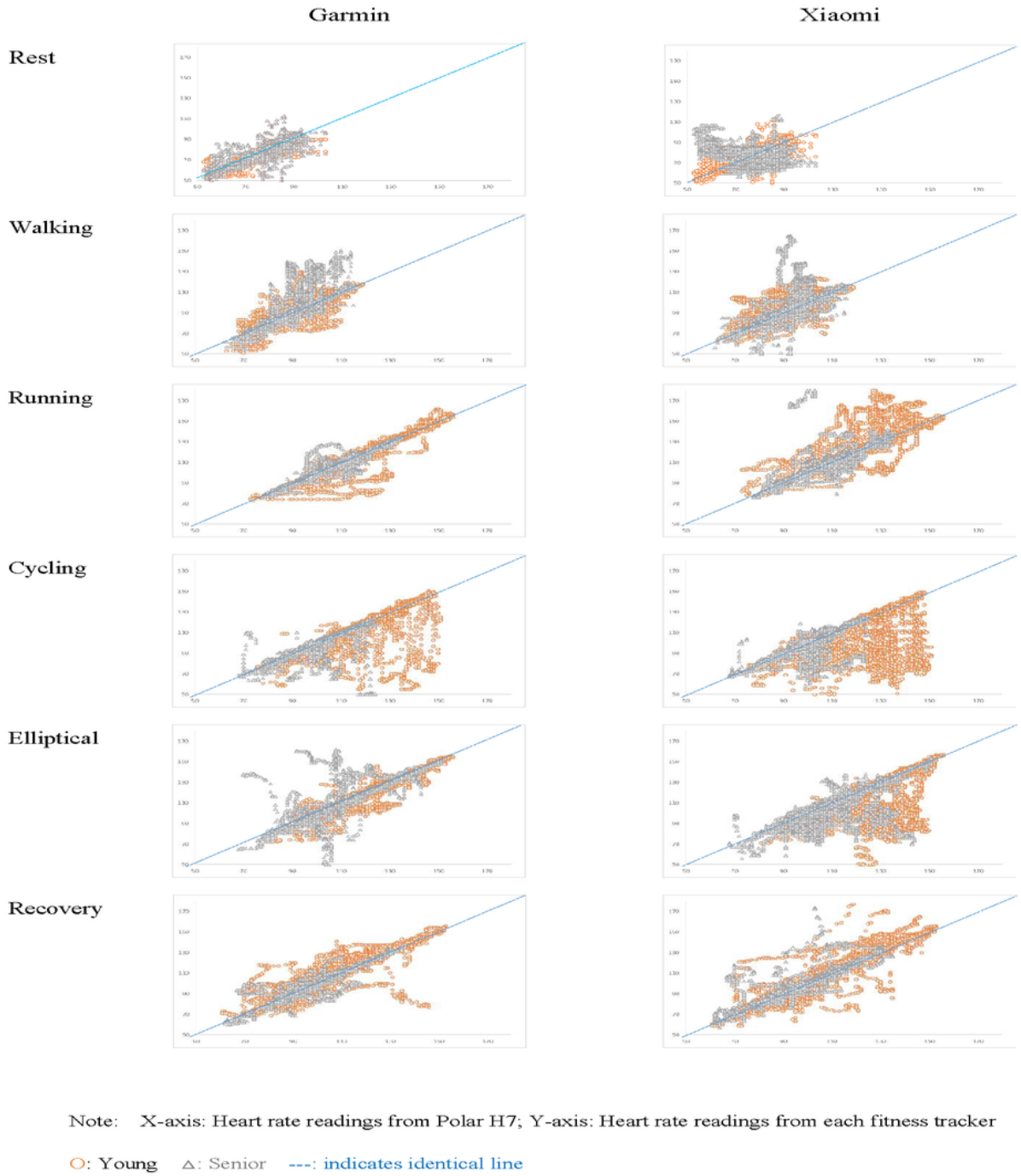
[c]PPMCC: Pearson product moment correlation coefficient.

[d]GA: Garmin Vivosmart HR^{+}.

[e]Mi: Xiaomi Mi Band 2.

To observe the overall trends and identify any apparent discrepancies in the correlation in different situations, each phase within the exercise sequence was plotted separately and color coded. The overlaid datasets of the different groups are represented in the scatter gram in Figure 2. Notably, the correlation of certain activities, such as cycling, was found to deviate from the criterion measurements much more frequently than activities such as walking. This was further confirmed using the Bland-Altman analysis (Table 1; see *Bland-Altman Analysis*).

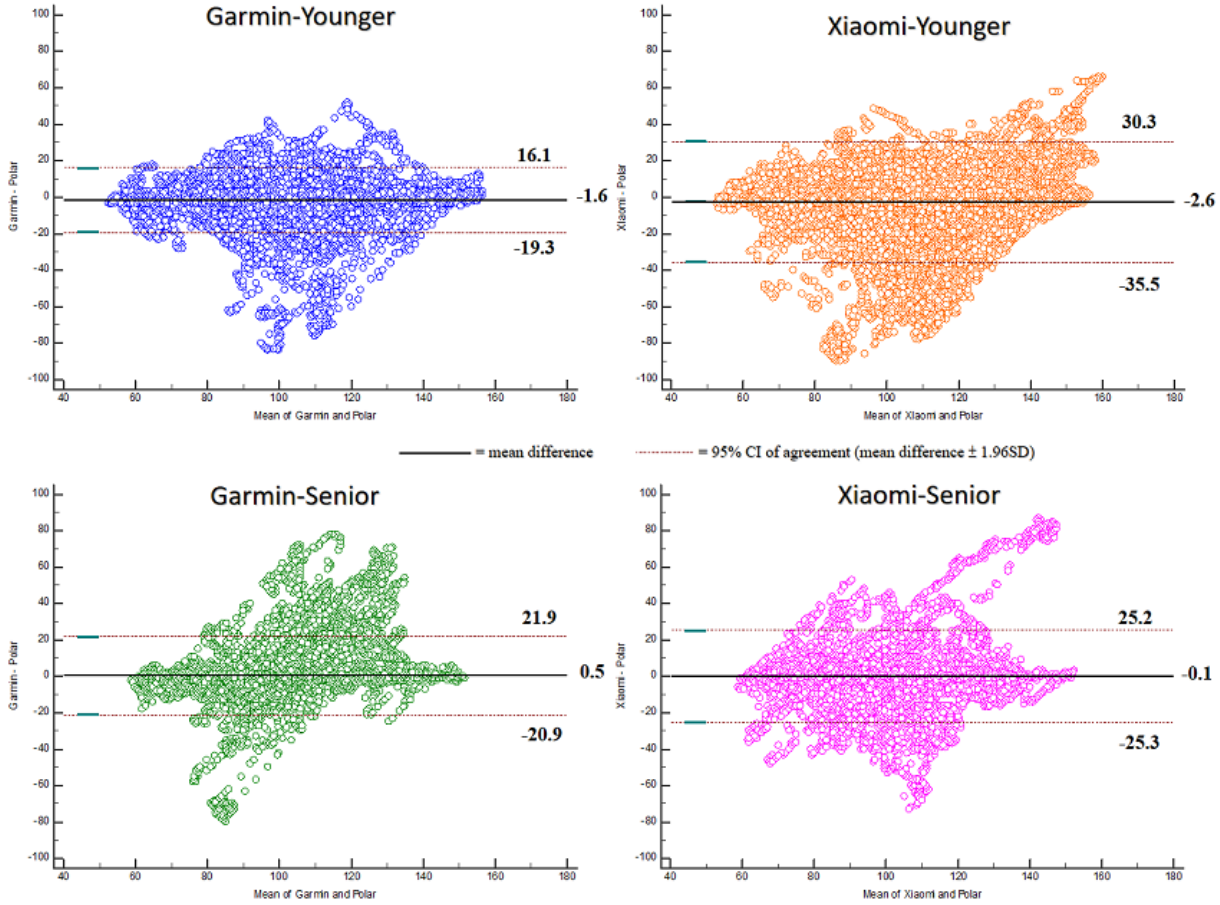
Figure 2. Scatter diagrams of the different phases of activities for different devices and groups.



Bland-Altman Analysis

Bland-Altman plots indicating the mean difference in heart rate detection between Garmin or Xiaomi and Polar H7 criterion measure and levels of agreement with 95% CIs for the Young and Senior groups are illustrated in Figure 3. The complete Bland-Altman analysis dataset is presented in Table 1 (the Bland-Altman plot for each activity phase is provided in Multimedia Appendix 1). The data indicated that both test devices achieved relatively higher variations during cycling phases compared with other activities (Table 1). These results suggest that both devices tended to underreport heart rates in certain situations, consistent with previous observations [16,24]. Notably, the Xiaomi device significantly underestimated heart rates during cycling and elliptical phases in the Young group (−13.4 bpm and −13.3 bpm, respectively). Moreover, the differences between the upper and lower limits during the recovery phase (rest sitting between active phases) were greater than those during the resting phase (rest sitting in the beginning; Table 1). This implies that the variation of differences was greater at the transitional phases in which participants changed their activities from dynamic exercise to recovery, and thus, the degree of errors might decrease gradually if the participants stay in the rest position.

Figure 3. Bland-Altman plots of overall phases for different age groups and devices.



Comparison of Correlation Tests

Various combinations of correlation tests are frequently adopted in evaluating the reliability or validity of examined devices [35]. As such, 3 independent statistical tests were employed in this study to compare whether the results from different correlation tests would deviate.

The obtained results (Table 2) revealed that the PPMCC test might compute a higher correlation coefficient than the CCC and ICC tests. The results of all the phases were quite identical; for example, the maximum difference was less than 0.01 (0.7258 and 0.7341 for Mi Band 2 in the Senior group). However, the difference between CCC or ICC and PPMCC was more obvious for activities; for example, a higher deviation was noted for activities such as cycling and elliptical exercise.

Discussion

Principal Findings

In line with previous studies [2,11,16,53], the combined results from this study indicated that both the Garmin and Xiaomi devices generally provided accurate heart rate readings. Both devices were also considered reliable in heart rate measurements with overall MAPE values below the 10% threshold. Notably, even though both devices achieved acceptable overall correlations in both age groups, they showed a tendency to modestly underestimate heart rates in many situations, as revealed by the Bland-Altman analysis. Similar findings were also reported in previous studies [11,12] and could represent a general characteristic of optical heart rate fitness trackers.

However, it is worth noting that significant discrepancies in device accuracy remained apparent between different physical activities. In general, these devices would be more accurate during sedentary behaviors such as sitting compared with active exercise [2]. Indeed, a previous study on a number of commercial wearable activity monitors have found that most devices exhibited low ICC values ($r<0.5$) when the activity intensity exceeded 100 watts in graded cycling exercise [13]. Similarly, our data revealed that the test devices generally had lower correlation coefficients and higher degrees of deviation during cycling and elliptical exercises compared with other activities.

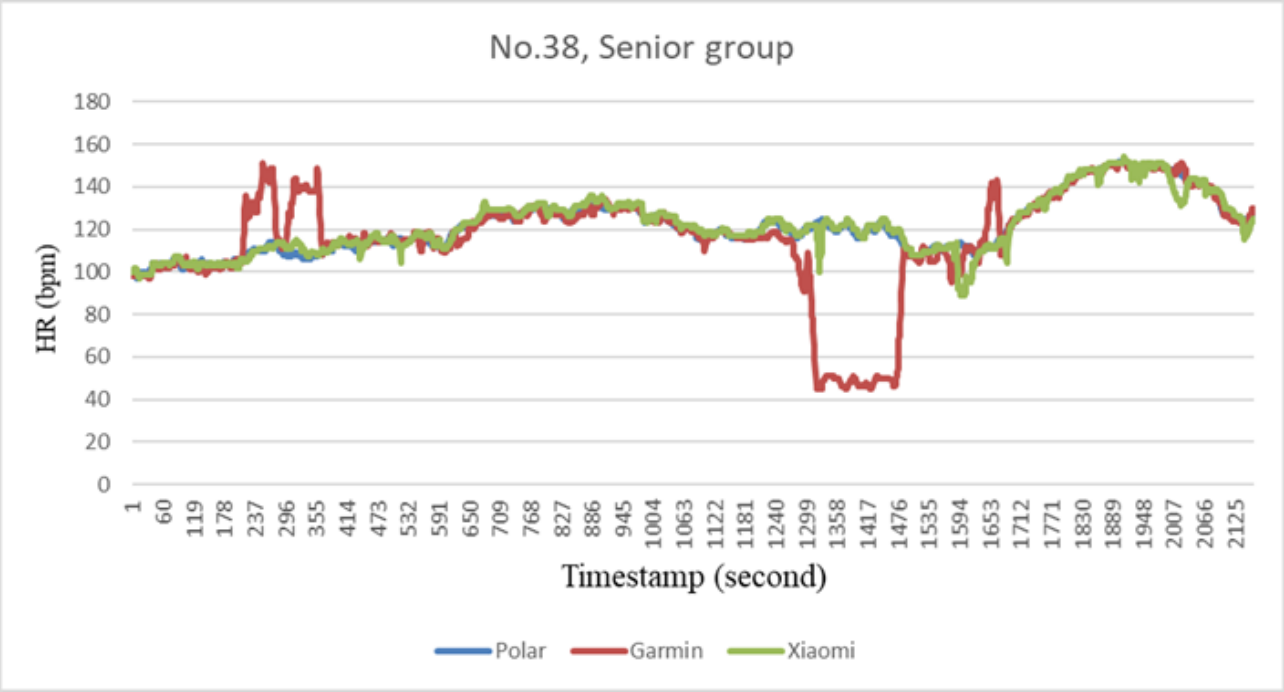
In addition to activity intensity, several other studies have identified that motion artifacts during exercise were negatively correlated with the accuracy of PPG heart rate–monitoring systems [32,38,46,54-56]. For example, in an experiment conducted by Gillinov et al [2], the optical devices exhibited more accuracy for exercise with fewer arm motion artifacts (cycling and elliptical exercise with no arms movement). It is somewhat surprising that the data collected in this study indicated the opposite (as cycling produced less motion artifact than running). Nevertheless, Benedetto et al [12] found that the Fitbit charge 2 had poor ICC values ($r=0.21$) and underestimated

the actual heart rate values when performing stationary cycling. Without further conclusions, users should be cautious when relying on optical heart rate readouts during various physical activities. Taken together, this study provides supporting evidence for a negative correlation between activity type and the accuracy of optical heart rate sensors but not between motion artifacts and the accuracy of optical heart rate sensors [2,13,24]. The precise mechanisms for such correlations currently remain unclear.

The profoundly expanding aging population worldwide is creating challenges for all sectors in the society. Promoting health condition of the older adult population and motivating them to engage in regular physical activity have become essential [57]. The adoption of new technology such as using health-related informatics technology (such as apps) or wearable fitness trackers is increasing [20,58,59], and the benefits are also observed in the senior population [20,60]. The fitness trackers validated in this study appear to exhibit similar accuracy for heart rate detection among different age groups.

Given its more thorough data acquisition method, this study had identified certain unexpected outliers. As shown in Figure 4, these extreme readings were unexpected, unpredictable, and transient. It is likely that these extreme readings did not represent the true heart rate values and that their displays were technical faults of the devices or the detection approach. Nonetheless, these random (or untrue) readings can skew the overall dataset and falsely represent the heart rate of an individual. Because these extreme heart rate readings were only observed for a short period, detecting these deviations while examining the heart rate readings every 15 seconds, every minute, or only at the end of the exercise, as in earlier studies, is difficult [11,15,16]. Given the transiency of such extreme readings, it is therefore recommended that future studies on optical heart rate sensors adopt a second-by-second approach demonstrated here and previously [12] to identify the outliers.

Figure 4. Subject-specific recording errors during the cycling phase.



Previous studies have proposed the use of different statistical methods to analyze the data correlation. These include the MAPE test, the Bland-Altman analysis, the correlation PPMCC, ICC, and CCC tests [2,12,13,15,53]. To minimize the insufficiencies of individual statistical tests, this study examined the second-by-second heart rate readings using all of the mentioned correlation tests. Our results showed that when given the same dataset, PPMCC tests would typically derive higher values than ICC or CCC tests. Although all of the correlation coefficients have previously been adopted in other studies on optical devices, future research should exercise caution when selecting correlation tests and interpreting test results. That said, ICC and CCC should nonetheless be the preferred tests, as they were initially used to assess the interrater reliability in related validation studies [61,62]. Sartor et al [38] also supported the use of the CCC test for validating wrist-based heart rate monitors. Another study has proposed standardization of exercise protocols to ensure that the aggregate data were reproducible [51]. Thus, a standard set of examining methods and statistical analyses should be developed and adopted in future validation studies of optical heart rate sensors.

In conclusion, this study revealed that both the Garmin and Xiaomi optical heart rate sensors were capable of producing fairly accurate heart rate readings for both young and older adults. In particular, these devices achieved better accuracy during sedentary behaviors compared with physical activities. The heart rate reading accuracy of both devices was influenced by different types of physical activities. Consistently, the results echoed the previously reported tendency for heart rate underestimation during cycling and elliptical training in both of the devices. Notably, both devices exhibited the tendency to transiently display erroneous extreme readings. Thus, cautions

should be exercised when using wrist-strapped fitness trackers to monitor the real-time heart rate during aerobic exercises.

Limitations

This study was limited by several factors. First, the test devices were chosen because of their popularity in Asia and the availability of the broadcasting heart rate mode on these devices. However, different brands would usually be integrated with different PPG modules or algorithms, which could lead to discrepancies among the different optical heart rate devices [2,11]. This makes direct interpretations of findings on other optical heart rate devices using the current results more difficult. Although this study strived to retrieve the second-by-second data, the heart rate signals derived from various devices were complex, and the time lag problem existed between the investigational and reference devices [38]; in addition, owing to the trade secrets pertaining to the PPG signal-processing algorithms and the receiving apps, we could only assume that the second-by-second data are from the nearest previous beat-to-beat waveform signal to represent the heart rate readings. Nevertheless, the PPG sensor provided satisfactory readings when it was worn on the wrist than on other body parts. Second, the exercise intensity in this study was set at a submaximal level because of the various physical conditions of the participants. Thus, performance of these examining devices during more vigorous intensity exercises remains to be examined. In addition, this study only selected healthy participants, that is, participants without any cardiovascular diseases (eg, coronary artery disease or abnormal heart rhythms) or neurological disorders (eg, Parkinson disease or essential tremor) because the abnormal heart rate might interfere in the accuracy of comparison [63,64]. Hence, the results cannot be generalized to the overall older adult population. The validity of PPG fitness trackers for a population with major disorders, such as patients with cardiac disorders, requires further investigation.

Suggestions

Future research on these topics should benefit from the standardization of the exercise protocol, selected statistical methods, and the threshold of acceptable accuracy. This will allow for better cross-study comparisons and more accurate interpretations [51]. Second, future studies can incorporate more participants with various health conditions to increase the representativeness of the cohort. Conducting multiple trials for the same cohort will control variability. This will also help identify erroneous readings, especially when they fall within the physiological range. For similar reasons, the second-by-second data acquisition method presented in this study should be adopted in all future studies. This will also help address the mechanisms of those conceivably erroneous displays. Third, future testing should include more contextual activities, such as outdoor walking, running, and cycling, to better mimic real-life events. This will allow for better comparisons of device performances under different settings.

Conclusions

Overall, the results of this study indicate that both the Garmin and Xiaomi optical heart rate sensors exhibit acceptable heart rate–sensing accuracy for yellow skin tone population (Fitzpatrick skin type III or IV). Both devices perform similar to the Polar H7 chest-strapped heart rate monitor. The results also indicate that the sensing reliability of both the Garmin and Xiaomi devices can be influenced by different types of physical activities and that the Garmin device generally outperformed the Xiaomi device. The accuracy of both devices was not significantly affected by the age of users which implies that both devices are suitable for use in older adults. This has significant implications for the increasing aging population because PPG fitness trackers are inexpensive and use a noninvasive technology to provide information regarding various parameters and they have a great potential for telemedicine use considering remote or home health monitoring, assisting the older adult population to monitor their health [32].

The accuracy levels of both devices were negatively correlated with the level of activity intensity. For both devices, the measurement accuracy deteriorated in individuals while cycling. For unknown reasons, this study also reports the occurrence of extreme errors in these heart rate–sensing devices. These relevant findings imply that users or exercise practitioners should be cautious when using wrist-strapped fitness trackers to monitor exercise performance.

Acknowledgments

This research was funded by Ministry of Science and Technology, Taiwan (ROC) grants #108-2410-H-006-099. This paper is developed from CY's master thesis—CY designed and performed experiments, analyzed data, and prepared the manuscript. HC conceived the study idea, supervised research progress, provided research materials, performed data analysis, critically edited the manuscript and tables or graphs, responded to reviewers' comments, and acquired financial support. The authors would like to thank all the participants who participated in this study. The authors also thank Dr. Tsang-Hai Huang for his help in this project. The authors would like to thank the editor and the 4 anonymous reviewers who provided valuable feedback and guidance on revising the manuscript before publishing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bland-Altman plots of each phase for different groups and devices.
[DOCX File , 934 KB - mhealth_v8i4e14707_app1.docx]

References

1. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
2. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
3. Thompson WR. Worldwide survey of fitness trends for 2016. ACSMs Health Fit J 2015;19(6):9-18. [doi: 10.1249/fit.0000000000000164]
4. Thompson WR. Worldwide survey of fitness trends for 2017. ACSMs Health Fit J 2016;20(6):8-17. [doi: 10.1249/fit.0000000000000252]
5. Thompson WR. Worldwide survey of fitness trends for 2019. ACSMs Health Fit J 2018;22(6):10-17. [doi: 10.1249/fit.0000000000000438]
6. Thompson WR. Worldwide survey of fitness trends for 2020. ACSMs Health Fit J 2019;23(6):10-18. [doi: 10.1249/fit.0000000000000526]
7. Laukkanen RM, Virtanen PK. Heart rate monitors: state of the art. J Sports Sci 1998 Jan;16 Suppl:S3-S7. [doi: 10.1080/026404198366920] [Medline: 22587712]
8. Böhm M, Reil J, Deedwania P, Kim JB, Borer JS. Resting heart rate: risk indicator and emerging risk factor in cardiovascular disease. Am J Med 2015 Mar;128(3):219-228. [doi: 10.1016/j.amjmed.2014.09.016] [Medline: 25447617]
9. Challoner AV, Ramsay CA. A photoelectric plethysmograph for the measurement of cutaneous blood flow. Phys Med Biol 1974 May;19(3):317-328. [doi: 10.1088/0031-9155/19/3/003] [Medline: 4445210]
10. 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]
11. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: Implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
12. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
13. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
14. 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]
15. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
16. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 1;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
17. Shih PC, Han K, Poole ES, Rosson MB, Carroll JM. Use and Adoption Challenges of Wearable Activity Trackers. In: iConference 2015 Proceedings. 2015 Presented at: iConference'15; March 24-27, 2015; Newport Beach, California, USA URL: https://www.ideals.illinois.edu/handle/2142/73649
18. Michaelis JR, Rupp MA, Kozachuk J, Ho B, Zapata-Ocampo D, McConnell DS, et al. Describing the user experience of wearable fitness technology through online product reviews. Proc Hum Factors Ergon Soc Annu Meet 2016;60(1):1073-1077. [doi: 10.1177/1541931213601248]
19. Yang R, Shin E, Newman MW, Ackerman MS. When Fitness Trackers Don't 'Fit': End-user Difficulties in the Assessment of Personal Tracking Device Accuracy. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 623-634 URL: https://dl.acm.org/doi/abs/10.1145/2750858.2804269 [doi: 10.1145/2750858.2804269]
20. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
21. Takacs J, Pollock C, Guenther J, Bahar M, Napier C, Hunt M. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport 2014 Sep;17(5):496-500. [doi: 10.1016/j.jsams.2013.10.241] [Medline: 24268570]
22. 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;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
23. 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]

24. Dondzila C, Lewis C, Lopez J, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):575-584 [FREE Full text]
25. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
26. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
27. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014 Jun;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006] [Medline: 24656871]
28. Lauritzen J, Muñoz A, Sevillano JL, Civit A. The usefulness of activity trackers in elderly with reduced mobility: a case study. Stud Health Technol Inform 2013;192:759-762. [Medline: 23920659]
29. Parak J, Korhonen I. Evaluation of Wearable Consumer Heart Rate Monitors Based on Photopletysmography. In: Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.: IEEE; 2014 Presented at: EMBC'14; Aug 26-30, 2014; Chicago, IL, USA p. 3670-3673. [doi: 10.1109/embc.2014.6944419]
30. Preejith S, Alex A, Joseph J, Sivaprakasam M. Design, Development and Clinical Validation of a Wrist-based Optical Heart Rate Monitor. In: Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA).: IEEE; 2016 Presented at: MeMeA'16; May 15-18, 2016; Benevento, Italy URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7533786 [doi: 10.1109/memea.2016.7533786]
31. Fokkema T, Kooiman TJ, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
32. 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]
33. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiol Meas 2010 Sep;31(9):1271-1290. [doi: 10.1088/0967-3334/31/9/015] [Medline: 20702919]
34. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 5;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
35. Wright SP, Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 1;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
36. Leagle. 2018 Jun 5. McLellan v. Fitbit, Inc. (N.D. Cal. 2016) (No. 16-cv-36), 2016 WL 64721, at *1-2 URL: https://www.leagle.com/decision/infdco20180606a63 [accessed 2019-02-15]
37. Leagle. 2016 May 10. Robb v. Fitbit, Inc., (Case No. 16-cv-00151-SI.) URL: https://www.leagle.com/decision/infdco20160511935 [accessed 2019-02-15]
38. Sartor F, Papini G, Cox LG, Cleland J. Methodological shortcomings of wrist-worn heart rate monitors validations. J Med Internet Res 2018 Jul 2;20(7):e10108 [FREE Full text] [doi: 10.2196/10108] [Medline: 29967000]
39. Delgado R, Parák J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of Accuracy and Reliability of PulseOn Optical Heart Rate Monitoring Device. In: Proceedings of 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).: IEEE; 2015 Presented at: EMBC'15; Aug 25-29, 2015; Milan, Italy p. 430-433. [doi: 10.1109/embc.2015.7318391]
40. McCullough JL, Kelly KM. Prevention and treatment of skin aging. Ann N Y Acad Sci 2006 May;1067:323-331. [doi: 10.1196/annals.1354.044] [Medline: 16804006]
41. Grimes PE, Sherrod Q. Structural and physiologic differences in the skin of darker racial ethnic groups. In: Grimes PE, editor. Aesthetics and Cosmetic Surgery for Darker Skin Types. New York: Lippincott Williams & Wilkins; 2008:15-26.
42. Australian Radiation Protection and Nuclear Safety Agency. Fitzpatrick Skin Phototype URL: https://www.arpansa.gov.au/sites/default/files/legacy/pubs/RadiationProtection/FitzpatrickSkinType.pdf [accessed 2019-11-02]
43. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012 Dec 20;31(29):3972-3981. [doi: 10.1002/sim.5466] [Medline: 22764084]
44. Cheatham SW, Kolber MJ, Ernst MP. Concurrent validity of resting pulse-rate measurements: a comparison of 2 smartphone applications, the polar H7 belt monitor, and a pulse oximeter with bluetooth. J Sport Rehabil 2015 May;24(2):171-178. [doi: 10.1123/jsr.2013-0145] [Medline: 24700420]
45. 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 Oct 20;5(10):e157 [FREE Full text] [doi: 10.2196/mhealth.8233] [Medline: 29055881]

46. Lang M. Beyond Fitbit: a critical appraisal of optical heart rate monitoring wearables and apps, their current limitations and legal implications. Alb LJ Sci Tech 2017;28(1):39-72 [FREE Full text]
47. Dorgo S, Robinson KM, Bader J. The effectiveness of a peer-mentored older adult fitness program on perceived physical, mental, and social function. J Am Acad Nurse Pract 2009 Feb;21(2):116-122. [doi: 10.1111/j.1745-7599.2008.00393.x] [Medline: 19228250]
48. Prosser LA, Stanley CJ, Norman TL, Park HS, Damiano DL. Comparison of elliptical training, stationary cycling, treadmill walking and overground walking. Electromyographic patterns. Gait Posture 2011 Feb;33(2):244-250 [FREE Full text] [doi: 10.1016/j.gaitpost.2010.11.013] [Medline: 21215636]
49. Stanish HI, Temple VA. Efficacy of a peer-guided exercise programme for adolescents with intellectual disability. J Appl Res Intellect Disabil 2012 Jul;25(4):319-328. [doi: 10.1111/j.1468-3148.2011.00668.x] [Medline: 22711480]
50. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Ninth Edition. Philadephia, PA: Lippincott Williams & Wilkins; 2014.
51. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current state of commercial wearable technology in Physical Activity Monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
52. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
53. Sartor F, Gelissen J, van Dinther R, Roovers D, Papini GB, Coppola G. Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. BMC Sports Sci Med Rehabil 2018;10:10 [FREE Full text] [doi: 10.1186/s13102-018-0098-0] [Medline: 29881626]
54. Alqaraawi A, Alwosheel A, Alasaad A. Heart rate variability estimation in photoplethysmography signals using Bayesian learning approach. Healthc Technol Lett 2016 Jun;3(2):136-142 [FREE Full text] [doi: 10.1049/htl.2016.0006] [Medline: 27382483]
55. Sweeney KT, Ward TE, McLoone SF. Artifact removal in physiological signals--practices and possibilities. IEEE Trans Inf Technol Biomed 2012 May;16(3):488-500. [doi: 10.1109/TITB.2012.2188536] [Medline: 22361665]
56. Ahmadi AK, Moradi P, Malihi M, Karimi S, Shamsollahi MB. Heart Rate Monitoring During Physical Exercise Using Wrist-Type Photoplethysmographic (PPG) Signals. In: Proceedings of 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.: IEEE; 2015 Presented at: EMBC'15; August 25-29, 2015; Milan, Italy p. 6166-6169 URL: https://ieeexplore.ieee.org/abstract/document/7319800 [doi: 10.1109/embc.2015.7319800]
57. Nied R, Franklin B. Promoting and prescribing exercise for the elderly. Am Fam Physician 2002 Feb 1;65(3):419-426 [FREE Full text] [Medline: 11858624]
58. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth 2016 Apr 13;4(2):e35 [FREE Full text] [doi: 10.2196/mhealth.5120] [Medline: 27076486]
59. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
60. Rasche P, Wille M, Theis S, Schäefer K, Schlick C, Mertens A. Activity Tracker and Elderly. In: Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing.: IEEE; 2015 Presented at: CIT/IUCC/DASC/PICOM'15; October 26-28, 2015; Liverpool, UK p. 1411-1416. [doi: 10.1109/cit/iucc/dasc/picom.2015.211]
61. Chen CC, Barnhart HX. Comparison of ICC and CCC for assessing agreement for data without and with replications. Comput Stat Data Anal 2008;53(2):554-564 [FREE Full text] [doi: 10.1016/j.csda.2008.09.026]
62. Chen CC, Barnhart HX. Assessing agreement with intraclass correlation coefficient and concordance correlation coefficient for data with repeated measures. Comput Stat Data Anal 2013;60:132-145 [FREE Full text] [doi: 10.1016/j.csda.2012.11.004]
63. Finsterer J, Wahbi K. CNS-disease affecting the heart: brain-heart disorders. J Neurol Sci 2014 Oct 15;345(1-2):8-14. [doi: 10.1016/j.jns.2014.07.003] [Medline: 25034054]
64. Dyer AR, Persky V, Stamler J, Paul O, Shekelle RB, Berkson DM, et al. Heart rate as a prognostic factor for coronary heart disease and mortality: findings in three Chicago epidemiologic studies. Am J Epidemiol 1980 Dec;112(6):736-749. [doi: 10.1093/oxfordjournals.aje.a113046] [Medline: 7457467]

Abbreviations

CCC: concordance correlation coefficient
ICC: intraclass coefficient correlation
MAE: mean absolute error
MAPE: mean absolute percentage error
PPG: photoplethysmography
PPMCC: Pearson product moment correlation coefficient

Edited by G Eysenbach; submitted 14.05.19; peer-reviewed by M Lang, J Parak, C Poon, A Gorny; comments to author 30.09.19; revised version received 08.12.19; accepted 04.02.20; published 28.04.20.

Please cite as:
Chow HW, Yang CC
Accuracy of Optical Heart Rate Sensing Technology in Wearable Fitness Trackers for Young and Older Adults: Validation and Comparison Study
JMIR Mhealth Uhealth 2020;8(4):e14707
URL: http://mhealth.jmir.org/2020/4/e14707/
doi:10.2196/14707
PMID:32343255

Original Paper

Step-Based Metrics and Overall Physical Activity in Children With Overweight or Obesity: Cross-Sectional Study

Jairo H Migueles[1], MSc; Cristina Cadenas-Sanchez[1], PhD; Elroy J Aguiar[2], PhD; Pablo Molina-Garcia[1,3], MSc; Patricio Solis-Urra[1,4], MSc; Jose Mora-Gonzalez[1], MSc; Eduardo García-Mármol[1], PhD; Eric J Shiroma[5], PhD; Idoia Labayen[6], PhD; Palma Chillón[1], PhD; Marie Löf[7,8], PhD; Catrine Tudor-Locke[2], PhD; Francisco B Ortega[1,7], PhD

[1]PROFITH (PROmoting FITness and Health through physical activity) Research Group, Department of Physical and Sports Education, Faculty of Sport Sciences, University of Granada, Granada, Spain

[2]Department of Kinesiology, University of Massachusetts Amherst, Amherst, MA, United States

[3]Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium

[4]IRyS (Investigación en Rendimiento Y Salud) Research Group, School of Physical Education, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

[5]Laboratory of Epidemiology and Population Science, National Institute on Aging, National Institutes of Health, Bethesda, MD, United States

[6]Institute for Innovation and Sustainable Development in the Food Chain (IS-FOOD), Public University of Navarra, Pamplona, Spain

[7]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[8]Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Jairo H Migueles, MSc
PROFITH (PROmoting FITness and Health through physical activity) Research Group
Department of Physical and Sports Education, Faculty of Sport Sciences
University of Granada
Carretera Alfacar s/n
Granada, 18011
Spain
Phone: 34 958244353
Email: jairohm@ugr.es

Abstract

Background: Best-practice early interventions to increase physical activity (PA) in children with overweight and obesity should be both feasible and evidence based. Walking is a basic human movement pattern that is practical, cost-effective, and does not require complex movement skills. However, there is still a need to investigate how much walking—as a proportion of total PA level—is performed by children who are overweight and obese in order to determine its utility as a public health strategy.

Objective: This study aimed to (1) investigate the proportion of overall PA indicators that are explained by step-based metrics and (2) study step accumulation patterns relative to achievement of public health recommendations in children who are overweight and obese.

Methods: A total of 105 overweight and obese children (mean 10.1 years of age [SD 1.1]; 43 girls) wore hip-worn accelerometers for 7 days. PA volumes were derived using the daily average of counts per 15 seconds, categorized using standard cut points for light-moderate-vigorous PA (LMVPA) and moderate-to-vigorous PA (MVPA). Derived step-based metrics included volume (steps/day), time in cadence bands, and peak 1-minute, 30-minute, and 60-minute cadences.

Results: Steps per day explained 66%, 40%, and 74% of variance for counts per 15 seconds, LMVPA, and MVPA, respectively. The variance explained was increased up to 80%, 92%, and 77% by including specific cadence bands and peak cadences. Children meeting the World Health Organization recommendation of 60 minutes per day of MVPA spent less time at zero cadence and more time in cadence bands representing sporadic movement to brisk walking (ie, 20-119 steps/min) than their less-active peers.

Conclusions: Step-based metrics, including steps per day and various cadence-based metrics, seem to capture a large proportion of PA for children who are overweight and obese. Given the availability of pedometers, step-based metrics could be useful in discriminating between those children who do or do not achieve MVPA recommendations.

Trial Registration: ClinicalTrials.gov NCT02295072; https://clinicaltrials.gov/ct2/show/NCT02295072

(JMIR Mhealth Uhealth 2020;8(4):e14841) doi:10.2196/14841

KEYWORDS

motion sensor; pedometer; sedentary behavior; MVPA; cadence

Introduction

Decreased physical activity (PA) is associated with increased risk of noncommunicable diseases [1-3] and is responsible for approximately 9% of premature mortality [4]. Worldwide PA deficits [2,5-7] and inequalities between countries regarding PA levels [8] require effective counteractive strategies, especially in populations at risk such as people with overweight and obesity. For example, evidence [9,10] suggests that low levels of PA initiated in childhood and perpetuated in adulthood set up adults with overweight and obesity for an increased array of comorbidities during their life span [11]. Best-practice early interventions should be both feasible and evidence based. Walking is a basic human movement pattern that is practical, cost-effective, and does not require complex movement skills. Thus, focusing on ambulatory activity could be the most accessible strategy to increase PA levels in children with overweight and obesity [12] who do not engage frequently in sports [13] and present poorer movement skills than normal-weight children [14]. However, there is still a need to investigate how much ambulatory activity is performed by children with overweight and obesity, as proportion of total PA level, in order to determine its utility as a public health strategy. Information on what type of PA children with overweight and obesity are more likely to perform could help to plan more effective public health strategies, since intervening on a behavior that is frequently occurring (eg, walking) would have a greater impact than generating a new behavior.

The ability to study health-related influences of PA has advanced in parallel with the increased use of accelerometer-based wearable technologies [15]. Accelerometers are capable of detecting human movement, but are primarily sensitive to ambulatory activity, the most common form of PA performed by adults [16,17]. However, children's movement patterns may be more variable and less is known about how predominant ambulatory activity, primarily walking, is relative to other types of PA behaviors. Time-stamped accelerometers are capable of detecting step-based metrics, including a tally of step accumulation over the day (ie, volume [steps/day]), the time spent in incremental cadence bands (eg, time spent walking at 80-99 steps/min), and/or peak 1-minute, 30-minute, and 60-minute cadence indices (ie, average steps/min of the highest 1, 30, or 60 nonconsecutive minutes in a day, respectively) [18-20]. Collectively, these metrics are referred to hereafter as step-based metrics.

Therefore, this study aimed to (1) investigate the proportion of overall PA that is explained by ambulatory activity (ie, step-based metrics) in children with overweight and obesity and (2) study step-based patterns relative to PA guidelines achievement in children with overweight and obesity.

Methods

Study Design and Participants

This cross-sectional analysis included data collected during the baseline assessment of the ActiveBrains project [21]. A detailed description of the study design, inclusion criteria, and methods have been published elsewhere [22]. Briefly, ActiveBrains is a randomized controlled trial intended to examine the effect of a 20-week PA intervention on brain structure, brain function, cognitive performance, academic achievement, and physical and mental health outcomes in children with overweight and obesity [22]. A total of 110 children (8.5-11 years old) with overweight and obesity, defined according to the World Obesity Federation cut points [23,24], were recruited from Granada, Spain. Data were collected from November 2014 to February 2016. Parents or legal guardians were informed of the purpose of the study and written informed parental consent was obtained. The ActiveBrains project was approved by the Human Research Ethics Committee of the University of Granada and was registered as a clinical trial at ClinicalTrials.gov (NCT02295072).

Procedures

As part of the protocol of the ActiveBrains project [22], body weight and height were measured to the nearest 0.1 kg and 0.1 cm using a seca 861 electronic scale (seca gmbh & co kg) and a seca 225 precision stadiometer (seca gmbh & co kg), respectively. BMI (kg/m^2) was then calculated. Overweight and obesity were classified based on the cutoffs of the World Obesity Federation [23].

Participants' overall PA and step-based metrics were measured with a GT3X+ accelerometer (ActiGraph) worn on their right hip for 7 complete days (24-hour wear-time protocol). Participants were encouraged to wear the accelerometers as many hours as possible and only remove them for water activities (ie, shower or swimming). Concurrently, participants logged the time they went to bed and woke up in a diary for the entire 7 days. All participants with at least 4 days, including 1 weekend day, with 16 hours or more of accelerometer wear time were included in the analyses (N=105).

Data Reduction

Raw .gt3x files (100 Hz) were loaded and processed with the ActiLife software (ActiGraph) to obtain activity counts (ie, metric intended to capture body movement), accumulated in the vertical axis over 15-second epochs, and steps accumulation over 60-second epochs using the default filter developed by ActiGraph. Nonwear time was detected based on the raw acceleration values of the three axes using a previously published algorithm [25]. Briefly, each 15-minute block was classified as nonwear time if the standard deviation of two out of the three axes was lower than 13 mg during the surrounding 60-minute moving window, or if the mean acceleration for two

out of the three axes was lower than 50 mg. Likewise, sustained abnormally high accelerations (ie, higher than 5.5 g; assumed to be related to device malfunction) were detected and labelled as nonwear time. The identified nonwear time, including sustained abnormally high accelerations, was imputed with the mean acceleration value for the corresponding time period over the remaining days of recording. Sleeping hours were identified using an automated algorithm guided by participants' logged times [26] and excluded from analyses. Nonwear time and sleeping hours identification were performed using functions included in the R package GGIR (The R Foundation) [25,27].

Each 15-second epoch was classified into sedentary time or time at different PA intensities using the activity-count cut points developed by Evenson et al [28]. Specifically, these were as follows: sedentary time (≤25 counts/15 sec), light intensity (26-573 counts/15 sec), moderate intensity (574-1002 counts/15 sec), and vigorous intensity (≥1003 counts/15 sec). Daily average acceleration (counts/15 sec), time spent at light-moderate-vigorous PA (LMVPA; >25 counts/15 sec), and time spent at moderate-to-vigorous PA (MVPA) intensity (>573 counts/15 sec) were included in the analyses as indicators of overall PA. Daily average acceleration (counts/15 sec) and MVPA are indicators commonly used to represent overall PA [29-31]. LMVPA was also included following the recommendations of the 2018 Physical Activity Guidelines Advisory Committee Scientific Report, which acknowledge the importance of any kind of PA for health [32]. Furthermore, light PA could be a stimulus worthy to consider in children with overweight and obesity since they usually engage in insufficient MVPA.

Total ambulatory activity volume was derived as the number of steps per day. Furthermore, ambulatory activity cadence patterns were estimated as described previously for adults [33] and children [19]. Briefly, cadences were organized into bands of approximately 20 steps per minute increments. These cadence bands have been previously associated with the following behavioral descriptors: incidental movement (1-19 steps/min), sporadic movement (20-39 steps/min), purposeful movement (40-59 steps/min), slow walking (60-79 steps/min), medium walking (80-99 steps/min), brisk walking (100-119 steps/min), and faster walking (≥120 steps/min). Time spent in each one of these bands, as well as time at zero cadence (TZC), were computed. In addition, the peak 60-minute, peak 30-minute, and peak 1-minute cadences were computed by rank-ordering each participants' data for each day and then computing the average steps per minute for the top 60, 30, and 1 minute, respectively. The ActiGraph GT3X+ accelerometer has been demonstrated to be valid for counting steps [34,35] and its identified cadence bands have been used to describe cadence patterns in large cohorts [19]. Mean daily counts per 15 seconds, sedentary time, and time-based and step-based metrics were then calculated as follows:

(mean of available weekdays × 5) + (mean of available weekend days × 2) / 7

Data Analyses

Descriptive characteristics of participants were presented as means and SD. We used simple linear regression models to study the proportion of overall PA indicators explained by each step-based metric, and stepwise regression models to study the proportion explained by using several step-based metrics as predictors. First, the variable that explained the highest proportion of the outcome variance was introduced. Then, those variables that significantly increased the proportion of variance explained were introduced. If any of the variables presented a variance inflation factor above 7, it was excluded from the model. TZC was not included in these models since it represents inactivity. In addition, we identified those participants who achieved the World Health Organization PA recommendations for this age group [36] (ie, at least 60 min/day of MVPA). Two-sample *t* tests were then used to compare time spent in different cadence bands; the peak 60-minute, the peak 30-minute, and the peak 1-minute cadences of children who accomplished the PA recommendations were also compared, using two-sample *t* tests, with their peers who did not. All analyses were performed in R [37]. The significance level was set at $P<.05$.

Results

Descriptive Characteristics

Table 1 presents anthropometric characteristics; sedentary time; light, moderate, and vigorous PA; as well as step-based metrics for all participants stratified by sex.

Multimedia Appendix 1 (Table A1) shows the same descriptive characteristics stratified by weight status group (ie, overweight, mild obesity, severe obesity, and morbid obesity).

Table 1. Anthropometry, sedentary time, time-based physical activity (PA) metrics, and step-based metrics of overweight and obese children.

Characteristic	All participants (N=105), mean (SD)	Boys (n=62), mean (SD)	Girls (n=43), mean (SD)
Age (years)	10.1 (1.1)	10.2 (1.2)	9.9 (1.1)
Anthropometry			
Weight (kg)	56.6 (11.1)	57.4 (11.1)	55.4 (11.1)
Height (cm)	144.4 (8.3)	145.0 (7.8)	143.6 (8.9)
BMI (z-score)	3.03 (0.87)	3.19 (0.97)	2.81 (0.64)
Awake and wear time (min/day)			
Awake time	919.6 (31.5)	921.3 (28.7)	917.0 (35.2)
Wear time during waking	903.1 (35.3)	905.1 (32.7)	900.4 (39.0)
Sedentary time and PA intensities (min/day)			
Sedentary time	600.8 (69.6)	593.6 (69.1)	611.1 (69.9)
Light-intensity PA	273.2 (51.7)	276.4 (51.2)	268.5 (52.5)
Moderate-intensity PA	34.0 (11.6)	37.9 (12.4)	28.2 (7.4)
Vigorous-intensity PA	10.7 (6.7)	12.3 (7.5)	8.3 (4.4)
Moderate-to-vigorous PA	44.7 (16.8)	50.3 (18.2)	36.6 (10.3)
Step-based metrics			
Volume (steps/day)	8676.8 (2202.9)	9257.6 (2431.9)	7836.9 (1485.4)
Peak 60-minute cadence (steps/min)	63.7 (13.6)	66.3 (14.4)	59.8 (11.4)
Peak 30-minute cadence (steps/min)	78.0 (14.5)	79.7 (15.2)	75.4 (13.2)
Peak 1-minute cadence (steps/min)	111.5 (13.3)	111.2 (13.3)	111.8 (13.3)
Time spent at different cadence bands (min/day)			
0 steps/minute	346.6 (78.1)	343.0 (79.5)	351.7 (76.7)
1-19 steps/minute	439.0 (63.4)	434.9 (62.2)	444.8 (65.4)
20-39 steps/minute	71.9 (18.2)	73.5 (19.1)	69.7 (16.8)
40-59 steps/minute	27.5 (9.2)	30.3 (9.8)	23.6 (6.6)
60-79 steps/minute	15.9 (7.9)	18.6 (8.5)	12.0 (4.6)
80-99 steps/minute	10.2 (6.2)	11.8 (7.0)	8.0 (3.8)
100-119 steps/minute	6.6 (6.0)	7.4 (7.0)	5.5 (4.0)
≥120 steps/minute	1.6 (2.6)	1.8 (2.7)	1.3 (2.3)

Proportion of Total Physical Activity Explained by Step-Based Metrics

Figure 1 depicts the proportion of variance in indicators of overall PA (ie, counts/15 sec, LMVPA, and MVPA) that each step-based metric explained (r^2) in separate linear regression models (ie, simple linear regression with each step-based metric as predictor and overall PA metric as outcome). Among the step-based metrics, steps per day explained the highest proportion of counts per 15 seconds and MVPA (66% and 74%, respectively), while time at 1-19 steps per minute explained the highest proportion of LMVPA (52%). Overall, peak cadence indicators explained a lower proportion of the variance than steps per day in overall PA indicators. Likewise, the shorter the time intervals used to calculate the specific peak cadence indicator, the lower the explanation capacity of the metric, which is to be expected given the shorter time frame represented (ie, 60 min > 30 min > 1 min).

Figure 1. Proportion of variance (r^2) in overall physical activity indicators, which is explained independently by each step-based metric. LMVPA: light-moderate-vigorous physical activity; MVPA: moderate-to-vigorous physical activity.

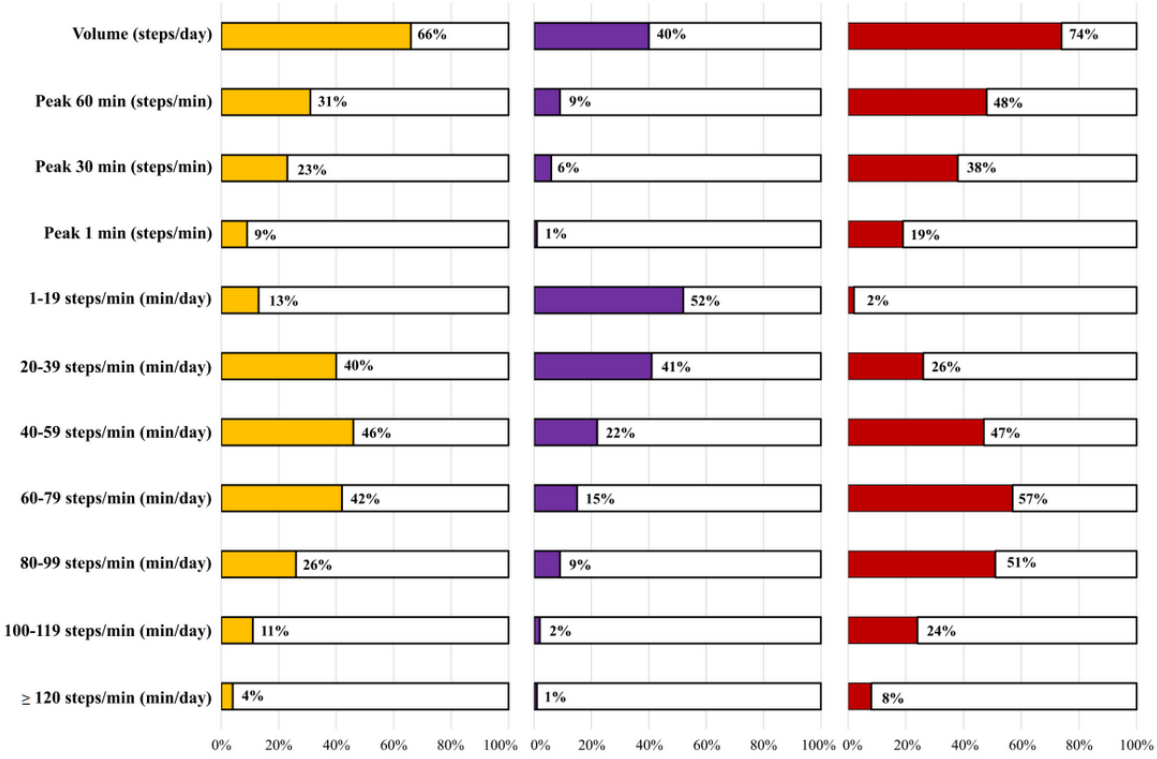
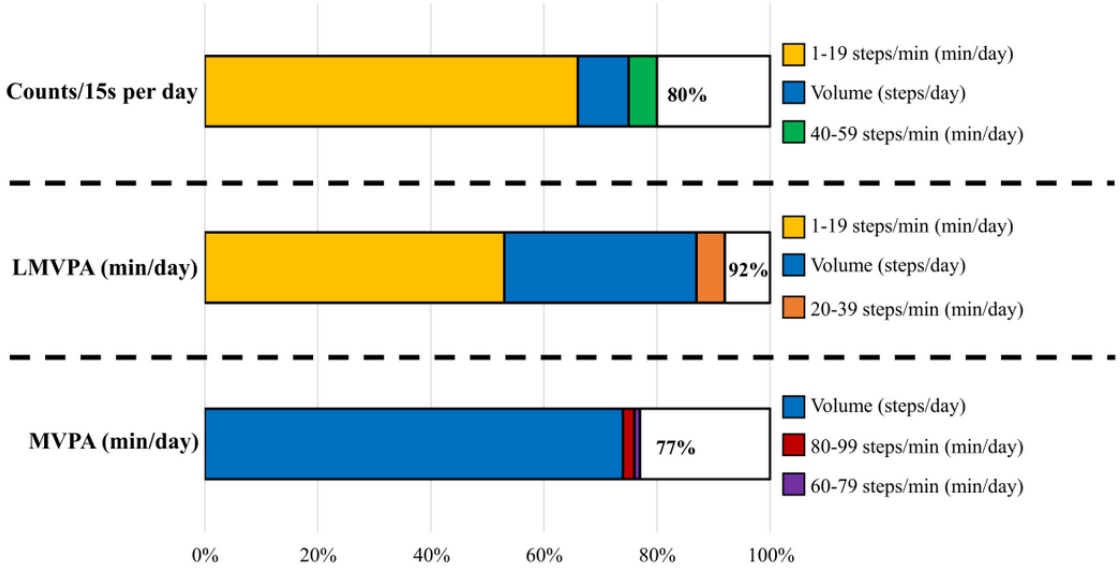


Figure 2. Proportion of variance (r^2) in overall physical activity (PA) indicators (ie, counts/15 sec, light-moderate-vigorous PA [LMVPA], and moderate-to-vigorous PA [MVPA]), which is explained by a combination of step-based metrics, calculated using stepwise linear regressions. All predictors presented variance inflation factors of <6 in the selected models.



Cadence Patterns According to Physical Activity Guidelines

Out of the 105 children, 20 (19.0%) achieved the recommended amount of MVPA (ie, ≥60 min/day). Children who performed less than 60 minutes per day of MVPA also had significantly higher values for TZC (P=.004) and less time in cadence bands from 20 to 120 steps per minute, compared with those who performed 60 minutes per day or more of MVPA (all P<.02) (see Figure 3). Likewise, peak 60-minute, peak 30-minute, and peak 1-minute cadences were higher in children who achieved the 60 minutes per day of MVPA.

Specifically, participants had to walk around 11,000 steps per day to achieve the recommended dose of MVPA (see Figure 4, panel A). Likewise, they had to spend 105 minutes per day walking at 20-39 steps per minute, 40 minutes per day at 40-59 steps per minute, 25 minutes per day at 60-79 steps per minute, 19 minutes per day at 80-99 steps per minute, 18 minutes per

day at 100-119 steps per minute, or 10 minutes per day at or above 120 steps per minute (see Figure 4, panel B). Finally, their peak cadences had to be higher than 140 steps per minute for peak 1-minute cadence, 100 steps per minute for peak 30-minute cadence, or 80 steps per minute for peak 60-minute cadence (see Figure 4, panel C).

Figure 3. Time spent in each cadence band (panel A) and peak cadences (panel B) across children meeting or not meeting the physical activity guidelines (ie, ≥60 minutes of MVPA/day). Error bars represent SDs. MVPA: moderate-to-vigorous physical activity.

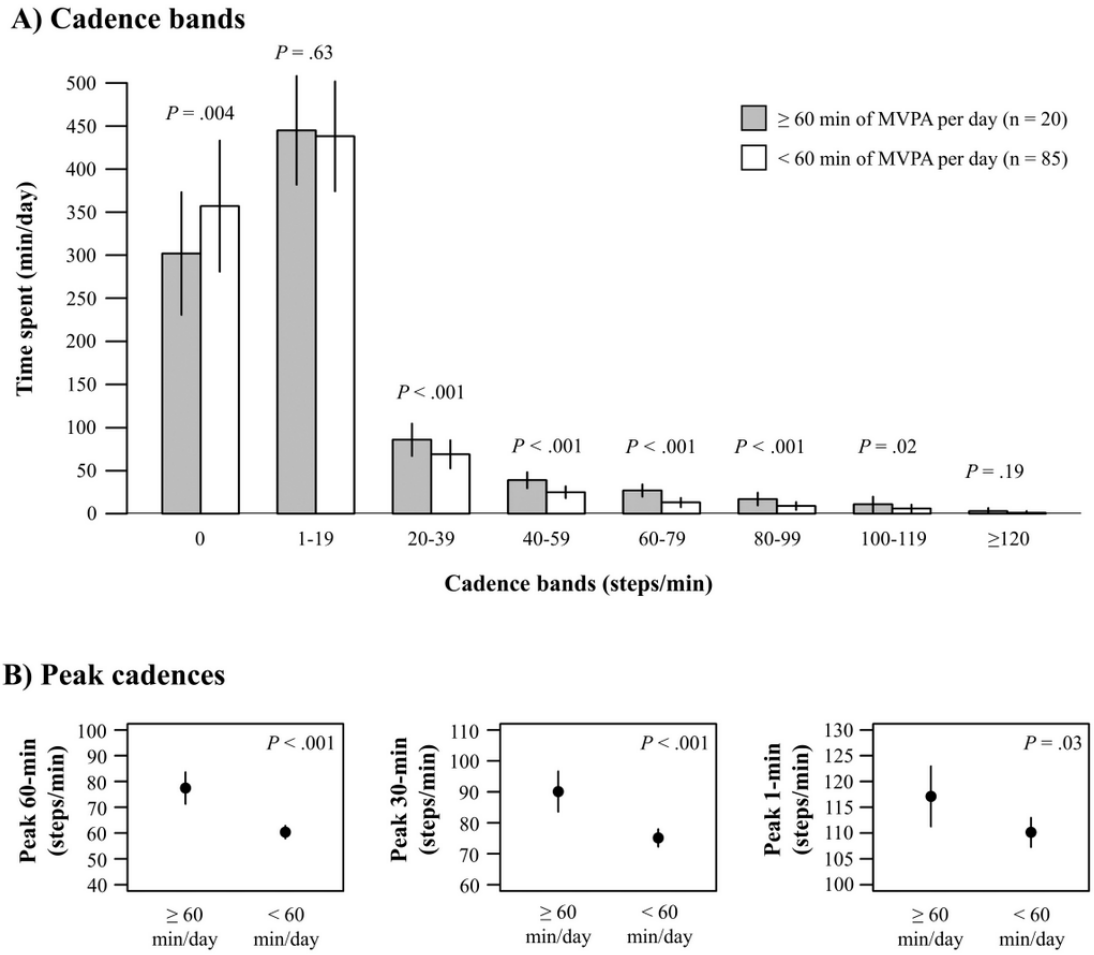
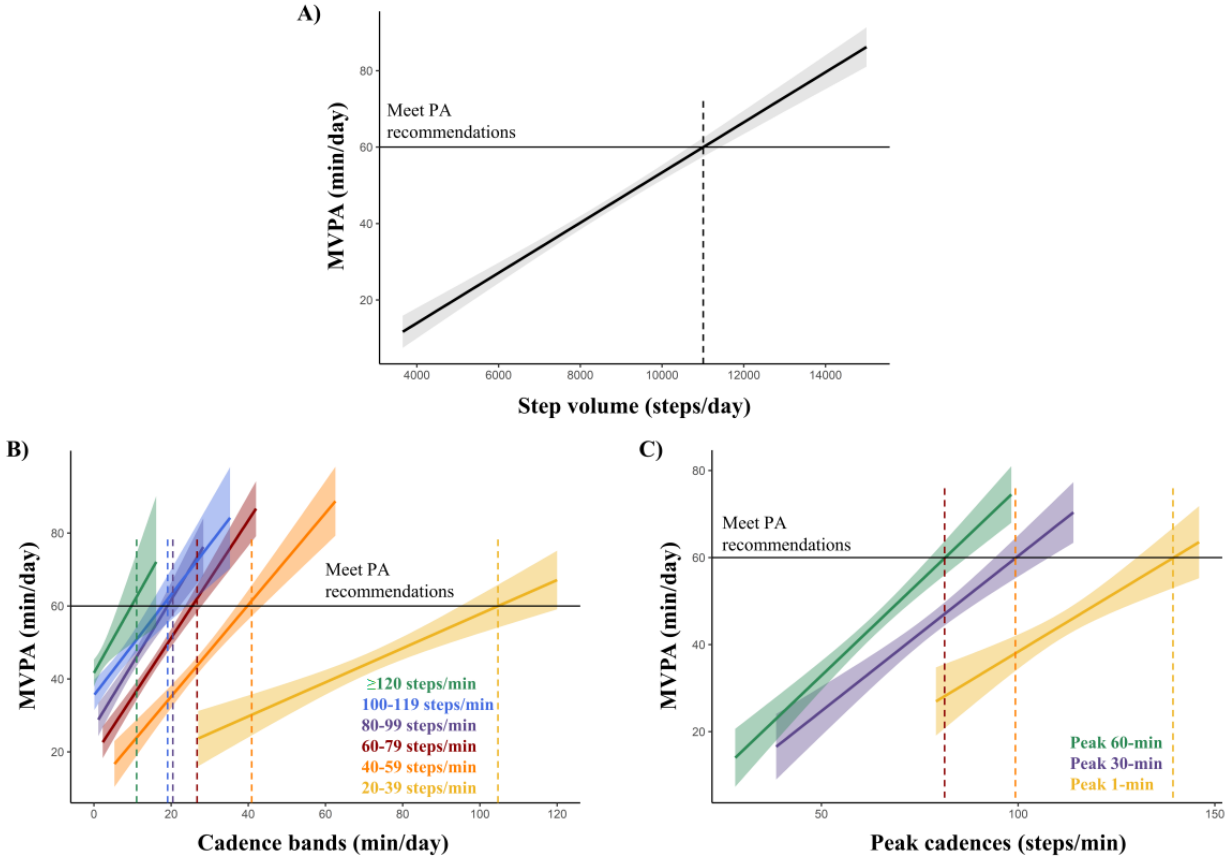


Figure 4. Linear regression line, with 95% CI (shaded areas), for the association between moderate-to-vigorous physical activity (MVPA) and step-based metrics: step volume (panel A), cadence bands (panel B), and peak cadences (panel C). PA: physical activity.



Discussion

The main findings of this study are as follows:

1. Steps per day and the 1-19 steps per minute cadence band explained the greatest amount of overall PA (ie, counts/15 sec, 66%; LMVPA, 52%; and MVPA, 74%) in children with overweight and obesity. The proportion of variance explained was further improved—by up to 77%-92%—by adding other step-based metrics to the models.
2. Cadence-based step patterns significantly differed between those children with overweight and obesity who achieved the PA recommendations compared with those who did not. Together, these findings seem to point out ambulatory activity as a major source of PA in children with overweight and obesity, as it has been previously reported in adults [16,17].

Further studies with larger and more representative samples should corroborate this finding. This finding can be leveraged to design appropriate PA interventions (ie, by investigating the amount of walking at a certain intensity needed to meet PA recommendations) as a strategy to lower lifespan health risks in this vulnerable population.

Steps per day and the 1-19 steps per minute cadence band were the best explanatory factors of overall PA. Specifically, more than half of the variation in overall PA could be explained by either steps per day or the 1-19 steps per minute cadence band in children with overweight and obesity, depending on the overall PA indicator regressed (ie, 66% for counts/15 sec [steps/day], 52% for LMVPA [1-19 steps/min], and 74% for MVPA [steps/day]). Furthermore, all of the stepwise models included steps per day to estimate either counts per 15 seconds, LMVPA, or MVPA. Accordingly, our sample was active for 5.3 hours per day (ie, LMVPA), during which time they spent around 2.2 hours per day in ambulatory activity (ie, from sporadic movement to faster cadences). This presumes around 40% of the time spent was in LMVPA, which is similar to the estimation obtained to predict LMVPA from steps per day (ie, 40%).

However, steps per day was not the only important factor in the prediction of overall PA indicators. Information regarding step accumulation pattern increased the prediction capacity up to 80%, 92%, and 77%, for counts per 15 seconds, LMVPA, and MVPA, respectively. These findings support the concept of also considering stepping rate, which has been associated with health-related intensity levels in children [20] in addition to steps per day. The fact that the explanation of the variability of LMVPA increased substantially (ie, from 40% to 92%) by including more metrics in the stepwise models is noteworthy. This suggests that considering both steps per day and certain cadence bands we can explain around 90% of their active minutes. However, the explanation of the variability of MVPA only increased from 74% to 77%, which means that almost all information on MVPA is already provided by steps per day. Together, it seems clear that step-based metrics are more powerful in explaining light-intensity PA than higher intensities.

To this end, Tudor-Locke et al found that walking at 115 steps per minute requires an energy expenditure of approximately 4 metabolic equivalents (ie, moderate PA intensity for children) in 9-11-year-old children, measured while walking on a treadmill [38]. However, the cadence-intensity relationship observed under laboratory-controlled conditions may not be generalizable to free-living data from children with overweight and obesity. Likewise, caution is advised since measurement tools differed between studies (ie, direct observation vs accelerometers). We observed around 34 minutes per day classified as moderate PA intensity by Evenson et al cut points [28] and, in turn, around only 2 minutes per day (SD 3) accumulated at a cadence of more than 115 steps per minute, which is indicative of MVPA intensity in this age group as measured in lab conditions [38]. A source for this difference could be the epoch length used to derive moderate intensity [39] (ie, 15 sec for Evenson et al cut points and 60 sec for time spent above 100 steps/min). Estimations based on Evenson cut points could be able to capture short bouts of moderate PA up to 15 seconds, while step-based estimations of moderate PA are limited to those bouts lasting at least 1 minute. We decided to use 60-second epochs for cadence to maintain consistency with previous studies, to ease comparability of findings and because there are no studies examining the cadence measured in 15-second epochs and intensity to date, making it more difficult to interpret the findings. It could be also argued that most of the moderate PA performed by our sample was not related to ambulatory activity, which seems unlikely because step-based metrics explained almost 80% of the variance in MVPA. We must also acknowledge that metabolic intensity is indirectly inferred from detected movement signals and is not a clear indicator of metabolic cost, so there are likely to be measurement differences attributable to differential definitions. Therefore, further research is needed to understand how free-living cadence bands relate to energy expenditure and accelerometer signals.

Peak cadence indices and cadence bands have been previously used as proxy indicators for ambulatory activity intensity and pattern, respectively, in children [19] and adults [33,40]. In congruence with Barreira et al [19], we found that most of the day was spent in low intensity or sedentary behaviors. Specifically, we found around 10 hours per day of sedentary time using Evenson et al [28] cut points or, in regard to step-based metrics, a TZC value of 5.8 hours per day and 7.3 hours per day in incidental movements (1-19 steps/min). Barreira et al reported similar step accumulation patterns in 6-11-year-old children from the US National Health and Nutrition Examination Survey (NHANES 2005-2006). Notably, only 38% of the NHANES population−based sample were overweight and obese [19]. In contrast, our sample of children who were overweight and obese accumulated more TZC, as well as time in incidental movements (1-19 steps/min), and less time in cadence bands, from sporadic movement to faster walking (20-120 steps/min). Likewise, differences in accelerometer models, study design, and socioenvironmental context should be considered when comparing these studies.

According to the Evenson et al cut-points definition of MVPA [28], 20 children out of 105 (19.0%) from our sample met the PA recommendation of at least 60 minutes per day of MVPA [36]. This finding should be interpreted with caution since quantification of time-based PA with accelerometers is notoriously challenging and is dependent on a variety of data collection and processing decisions [39], including those related to selecting appropriate analytical cut points [41]. We have previously reported that changing cut points can derive extremely different estimations of the proportion of children meeting PA recommendations in this sample [41]. It is also important to consider that PA recommendations are mainly based on self-reported data, which could bias interpretation over objective data. Additionally, when compared to normative values from NHANES 2005-2006 [42], our sample can be considered *below average* for steps per day for 8-9-year-old children (ie, 7647-9398 steps/day) or *average* for 10-11-year-olds (ie, 8504-10,066 steps/day). Likewise, our sample presented *below average* values for peak 60-minute cadence (ie, 62-71 steps/min). Nevertheless, a large proportion of the count-based MVPA performed was related to step-based metrics, which suggests that limited ambulatory behaviors could be responsible for the low prevalence of children meeting the PA recommendations. Furthermore, we found significant differences in ambulatory activity intensity (ie, time spent in almost every band cadence was significantly different) between those who met and those who did not meet the PA recommendations. Specifically, children who met the recommendations spent around 55 minutes per day less in TZC, 17 minutes per day more in sporadic movement (ie, 20-39 steps/min), 14 minutes per day more in purposeful movement (ie, 40-59 steps/min), 13 minutes per day more in slow walking (ie, 60-79 steps/min), 8 minutes per day more in medium walking (ie, 80-99 steps/min), and 5 minutes per day more in brisk walking (ie, 100-119 steps/min). Additionally, peak 1-minute, 30-minute, and 60-minute cadences seem to be able to discern between children achieving or not achieving the recommended dose of MVPA per day (ie, 60 min/day).

Findings of this study have several practical implications to consider. Two examples are as follows:

1. As a large proportion of overall PA identified by accelerometers is explained by step-based metrics in children with overweight and obesity, these measures could be used to describe and compare PA patterns in this population.
2. It could be assumed that increasing ambulatory activity volume and intensity is a feasible form of PA that can increase the chances of meeting PA recommendations in this population. This is especially important to consider as ambulatory activity is a feasible PA strategy that may lead to several health benefits (eg, improved body composition and mental health) in children with overweight and obesity [43-45]. Notably, walking does not require complex movement skills, so it can be performed by most populations, including children with overweight and obesity who frequently do not engage in sports because of their low physical competence [14].

Several limitations should be acknowledged. First and foremost, accelerometer measurements of PA are influenced by a variety of data collection and processing decisions [39]. This means that it cannot be considered a gold standard for overall PA

measurement and that changes in the quantification of PA could change the findings observed in this study. However, we were as consistent as possible regarding the measurement of both overall PA and step-based metrics. Both outcomes come from the same hip-worn accelerometer, and cut points used are based on the vertical axis acceleration, which is consistent with the ActiGraph procedures to detect steps. This would reduce the methodological inconsistencies between the overall PA and the step-based metrics estimations, which, in turn, can be considered as a strength of this study. Note that epoch length discrepancies between count-based and step-based metrics may be partially responsible for the differences observed. However, our findings should be interpreted with caution since overall PA refers to accelerometer-determined PA, which is not a gold standard and could ignore certain activities such as swimming. Note that step-based metrics derived from pedometers could vary the findings from this study and their relationship with overall PA should be investigated. Likewise, another strength to highlight is that we are focusing on a population who may benefit greatly from increases in ambulatory activity; for example, this study demonstrates that they could have substantially increased chances of meeting PA recommendations by only focusing on ambulatory activity.

In conclusion, step-based metrics including steps per day and various cadence-based intensity indicators seem to capture the majority of PA in children with overweight and obesity. Given that pedometers are more affordable than accelerometers, step-based metrics could be useful for discriminating between those children who do or do not achieve MVPA recommendations.

Acknowledgments

This study is part of a PhD thesis conducted in the Biomedicine Doctoral Studies program at the University of Granada, Spain. This study was conducted under the umbrella of the ActiveBrains and the SmarterMove projects supported by the MINECO (Ministerio de Economía y COmpetitividad)/FEDER (Fondo Europeo de DEsarrollo Regional) (DEP2013-47540, DEP2016-79512-R, and RYC-2011-09011). JHM and JMG are supported by grants from the Spanish Ministry of Education, Culture and Sport (FPU15/02645 and FPU14/06837). CCS is supported by a grant from the Spanish Ministry of Economy and Competitiveness (BES-2014-068829). PMG is supported by a grant from the European Union's Horizon 2020 research and innovation program (No. 667302). This study has been partially funded by the University of Granada, Plan Propio de Investigación 2016, Excellence actions: Units of Excellence; Unit of Excellence on Exercise and Health (UCEES), and by the Junta de Andalucía, Consejería de Conocimiento, Investigación y Universidades, and European Regional Development Fund (ERDF) (reference SOMM17/6107/UGR and reference RD16/0022). In addition, funding was provided by the SAMID III (red de SAlud Materno Infantil y Desarrollo) network, RETICS (REdes Temáticas de Investigación Cooperativa en Salud), funded by the PN (Plan Nacional) I+D+I (Investigación + Desarrollo + Innovación) 2017-2021 (Spain), ISCIII (Instituto de Salud Carlos III)–Sub-Directorate General for Research Assessment and Promotion and the EXERNET Research Network on Exercise and Health in Special Populations (DEP2005-00046/ACTI), and the European Union's 2020 research and innovation program under grant agreement No. 667302. Likewise, this research was supported, in part, by the Intramural Research Program at the National Institute on Aging, USA.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Anthropometry, sedentary time, time-based physical activity metrics, and step-based metrics of overweight and obese children stratified by weight status, according to the World Obesity Federation standards.
[DOCX File , 18 KB - mhealth_v8i4e14841_app1.docx]

References

1. Booth FW, Laye MJ, Lees SJ, Rector RS, Thyfault JP. Reduced physical activity and risk of chronic disease: The biology behind the consequences. Eur J Appl Physiol 2008 Mar;102(4):381-390. [doi: 10.1007/s00421-007-0606-5] [Medline: 17987311]
2. Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, Brauer M, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Dec 05;386(10010):2287-2323 [FREE Full text] [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
3. Myers J, McAuley P, Lavie CJ, Despres J, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: Their independent and interwoven importance to health status. Prog Cardiovasc Dis 2015;57(4):306-314. [doi: 10.1016/j.pcad.2014.09.011] [Medline: 25269064]
4. 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]

5. 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]
6. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: Global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
7. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
8. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
9. Dennison BA, Erb TA, Jenkins PL. Television viewing and television in bedroom associated with overweight risk among low-income preschool children. Pediatrics 2002 Jun;109(6):1028-1035. [doi: 10.1542/peds.109.6.1028] [Medline: 12042539]
10. Vale S, Santos R, Silva P, Soares-Miranda L, Mota J. Relationship of objective measurement of physical activity during school hours and BMI in preschool children. Int J Pediatr Obes 2011 Oct;6 Suppl 2:37-38. [doi: 10.3109/17477166.2011.613663] [Medline: 21923294]
11. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
12. Ding D, Nguyen B, Learnihan V, Bauman AE, Davey R, Jalaludin B, et al. Moving to an active lifestyle? A systematic review of the effects of residential relocation on walking, physical activity and travel behaviour. Br J Sports Med 2018 Jun;52(12):789-799. [doi: 10.1136/bjsports-2017-098833] [Medline: 29858466]
13. Kobel S, Kettner S, Kesztyüs D, Erkelenz N, Drenowatz C, Steinacker JM. Correlates of habitual physical activity and organized sports in German primary school children. Public Health 2015 Mar;129(3):237-243 [FREE Full text] [doi: 10.1016/j.puhe.2014.12.002] [Medline: 25700788]
14. Cliff DP, Okely AD, Morgan PJ, Jones RA, Steele JR, Baur LA. Proficiency deficiency: Mastery of fundamental movement skills and skill components in overweight and obese children. Obesity (Silver Spring) 2012 May;20(5):1024-1033 [FREE Full text] [doi: 10.1038/oby.2011.241] [Medline: 21799480]
15. Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chastin SF, et al. Utilization and harmonization of adult accelerometry data: Review and expert consensus. Med Sci Sports Exerc 2015 Oct;47(10):2129-2139 [FREE Full text] [doi: 10.1249/MSS.0000000000000661] [Medline: 25785929]
16. Welk GJ, Kim Y. Context of physical activity in a representative sample of adults. Med Sci Sports Exerc 2015 Oct;47(10):2102-2110 [FREE Full text] [doi: 10.1249/MSS.0000000000000641] [Medline: 25699482]
17. Tudor-Locke C, Ham SA. Walking behaviors reported in the American Time Use Survey 2003-2005. J Phys Act Health 2008 Sep;5(5):633-647. [doi: 10.1123/jpah.5.5.633] [Medline: 18820341]
18. Tudor-Locke C, Rowe DA. Using cadence to study free-living ambulatory behaviour. Sports Med 2012 May 01;42(5):381-398. [doi: 10.2165/11599170-000000000-00000] [Medline: 22462794]
19. Barreira TV, Katzmarzyk PT, Johnson WD, Tudor-Locke C. Cadence patterns and peak cadence in US children and adolescents: NHANES, 2005-2006. Med Sci Sports Exerc 2012 Sep;44(9):1721-1727. [doi: 10.1249/MSS.0b013e318254f2a3] [Medline: 22453245]
20. Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna JM, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: A narrative review. Br J Sports Med 2018 Jun;52(12):776-788 [FREE Full text] [doi: 10.1136/bjsports-2017-097628] [Medline: 29858465]
21. Ortega FB. PROFITH. ActiveBrains URL: http://profith.ugr.es/activebrains?lang=en [accessed 2019-10-24]
22. Cadenas-Sánchez C, Mora-González J, Migueles JH, Martín-Matillas M, Gómez-Vida J, Escolano-Margarit MV, et al. An exercise-based randomized controlled trial on brain, cognition, physical health and mental health in overweight/obese children (ActiveBrains project): Rationale, design and methods. Contemp Clin Trials 2016 Mar;47:315-324. [doi: 10.1016/j.cct.2016.02.007] [Medline: 26924671]
23. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes 2012 Aug;7(4):284-294. [doi: 10.1111/j.2047-6310.2012.00064.x] [Medline: 22715120]
24. Bervoets L, Massa G. Defining morbid obesity in children based on BMI 40 at age 18 using the extended international (IOTF) cut-offs. Pediatr Obes 2014 Oct;9(5):e94-e98. [doi: 10.1111/j.2047-6310.2014.00217.x] [Medline: 24578314]
25. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
26. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One 2015;10(11):e0142533 [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]

27. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. J Meas Phys Behav 2019;2(3):188-196. [doi: 10.1123/jmpb.2018-0063]
28. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008 Dec;26(14):1557-1565. [doi: 10.1080/02640410802334196] [Medline: 18949660]
29. Fairclough SJ, Taylor S, Rowlands AV, Boddy LM, Noonan RJ. Average acceleration and intensity gradient of primary school children and associations with indicators of health and well-being. J Sports Sci 2019 Sep;37(18):2159-2167. [doi: 10.1080/02640414.2019.1624313] [Medline: 31156048]
30. Delisle Nyström C, Pomeroy J, Henriksson P, Forsum E, Ortega FB, Maddison R, et al. Evaluation of the wrist-worn ActiGraph wGT3x-BT for estimating activity energy expenditure in preschool children. Eur J Clin Nutr 2017 Oct;71(10):1212-1217. [doi: 10.1038/ejcn.2017.114] [Medline: 28745334]
31. Gomez-Marcos MA, Recio-Rodríguez JI, Patino-Alonso MC, Martinez-Vizcaino V, Martin-Borras C, de-la-Cal-Dela-Fuente A, EVIDENT Study Investigators. Relationship between physical activity and plasma fibrinogen concentrations in adults without chronic diseases. PLoS One 2014;9(2):e87954 [FREE Full text] [doi: 10.1371/journal.pone.0087954] [Medline: 24498413]
32. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: US Department of Health and Human Services; 2018. URL: https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf [accessed 2020-03-29]
33. Tudor-Locke C, Camhi SM, Leonardi C, Johnson WD, Katzmarzyk PT, Earnest CP, et al. Patterns of adult stepping cadence in the 2005-2006 NHANES. Prev Med 2011 Sep;53(3):178-181. [doi: 10.1016/j.ypmed.2011.06.004] [Medline: 21708187]
34. De Craemer M, De Decker E, Santos-Lozano A, Verloigne M, De Bourdeaudhuij I, Deforche B, et al. Validity of the Omron pedometer and the actigraph step count function in preschoolers. J Sci Med Sport 2015 May;18(3):289-293. [doi: 10.1016/j.jsams.2014.06.001] [Medline: 24994695]
35. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015 Apr;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]
36. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: WHO Press; 2010 Jan. URL: https://apps.who.int/iris/bitstream/handle/10665/44399/9789241599979_eng.pdf?sequence=1 [accessed 2020-03-29]
37. The Comprehensive R Archive Network (CRAN). URL: https://cran.r-project.org/ [accessed 2020-03-29]
38. Tudor-Locke C, Schuna JM, Han H, Aguiar E, Larrivee S, Hsia D, et al. Cadence (steps/min) and intensity during ambulation in 6-20 year olds: The CADENCE-kids study. Int J Behav Nutr Phys Act 2018 Feb 26;15(1):20 [FREE Full text] [doi: 10.1186/s12966-018-0651-y] [Medline: 29482554]
39. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
40. Tudor-Locke C, Brashear MM, Katzmarzyk PT, Johnson WD. Peak stepping cadence in free-living adults: 2005-2006 NHANES. J Phys Act Health 2012 Nov;9(8):1125-1129. [doi: 10.1123/jpah.9.8.1125] [Medline: 22207030]
41. Migueles JH, Cadenas-Sanchez, Tudor-Locke C, Löf M, Esteban-Cornejo I, Molina-Garcia P, et al. Comparability of published cut-points for the assessment of physical activity: Implications for data harmonization. Scand J Med Sci Sports 2019 Apr;29(4):566-574. [doi: 10.1111/sms.13356] [Medline: 30548545]
42. Barreira TV, Schuna JM, Mire EF, Broyles ST, Katzmarzyk PT, Johnson WD, et al. Normative steps/day and peak cadence values for United States children and adolescents: National Health and Nutrition Examination Survey 2005-2006. J Pediatr 2015 Jan;166(1):139-143. [doi: 10.1016/j.jpeds.2014.09.014] [Medline: 25311710]
43. Barreira TV, Katzmarzyk PT, Johnson WD, Tudor-Locke C. Walking cadence and cardiovascular risk in children and adolescents: NHANES, 2005-2006. Am J Prev Med 2013 Dec;45(6):e27-e34. [doi: 10.1016/j.amepre.2013.08.005] [Medline: 24237926]
44. Martínez-Gómez D, Ruiz JR, Gómez-Martínez S, Chillón P, Rey-López JP, Díaz LE, AVENA Study Group. Active commuting to school and cognitive performance in adolescents: The AVENA study. Arch Pediatr Adolesc Med 2011 Apr;165(4):300-305. [doi: 10.1001/archpediatrics.2010.244] [Medline: 21135316]
45. Mendoza JA, Liu Y. Active commuting to elementary school and adiposity: An observational study. Child Obes 2014 Feb;10(1):34-41 [FREE Full text] [doi: 10.1089/chi.2013.0133] [Medline: 24443901]

Abbreviations

ERDF: European Regional Development Fund
FEDER: Fondo Europeo de DEsarrollo Regional
I+D+I: Investigación + Desarrollo + Innovación
ISCIII: Instituto de Salud Carlos III

LMVPA: light-moderate-vigorous physical activity
MINECO: Ministerio de Economía y COmpetitividad
MVPA: moderate-to-vigorous physical activity
NHANES: National Health and Nutrition Examination Survey
PA: physical activity
PN: Plan Nacional
RETICS: REdes Temáticas de Investigación Cooperativa en Salud
SAMID III: red de SAlud Materno Infantil y Desarrollo
TZC: time at zero cadence
UCEES: Unit of Excellence on Exercise and Health

Edited by G Eysenbach; submitted 28.05.19; peer-reviewed by C Moore, G Hu, B Larsen, M Gaman, A Tathe; comments to author 30.09.19; revised version received 06.11.19; accepted 16.12.19; published 28.04.20.

Please cite as:
Migueles JH, Cadenas-Sanchez C, Aguiar EJ, Molina-Garcia P, Solis-Urra P, Mora-Gonzalez J, García-Mármol E, Shiroma EJ, Labayen I, Chillón P, Löf M, Tudor-Locke C, Ortega FB
Step-Based Metrics and Overall Physical Activity in Children With Overweight or Obesity: Cross-Sectional Study
JMIR Mhealth Uhealth 2020;8(4):e14841
URL: http://mhealth.jmir.org/2020/4/e14841/
doi:10.2196/14841
PMID:32343251

Corrigenda and Addenda

Correction: A Smartphone App to Improve Medication Adherence in Patients With Type 2 Diabetes in Asia: Feasibility Randomized Controlled Trial

Zhilian Huang[1,2], MPH; Eberta Tan[3], MRCP (UK), MMed, MBBS; Elaine Lum[1,4,5], PhD, MClinPharm, BPharm; Peter Sloot[6,7,8], PhD; Bernhard Otto Boehm[9,10], FRCP, MD; Josip Car[1], FRCP, FFPH, MSc, PhD, MD

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[2]NTU Institute for Health Technologies, Interdisciplinary Graduate School, Nanyang Technological University, Singapore, Singapore
[3]Department of Endocrinology, Changi General Hospital, Singapore, Singapore
[4]Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia
[5]School of Clinical Sciences, Faculty of Health, Queensland University of Technology, Queensland, Australia
[6]Complexity Institute, Nanyang Technological University, Singapore, Singapore
[7]Institute for Advanced Science, University of Amsterdam, Amsterdam, Netherlands
[8]ITMO University, Saint Petersburg, Russian Federation
[9]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[10]Department of Endocrinology, Tan Tock Seng Hospital, Singapore, Singapore

Corresponding Author:
Zhilian Huang, MPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
Clinical Sciences Building, Level 18
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 65141221
Email: zhuang014@e.ntu.edu.sg

Related Article:

Correction of: https://mhealth.jmir.org/2019/9/e14914/

(JMIR Mhealth Uhealth 2020;8(4):e18411) doi:10.2196/18411

The authors of "A Smartphone App to Improve Medication Adherence in Patients With Type 2 Diabetes in Asia: Feasibility Randomized Controlled Trial" (JMIR Mhealth Uhealth 2019;7(9):e14914) noticed an error in the Results section of their published article. Under the "Adherence to Trial Participation" subsection of the Results section, the following sentence:

> *Eight participants had 100% adherence for the first 2 weeks of the intervention, which was decreased to 4% by the third week of the intervention.*

Has been changed to:

> *Eight participants had 100% adherence for the first 2 weeks of the intervention, which was decreased to four participants by the third week of the intervention.*

This change does not affect any of the data presented in the Results section of the paper.

The correction will appear in the online version of the paper on the JMIR website on April 29, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.02.20; this is a non–peer-reviewed article;accepted 25.02.20; published 29.04.20.

Please cite as:
Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J
Correction: A Smartphone App to Improve Medication Adherence in Patients With Type 2 Diabetes in Asia: Feasibility Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(4):e18411
URL: https://mhealth.jmir.org/2020/4/e18411
doi:10.2196/18411
PMID:32348282

Corrigenda and Addenda

Correction: Treatment Adherence and Secondary Prevention of Ischemic Stroke Among Discharged Patients Using Mobile Phone- and WeChat-Based Improvement Services: Cohort Study

Yuanjin Zhang[1], MD; Dongsheng Fan[1], MD, PhD; Hong Ji[2], PhD; Shudong Qiao[3], MD; Xia Li[3], MD

[1]Neurology Department, Peking University Third Hospital, Beijing, China

[2]Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

[3]Neurology Department, Peking University Shougang Hospital, Beijing, China

Corresponding Author:
Dongsheng Fan, MD, PhD
Neurology Department
Peking University Third Hospital
49 North Garden Road
Haidian District
Beijing
China
Phone: 86 01082265694
Email: dsfan2010@aliyun.com

Related Article:

Correction of: https://mhealth.jmir.org/2020/4/e16496/

(JMIR Mhealth Uhealth 2020;8(4):e19454) doi:10.2196/19454

The authors of "Treatment Adherence and Secondary Prevention of Ischemic Stroke Among Discharged Patients Using Mobile Phone- and WeChat-Based Improvement Services: Cohort Study" (JMIR Mhealth Uhealth 2020;8(4):e16496) noticed several errors in their published article. The following corrections have been implemented:

In the Abstract, the sentence:

> *Of the platform registry participants, 88.9% (209/234: 167 hospital-based and 42 community-based participants) adhered to inputting information into WeChat for 8-96 weeks.*

Has been corrected to:

> *Of the platform registry participants, 89.7% (210/234: 167 hospital-based and 43 community-based participants) adhered to inputting information into WeChat for 8-96 weeks.*

This change does not impact the findings of the paper.

The following additional changes have also been made within the manuscript. The caption of Figure 3 has been corrected from "Untitled." to:

> *Synchronous web terminal of the WeChat information platform part 1.*

Figure 4 was formerly captioned:

> *Synchronous web terminal of the WeChat information platform.*

This has been revised to:

> *Synchronous web terminal of the WeChat information platform part 2.*

Figure 6 was formerly captioned:

> *Flowchart of the study design. PUSH: Peking University Shougang Hospital; PUTH: Peking University Third Hospital; ITT: intention-to-treat.*

This has been revised to:

> *Selection bias analysis of the WeChat platform registry population. DM: diabetes mellitus; HLP: hyperlipidemia; HP: hypertension; TIA: transient ischemic attack.*

Figure 7 was formerly captioned:

> *Selection bias analysis of the WeChat platform registry population. DM: diabetes mellitus; HLP: hyperlipidemia; HP: hypertension; TIA: transient ischemic attack.*

This has been revised to:

> *Participants' usage of the WeChat-based modules.*

Figure 8 was formerly captioned:

> *Participants' usage of the WeChat-based modules.*

This has been revised to:

Survival functions: Kaplan-Meier analysis of the rate of endpoint events during the 1-year follow-up.

Finally, Figure 9 was formerly captioned:

Survival functions: Kaplan-Meier analysis of the rate of endpoint events during the 1-year follow-up.

This has been revised to:

Demographics of the participants with good and poor adherence.

The correction will appear in the online version of the paper on the JMIR website on April 29, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.04.20; this is a non–peer-reviewed article;accepted 24.04.20; published 29.04.20.

Please cite as:
Zhang Y, Fan D, Ji H, Qiao S, Li X
Correction: Treatment Adherence and Secondary Prevention of Ischemic Stroke Among Discharged Patients Using Mobile Phone- and WeChat-Based Improvement Services: Cohort Study
JMIR Mhealth Uhealth 2020;8(4):e19454
URL: http://mhealth.jmir.org/2020/4/e19454/
doi:10.2196/19454
PMID:32348276

Corrigenda and Addenda

Correction: Use of Health Apps by Nurses for Professional Purposes: Web-Based Survey Study

Miguel Angel Mayer[1*], MPH, PhD, MD; Octavi Rodríguez Blanco[2*], RNC, MSci; Antonio Torrejon[2*], RNC, BSciS, MSci

[1]Research Programme on Biomedical Informatics. Hospital del Mar Medical Research Institute, Universitat Pompeu Fabra, Barcelona, Spain
[2]Col·legi Oficial d'Infermeres i Infermers de Barcelona, Barcelona, Spain
[*]all authors contributed equally

Corresponding Author:
Antonio Torrejon, RNC, BSciS, MSci
Col·legi Oficial d'Infermeres i Infermers de Barcelona
Pujades 350
Barcelona, 08019
Spain
Phone: 34 932128108
Email: atorrejon@coib.cat

Related Article:

Correction of: https://mhealth.jmir.org/2019/11/e15195

(JMIR Mhealth Uhealth 2020;8(4):e18516) doi:10.2196/18516

In the manuscript “Use of Health Apps by Nurses for Professional Purposes: Web-Based Survey Study” (JMIR Mhealth Uhealth 2019;7(11):e15195) there was an error in the original published PDF. The original published PDF version of the manuscript available on the JMIR mHealth and uHealth website displayed the title as “Use of Health Apps by Nurses for Professional Purposes in Catalonia, Spain: Web-Based Survey Study”. The correct title is “Use of Health Apps by Nurses for Professional Purposes: Web-Based Survey Study”. The PDF has now been revised to remove “in Catalonia, Spain”. This correction notice will appear on the JMIR mHealth and uHealth website on April 30, 2020. Because the XML had been correctly submitted to all other repositories, it will not need to be resubmitted.

Submitted 02.03.20; this is a non–peer-reviewed article;accepted 06.03.20; published 30.04.20.

Please cite as:
Mayer MA, Rodríguez Blanco O, Torrejon A
Correction: Use of Health Apps by Nurses for Professional Purposes: Web-Based Survey Study
JMIR Mhealth Uhealth 2020;8(4):e18516
URL: https://mhealth.jmir.org/2020/4/e18516
doi:10.2196/18516
PMID:32352925

Original Paper

Bridging the Digital Divide Among Racial and Ethnic Minority Men Who Have Sex With Men to Reduce Substance Use and HIV Risk: Mixed Methods Feasibility Study

Elaine Hsiang[1], BA; Claudine Offer[2], MPH; Maximo Prescott[2], MPH; Amy Rodriguez[2], BS; Emily Behar[2], MS, PhD; Tim Matheson[2], PhD; Diane Santa Maria[3], MSN, RN, DrPH; Glenn-Milo Santos[2,4], MPH, PhD

[1]School of Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Substance Use Research Unit, Center for Public Health Research, San Francisco Department of Public Health, San Francisco, CA, United States

[3]Cizik School of Nursing, University of Texas Health Science Center at Houston, Houston, TX, United States

[4]Department of Community Health Systems, School of Nursing, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Elaine Hsiang, BA
School of Medicine
University of California San Francisco
Medical Student Center, S-245
513 Parnassus Avenue
San Francisco, CA, 94122
United States
Phone: 1 626 632 1904
Email: elaine.hsiang@ucsf.edu

Abstract

Background: Ecological momentary assessment (EMA) is a promising data collection tool for mobile health interventions targeting episodic health behaviors. For substance-using men who have sex with men (SUMSM), EMA is becoming more widely utilized in efforts to characterize substance use and sexual risk factors for HIV transmission. However, recent literature demonstrates emerging concerns over compliance and lower EMA engagement and data concordance among racial and ethnic minority SUMSM.

Objective: This study aimed to provide a qualitative evaluation of the barriers and facilitators of EMA as a data collection tool among racial and ethnic minority SUMSM.

Methods: Between October and November 2017, 45 racial and ethnic minority SUMSM were recruited from a list of prior research participants at the San Francisco Department of Public Health to participate in daily EMA surveys on their substance use and sexual health behaviors for 1 week, followed by in-person focus groups (FGs). A total of 4 FGs explored the participants' experiences with the surveys, issues regarding privacy and confidentiality, and suggestions for improvement. Qualitative analysis was performed using content analysis. Descriptive statistics and Fisher exact tests were used to assess the associations between demographics or substance use behaviors and EMA completion.

Results: Overall, 93.9% (295/314) of all delivered surveys were initiated, and of those, 98.0% (289/295) were completed. Neither participant demographics, including race (P=.65) or age (P=.43), nor substance use behaviors, including the frequency of alcohol (P=.40) or methamphetamine (P=.91) use or any cocaine (P=.28), crack (P=.99), or polysubstance use (P=.24), were found to be associated with survey completion. Overall, participants were receptive to the text message–based EMA surveys. Facilitators included survey timing, user-friendly survey design, survey-stimulated self-reflection, coding of sensitive phrases, and other privacy benefits of a mobile survey. Barriers included an inability to correct texting errors and participants' perception of judgment or stigmatization related to questions about condomless sex. To improve EMA compliance and uptake, participants suggested adding response confirmations, clarifying survey language, and continuing to diversify the study audience.

Conclusions: EMA appears to be feasible and acceptable among this sample of racial and ethnic minority SUMSM. Close attention to EMA study design and the development of nonjudgmental, contextualized questions regarding stigmatized health behaviors may be critical to further improve EMA compliance.

(JMIR Mhealth Uhealth 2020;8(4):e15282) doi:10.2196/15282

KEYWORDS

ecological momentary assessment; men who have sex with men; text messaging; substance use; HIV; digital divide; focus group

Introduction

In the United States, men who have sex with men (MSM) experience higher rates of substance use compared with the general population [1-3]. A growing body of literature independently links binge drinking, methamphetamine, and injection drug use to sexual risk behaviors and HIV transmission among MSM [2,4]. Although the incidence of HIV in MSM with injection drug use decreased slightly between 2012 and 2016, MSM remained the single largest demographic, accounting for 70% of all new HIV infections in the country (70%) [2]. HIV continues to disproportionately impact the health of substance-using men who have sex with men (SUMSM) [5-7] as well as MSM who identify as racial and/or ethnic minorities [8-11]. It is unclear whether substance use is associated with racial disparities among SUMSM living with HIV [12,13].

Understanding the prevalence, patterns, and frequency of substance use in SUMSM is necessary for the development of effective interventions to address substance use and HIV infection in this population. This research often relies on self-report of substance use behaviors and is vulnerable to social desirability bias, limitations in recall ability, and other mechanisms that introduce variations in validity [14,15]. In addition, lower data reliability among racial and ethnic minorities has been associated with the fear of legal repercussions because of the disproportionate criminalization of substance use, particularly among black/African American adults [12,16-19].

Ecological momentary assessment (EMA) has emerged as a promising tool for substance use and mobile health intervention research. EMA utilizes mobile technologies such as text messaging to collect real-time data and can reduce recall bias when characterizing substance use and other episodic health behaviors [20]. The tremendous potential of EMA as a robust data collection method is associated with the widespread ownership and use of mobile devices [21,22], and EMA has been leveraged in a number of studies describing substance use patterns, sex events, and the delivery of health interventions [23-27]. Despite its many benefits, EMA remains underutilized in substance use research focused on sexual and gender minorities, including SUMSM. In addition, concerns about response compliance feature prominently in the EMA literature, with a recent meta-analysis of EMA studies related to substance use reporting a pooled compliance rate lower than the recommended 80% [28]. Beyond evidence that compliance may differ between those with and without a clinical diagnosis of a substance use disorder [28], few studies have explored EMA engagement among SUMSM [29-31] and potential sociodemographic correlates. Previous studies by our research group found a significantly lower adjusted odds of responding to EMA text messages among racial and ethnic minority participants [32] as well as lower concordance in methamphetamine and alcohol reporting via EMA compared with data provided on timeline follow-back assessments [33]. These findings provide limited data to suggest a digital divide between racial and ethnic minority SUMSM and white SUMSM with regard to data reporting in EMA.

As EMA becomes more widely employed in substance use literature among SUMSM, it is important to explore the differences in EMA engagement and data concordance between white and racial and ethnic minority SUMSM and develop strategies to ensure that racial and ethnic minority SUMSM can fully benefit from interventions utilizing EMA data to reduce substance use and related HIV risk factors. Involving racial and ethnic minorities in feasibility and acceptability studies is a critical step in this process, yet few studies have done so [29,30]. Therefore, this study aimed to evaluate the acceptability of a text message–based survey leveraging EMA data among a sample of racial and ethnic minority SUMSM and elucidate barriers to and facilitators of EMA engagement and utilization.

Methods

Participant Recruitment and Data Collection

The Digital Divide study recruited 45 participants over phone in October 2017 using a list of previous SUMSM study participants at the San Francisco Department of Public Health who were willing to be contacted for additional studies. Eligibility was limited to participants living in the Bay Area who identified as men, were aged 18 years or older, belonged to a racial and ethnic minority, reported having sex with men, and were using at least one of four target substances (alcohol, methamphetamine, cocaine, and/or crack). Participants were required to be English speaking, have a phone that could receive and send text messages, and agree to participate in a focus group (FG). All participants provided informed consent. Study procedures were reviewed and approved by the Institutional Review Board at the University of California, San Francisco (CHR 17-22897).

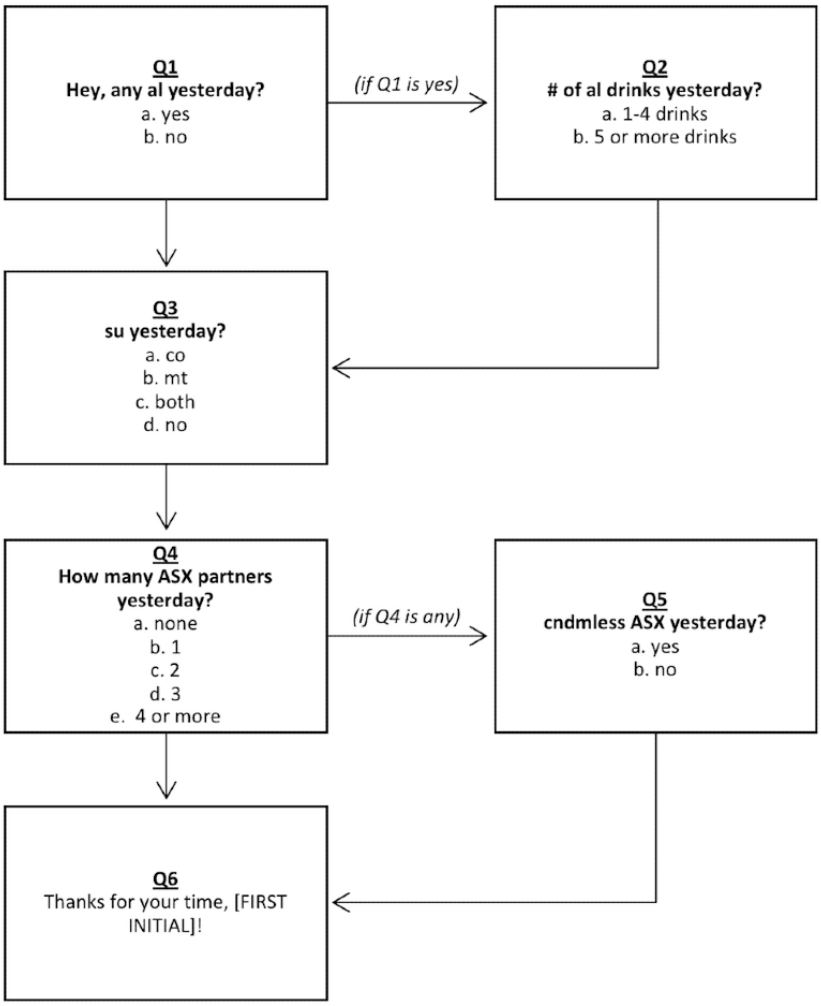
At baseline, a phone survey was conducted to collect the participants' demographic and substance use information. Participants were asked to quantify their frequency of alcohol use, binge drinking, cocaine use, crack use, and methamphetamine use in the previous 6 months. A 6-month recall period is a validated recall window used in prior epidemiological studies to gather self-reported substance use data in a feasible manner [34,35]. Once enrolled, participants received a confirmatory email with the date and time of their FG, an instructional guide to the text message surveys, and individualized technical support. Participants then received text message surveys on 7 consecutive days. Surveys were routed through the Health Insurance Portability and Accountability Act-compliant CareSpeak mobile health platform (OptimizeRx Corp). The surveys included 3 to 5 questions (depending on each individual's self-reported behaviors on a given day) and were estimated to take less than 5 min (Figure 1). Abbreviations (eg, *al* for alcohol, *su* for substance use, and *asx* for anal sex) were used in text message prompts to provide confidentiality, and a key was provided in the initial guide. Participants had a

choice between receiving the text messages at the default time of 10:30 AM each day or another time of their choice. All except one participant chose to receive messages at the default time.

After the weeklong EMA study, all 45 participants attended 1 of the 4 FGs in November 2017. There were 8 to 15 participants per FG, each lasting between 1.5 to 2 hours. A semistructured interview guide was designed to explore the participants' experiences with the surveys (Multimedia Appendix 1). Predefined topics included general barriers and facilitators to initiating or completing the daily text message surveys; suggestions for improving survey delivery, questionnaire design, and incentives; and issues with privacy or confidentiality. FGs were conducted by trained research staff who identified as MSM and formerly used methamphetamine.

Participants received US $2 for each text message survey that was completed and a bonus of US $6 for completing all 7 surveys. Participants were paid US $70 for attending the FG. Hence, participants received up to US $90 for completing all the study procedures.

Figure 1. Daily text message ecological momentary assessments for substance-using men who have sex with men. Abbreviations: al: alcohol; ASX: anal sex; cndmless: condomless; co: cocaine; mt: methamphetamine; Q: question; su: substance use.



Analytic Approach

Qualitative Data

The FGs were audiorecorded and transcribed verbatim. The transcripts were coded using Dedoose 8.0.35 (SocioCultural Research Consultants, LLC). Data saturation, in which additional data points do not contribute new information, was attained in the 4 FGs [36]. Content analysis was used to develop a codebook of emerging themes and perform subsequent analyses [37]. Two of the authors coded the transcripts and conferred frequently to discuss emerging themes and develop additional codes or resolve disagreements. Coded excerpts were then extracted, organized by category, and reviewed iteratively and corroborated by a third author before thematic compilation.

Quantitative Data: Demographics

Frequency of alcohol use, as well as binge drinking, was assessed using 4-point scales; however, the use of cocaine, crack, and methamphetamine was recorded as any or no use in the past 6 months because of the small sample sizes across the frequency categories. A variable for polysubstance use was generated to categorize the number of aforementioned substances the participants had used in the past 6 months (*only 1 substance* or *2 or more substances*).

Quantitative Data: Survey Completion

The primary outcome of interest was the completion of all text messaging surveys, dichotomized as successfully or unsuccessfully completing all 7 daily surveys. The completion of a survey required participants' responses to all applicable questions; surveys in which participants responded to at least one question but did not reach the final text thanking them for their time were classified as initiated but not completed. Owing to a technical issue, 1 participant received only 6 surveys in total; the authors considered this participant's completion of the 6 surveys as successful completion of all surveys. Descriptive analyses and Fisher exact tests were conducted to evaluate the associations between demographics or substance use behaviors and the completion of EMAs. Analyses were performed with Stata 14 (StataCorp), and statistical significance was determined using a *P* value of <.05. Given the scope of the study and the intent to elucidate baseline correlates of survey completion as a measure of feasibility and acceptability, the substance use and sexual behavior data obtained from the EMA surveys themselves were not analyzed as outcomes.

Results

Participant Characteristics

Demographics and substance use behaviors collected from the baseline phone surveys are presented in Table 1. Participants (n=45) had a mean age of 44 years (median 43 years; range 21-71 years). All participants reported being cisgender men (ie, people assigned male sex at birth and currently identifying as men). The majority of participants were African American (24/45; 53%) or Hispanic/Latino (16/45; 36%) and had sex exclusively with men in the past 6 months (35/45; 78%). Most (35/45; 78%) SUMSM in our sample reported using at least two substances in the past 6 months. The most commonly used substance was alcohol (40/45; 89%), followed by methamphetamine (31/45; 69%). The frequency of methamphetamine use was relatively high, with 22% (10/45) of all participants reporting methamphetamine use on 2 or 3 days per week and 27% (12/45) reporting use on 4 or more days per week in the past 6 months. Recent use of cocaine was more common (12/45; 27%) than crack (5/45; 11%).

Table 1. Participant demographics, sexual behaviors, and substance use behaviors collected at baseline (n=45).

Characteristics	Participants, n (%)
Age (years)	
18-24	1 (2)
25-34	9 (20)
35-44	14 (31)
45+	21 (47)
Race/ethnicity	
Asian/Pacific Islander	3 (7)
Black/African American	24 (53)
Hispanic/Latino	16 (36)
Mixed	2 (4)
Number of sexual partners in the past 6 months	
1	3 (7)
2-5	15 (33)
6 or more	27 (60)
Gender of sexual partners in the past 6 months[a]	
Only men	35 (78)
Multiple	10 (22)
Frequency of alcohol use in the past 6 months	
None	5 (11)
Once a week or less	8 (18)
2-3 days per week	23 (51)
4 or more days per week	9 (20)
Frequency of binge drinking in the past 6 months	
Never	13 (29)
Once a month or less	12 (27)
Weekly	17 (38)
Daily or almost daily	3 (7)
Cocaine use in the past 6 months	
Not used	33 (73)
Used at least once	12 (27)
Crack use in the past 6 months	
Not used	40 (89)
Used at least once	5 (11)
Methamphetamine use in the past 6 months	
Not used	14 (31)
Used at least once	31 (69)
Frequency of methamphetamine use in the past 6 months	
None	14 (31)
Once a week or less	9 (20)
2-3 days per week	10 (22)
4 or more days per week	12 (27)
Number of substances used in the past 6 months	

Characteristics	Participants, n (%)
1	10 (22)
2 or more	35 (78)

[a]Participants were not asked whether their sexual partners were cisgender or transgender.

Survey Compliance

Over the 7-day sampling frame, 314 text message surveys were successfully delivered to the participants. Of these 314 surveys, 295 (93.9%) were initiated. Overall, 9 of the 19 (47%) noninitiated surveys were attributable to 2 participants alone; one of the participants temporarily misplaced his or her phone for part of the study and another did not initiate any of the surveys. Among the 295 surveys that were initiated, 289 (98.0%) were completed. The baseline data of the participants completing all 7 text messaging surveys in their entirety are presented in Table 2. Neither the participants' race/ethnicity (P=.65) and age (P=.43) nor any of the substance use behaviors, including frequency of alcohol (P=.40) or methamphetamine use (P=.91) or any cocaine (P=.28), crack (P=.99), or polysubstance use (P=.24), were significantly associated with the completion rates.

Table 2. Demographics of participants (n=45) based on the completion of all 7 text messaging surveys.

Demographics	Completed[a] (n=32), n (%)	Did not complete (n=13), n (%)	*P* value
Race			**.65**
Asian/Pacific Islander	3 (100)	0 (0)	
Black/African American	17 (71)	7 (29)	
Hispanic/Latino	10 (63)	6 (37)	
Mixed	2 (100)	0 (0)	
Age (years)			**.43**
18-24	1 (100)	0 (0)	
25-34	5 (56)	4 (44)	
35-44	9 (64)	5 (36)	
45+	17 (81)	4 (19)	
Number of sexual partners in the past 6 months			**.99**
1	2 (67)	1 (33)	
2-5	11 (73)	4 (27)	
6+	19 (70)	8 (30)	
Gender of sexual partners in the past 6 months[b]			**.99**
Only men	25 (71)	10 (29)	
Multiple	7 (70)	3 (30)	
Frequency of alcohol use in the past 6 months			**.40**
None	5 (100)	0 (0)	
Once a week or less	6 (75)	2 (25)	
2-3 days per week	14 (61)	9 (39)	
4 or more days per week	7 (78)	2 (22)	
Frequency of binge drinking in the past 6 months			**.64**
Never	11 (85)	2 (15)	
At least once a month	8 (67)	4 (33)	
Weekly	11 (65)	6 (35)	
Daily or almost daily	2 (67)	1 (33)	
Use of cocaine in the past 6 months			**.28**
Not used	25 (76)	8 (24)	
Used at least once	7 (58)	5 (42)	
Use of crack in the past 6 months			**.99**
Not used	28 (70)	12 (30)	
Used at least once	4 (80)	1 (20)	
Use of methamphetamine in the past 6 months			**.99**
Not used	10 (71)	4 (29)	
Used at least once	22 (71)	9 (29)	
Frequency of methamphetamine use in the past 6 months			**.91**
None	10 (71)	4 (29)	
One a week or less	6 (67)	3 (33)	
2-3 days per week	8 (80)	2 (20)	
4 or more days per week	8 (67)	4 (33)	

Demographics	Completed[a] (n=32), n (%)	Did not complete (n=13), n (%)	*P* value
Number of substances used in the past 6 months			**.24**
Only 1 substance	9 (90)	1 (10)	
2 or more substances	23 (66)	12 (34)	

[a]Initiated but unfinished surveys were not considered *completed*.

[b]Participants were not asked whether their sexual partners were cisgender or transgender.

Qualitative Results

Overall, participants expressed ease with and receptivity of the text message–based EMA surveys. The 3 overarching themes of barriers to survey completion, facilitators of survey completion, and suggestions for study improvement covered a variety of aspects related to participants' experiences with the surveys. Selected quotes are illustrative of both the range of experiences as well as common perspectives.

Facilitators

Survey Scheduling and Delivery

FG discussions highlighted the consistent, timely scheduling of survey delivery as a major strength of text messaging as a survey modality. Over the course of the study, some participants reported overcoming an initial hesitance about the reliability of texted surveys:

> *I was like, maybe it's going to come through at 10:30 today, and then tomorrow it's going to come in at like 10:45 or something like that. And because there's always glitches in technology, but—for the seven days, when—every day when it came through at the same time, I was like, you know what? Someone nailed it on the head.* [FG1—Participant M]

The ability to customize the time of the survey delivery provided a sense of ownership and empowerment, but many agreed that 10:30 AM granted a sufficient timeframe to respond to the surveys. The participants also appreciated the ability to incorporate the surveys into their daily routine and ensure completion of as many surveys as possible:

> *I like the option of being in power...being able to choose the time I wanted it to actually come, that makes me feel like I'm more in control of it.* [FG4—Participant M]

Some participants explained that having a full day to respond to the surveys allowed them to wait and do so when they felt they could give accurate answers. For example, if participants received survey prompts when they were using substances, they could decide whether they had enough capacity to complete the surveys then or later:

> *I was usually "unavailable" (laughing). But that's the truth. I was tripping. During the time I was kind of high. But I was coherent [on the survey], because this was business.* [FG3—Participant F]

Survey Design

The flexibility of not having to complete the survey in one sitting was frequently characterized as a user-friendly feature. Participants commented on the value of being able to return to the surveys at a more convenient time, especially for those unable to respond during work hours. Many found this to be a benefit even when the surveys took little time to complete:

> *Because you have to be in that professional environment, and I'm getting these text messages while I'm trying to do work...a couple of times I did one or two questions and then at the very end, like at 5 o'clock in the afternoon, I was like, "oh shit, oh shit, better go and check that thing again," and there it was still.* [FG3—Participant D]

One participant described how participating in the surveys changed his relationship with his mobile phone. The ease of completing surveys with just a few keystrokes, combined with the responsiveness of the automated system, helped this participant become more comfortable with texting:

> *Here comes another question, and then you answer it; about 30 seconds later, it beeps, "Thank you for your time!" You know, it was great. I actually like texting more—I hated it before—and now I actually don't hate it. So this has actually converted a person who hates texting to someone who is open to it.* [FG1—Participant F]

Participants agreed that the survey was of an appropriate, and even optimal, length. Many participants could complete the survey within 5 min. One FG discussed the optimal number of questions. Although most said that they preferred the current number of 3 to 5 questions, others expressed a willingness to complete up to 10 questions at a time. Across FGs, it became evident that the speedy delivery of questions with defined, unambiguous answer choices was a stronger contributor to a positive experience than the actual number of survey questions:

> *It was very quick, and it seemed as though, like, the categories [had] a nice gap within it...So it gave you room if you was moderate, or social, or what have you.* [FG2—Participant P]

Personalization and Privacy

The personalization of texts using participants' initials (Figure 1) was often cited as a positive aspect of the surveys. Participants felt that they were being addressed personally and engaged in conversation, rather than simply filling out a form. Many appreciated the conversational language, stating that an interactive experience complemented the text messaging format and made them more likely to respond, as if they were texting with a live recipient:

> *At the very end, when they said, "Thank you for your time," I said, "No problem" (laughing).* [FG1—Participant M]

Several participants mentioned that they preferred being addressed by their initials, rather than their full name, because of increased privacy. Privacy was important to most participants when answering questions about sexual activity and substance use, but opinions differed with regard to other privacy safeguards in the survey, particularly the abbreviations used in the text messages. Although some felt that the abbreviations should have been spelled out to improve readability, participants ultimately agreed that having a code for terms such as *methamphetamine* or *condomless anal sex* protected their privacy:

> *[The abbreviations] were just for you, because you knew what they were...but somebody just picking up your phone and looking, they'd never know.* [FG3—Participants F and U]

> *I do [need codes] if you're living with a significant other. This could break up our relationship, marriage, such-and-such.* [FG1—Participant F]

Another participant shared his preference for coded messages because of past experiences with security breaches:

> *I need that option, too, because sometimes my phone has been tapped in the past.* [FG2—Participant G]

Participants spoke of completing surveys on their personal phones as a major facilitator of both privacy and convenience. Compared with in-person questionnaires, surveys that could be completed alone or in any chosen location minimized the stress and stigma of giving truthful answers:

> *I appreciated the ease...it wasn't really about anybody looking over my shoulder. So, I just appreciated that it was short, it was easy, and I could do it on the run, or waiting for a bus, or doing it at work.* [FG4—Participant E]

Facilitator of Personal Reflection

A number of participants reported that participating in the study provided a beneficial exercise in understanding their own substance use behaviors. Some individuals were surprised after quantifying their substance use on a day-to-day basis:

> *It also helped me self-reflect, because it was a busy week. And I was like, damn, I've been drinking a lot...Because I kept answering each day how many drinks I've had then, so it kind of kept me in check.* [FG3—Participant C]

For 1 participant, the reflective benefits of the surveys extended beyond individual substance use patterns. Recognizing that text messaging was a relatively new modality in collecting substance use data, this participant felt that he could help make a meaningful contribution to public health research:

> *It taught me [to] be more complete with things. It actually teaches you a few things, this survey. It opens up your mind. It's not like just, "Okay, let's go get this money." There's more to it. I made it more useful and utilized it in credence to some kind of meaning in my life...[helping] set a precedence for the rest of America, for the rest of any other public health service.* [FG1—Participant A]

Other participants commented on the deliberate representation of racial and ethnic minority MSM in this study and the impact of seeing other racial and ethnic minorities participate in research:

> *I like the galvanizing of people, gathering the vibes, and gathering the tribes of San Francisco. I like the connectivity...You can come to a focus group, and you're like, "Okay, where is everybody at that's supposed to be there?" It's good to see an actual turnout for some focus group to actually see more than one or two faces other than the people who are supposed to be there. It brings back some hope. Like, you're giving me back some hope, thinking that maybe there is a difference that you'd be making.* [FG1—Participant E]

Barriers

Inability to Correct Errors

The major barrier to accurate and complete documentation of participants' responses was the inability to change the previously submitted responses. Although this was not identified as a barrier to engagement, participants spoke widely of the impact of typographical errors (typos) on their survey responses. Participants discussed common experiences where either mistyping a letter or the autocorrect feature on their mobile phones resulted in a different answer choice than intended:

> *I would like constantly just do typos, and I was just wondering like, instead of "B," I put a "C." Was there any way to fix it?* [FG1—Participant J]

Another individual described a situation where he had incorrectly recalled his substance use from the day prior:

> *There were a couple of times that I put the wrong answer...And when I first wake up, and I get that survey, I remember answering the questions, but not remembering, "oh, I did have a glass of wine last night. Why did I say no?" And I couldn't change the answer once I submitted it.* [FG3—Participant B]

Participants also highlighted some confusion over response options. For survey questions structured to receive *yes* or *no* responses, some submitted *Y* or *N* instead of the provided response options *A* or *B*, respectively, and found themselves unable to verify their answers or correct the errors:

> *I kept putting "Y" or "N" for answering the questions because it was just what I was used to. The whole "A," "B," "C" part—sometimes I kind of got confused for a minute.* [FG3—Participant C]

In addition, unrecognized survey responses occasionally prevented the display of subsequent questions, thus truncating the surveys unless a CareSpeak representative manually texted the participants to input a valid response.

Sensitive Topics

The FGs drew mixed responses when asked about questions they would have preferred to skip. Not all participants felt comfortable answering all the survey questions. One participant described feeling judged for his responses on condomless anal sex:

> *It's an insecurity thing, you know. I'm condomless most of the time, so, it just kind of brings out self-doubt. It's one of those things—because I'm aware and it's kind of looked down upon.* [FG2—Participant A]

Multiple participants from this FG responded that because they regularly engaged in condomless sex or assumed sex would be condomless, the goals of this question were unclear and risked bringing up unpleasant or unwanted memories:

> *It's funny, because that's a stupid question to ask me. That's what I would see from it, honestly. Like, why you asking that?* [FG2—Participant R]

> *The questions forced you to remember things that you—probably wanted to forget.* [FG4—Participant B]

Other participants felt that the substance use questions were subtly structured to screen for substance use disorders, and were reluctant to respond without the ability to provide context:

> *I actually thought that the survey was trying to see how much you really drink. You know, to see, "oh well, she's an alcoholic," or "he's an alcoholic." That's what I thought that's what the survey was doing.* [FG4—Participant A]

In response to this discussion, 1 individual expressed his frustration that others would skip or fabricate answers:

> *I understand what the question is, but once again, you're an adult and you signed up for this, and you knew what you were signing up for...So there's no way you should be saying, I don't want to answer this question or that question because you explained it from the beginning, what the survey is all about. So how can you backtrack now? But we'll backtrack some of that money that you get and everything will be okay. Put it to them that way.* [FG3—Participant U]

Suggestions

Additional Survey Features to Improve Usability

Given the participants' earlier discussions on errors in response, many felt that the survey could be improved by adding a final review and confirmation of answers before submission:

> *Some people are really trying to be as honest as possible when they're trying to respond to these messages, and if they respond with the wrong answer and they want to go back and change it, I think that makes sense to summarize it at the end and see if all the answers are correct.* [FG2—Participant S]

Another suggestion included sending a *bump* to remind participants to complete the survey:

> *I don't want to disrespect your study, but your text was deprioritized to me. Yeah, whether you pay me or not, I was like, "Yeah, I want to do this because I committed to it." And when I commit to something, it's like a job and I want to, you know, do my best to complete the assignment. But, honestly, I have higher priorities...I get an average of about sixty emails and, probably twenty texts on average in a day...That's why having a reminder would be helpful.* [FG4—Participant J]

Participants commented on the large volume of texts, emails, and push notifications they received each day, debating whether a reminder text, email, or even call would increase the survey completion rates or add to their notification burden and contribute to survey noncompletion.

Clarifying Abbreviations

Participants brainstormed ways to facilitate the uptake of the abbreviations used in the study, especially after their peers raised concerns regarding privacy. As the abbreviation key was emailed to participants days to weeks before the start of the study, many reported having to search through their emails when attempting to complete the first survey:

> *The first text was just a question right away. So, there was no kind of priming. Even though I was primed a week ago, I totally forgot about it.* [FG4—Participant N]

Therefore, one suggestion was to include the survey instructions and abbreviation guide within the text message surveys themselves:

> *That way I can see, "Oh, okay, this is what I'm about to do. These are the codes." And then I can begin to answer the questions. But that would be super, super easy. I don't see how anyone could be confused that way, if it's embedded in the first question.* [FG3—Participant C]

Other suggestions included streamlining the abbreviation format by uniformly using uppercase or lowercase (eg, rather than the *asx* used for anal sex and *su* for substance use) to prevent confusion between abbreviations and acronyms as well as using alternative codes altogether. One FG discussed using emojis to represent different substances or sex behaviors, such as an unpeeled banana for condomless sex, sugar or salt for cocaine, and a cloud or blowing wind for methamphetamine.

Broadening the Study Audience and Focus Group Outreach

The FGs highlighted the need to continue recruiting diverse participants to increase the uptake of text messaging survey among MSM. Diversity extending beyond race and even sexual orientation or behaviors featured prominently in the dialog:

> *One thing I would love: it's good to come together, and get a gathering of the minds or consensus of what we all were just involved in, and then it also shows that it's not just one age group or a demographic that is participating; it's showing that it's not just us young folks—but it's also all ages that are getting involved in it.* [FG1—Participant M]

One participant proposed that EMA acceptability could be bolstered through explicit recruitment of MSM who identify as heterosexual or by expanding the study population to include people of all genders and sexualities who use substances:

> *If you were doing this survey with straight people, you might actually find out more. Like, on the down-low, how many people [are] having unprotected sex? And they may be HIV positive, and they're using the meth, and they got a wife at home, so there's a whole range of things that you all could cover as opposed to just asking predominantly gay and bisexual people.* [FG2—Participant G]

Discussion

Principal Findings

Few studies have explored the acceptability and feasibility of EMA among racial and ethnic minority SUMSM, and to our knowledge, this is one of the first studies to provide an in-depth, qualitative analysis that centers participants' input and recommendations. The study participants demonstrated an overall positive reception to the EMA surveys. A high daily survey completion rate (92%) in our study mirrors the trends in recent literature supporting EMA as a substance use data collection tool among MSM [28,29,32,33]. Although race and age have previously been found to correlate with EMA engagement among MSM [32], no demographic or substance use variables were significantly associated with EMA engagement in this study with racial and ethnic minority MSM, further suggesting the utility of EMA in diverse MSM populations.

A short survey of 3 to 5 questions, delivered at midmorning each day for 7 days, and allowing a full day for response was well received by participants. Previous research has demonstrated the potential of personalized text messages to increase response rates and influence substance use and sexual behaviors [26,38,39], as bidirectional communication allows for the delivery of on-demand resources or interventions. As suggested by participants, EMA recordings may also encourage self-reflection on substance use behaviors [40]. Additional studies should evaluate the impact of EMA on substance use as well as sexual risk behaviors [26] as our findings do not corroborate the latter.

In fact, the inclusion of sexual behavior questions solely on condomless sex was not well received by some participants. As many assumed that sex under the influence of substances, often termed *chemsex*, would be condomless, survey questions asking about condomless sex appeared redundant and even stigmatizing. Although recent literature documents an epidemiologically significant rise in chemsex among MSM [41,42], successful interventions on substance use and sexual health must consider the stigma associated with substance use, HIV, and identifying as MSM as well as racism and other forms of structural violence against racial and ethnic minority MSM [43]. Our findings provide evidence that participants' perception of judgment or stigmatization by the study design can be a barrier to participants' engagement and honest reporting of sensitive behaviors. Future research should consider outcomes beyond condomless sex and explore whether collecting data on other sexual behaviors can help destigmatize questions regarding condomless sex. The inability to correct erroneous responses was also an important barrier to accurate data collection; future EMA studies should develop data collection systems that provide participants with the ability to make corrections to their responses to address this limitation.

Previous studies have discussed concerns over the confidentiality of substance use data, particularly if mobile devices are lost, stolen, or accessed by parties such as law enforcement [24]. Discussions in our FGs revealed that the benefits of completing the surveys on a mobile phone at any location and time, combined with abbreviation codes used for sensitive information, may afford sufficient privacy for EMA engagement. A more comprehensive understanding of the ways to maximize participants' privacy is needed, including making data inaccessible on phones after submission.

An additional consideration of privacy and EMA engagement was reflected by suggestions to include MSM who do not identify as gay or bisexual but would otherwise benefit from these interventions. For racial and ethnic minority MSM who identify as heterosexual, sometimes labeled with the racialized term of being on the *down-low*, research studies that call for MSM participants and involve FGs or other public appearances may not be appealing due to concerns over privacy and confidentiality [44]. The use of mobile health interventions and creation of intentional safe spaces may help bridge the gap in understanding HIV transmission and other health disparities in these subgroups.

Limitations

The limitations of this study include issues with generalizability. Owing to the small sample size, our quantitative results demonstrated relatively low power to detect small differences in compliance by demographics or substance use patterns. In addition, the EMA intervention in this study was only for 7 days; it remains unclear whether participants would have shared similar perceptions on acceptability with a longer study. Although the participants were encouraged to share experiences and opinions that differed from their peers, FG discussions tend toward normativity [45]. Furthermore, a data sample drawn mainly from prior substance use intervention research participants at the San Francisco Department of Public Health may introduce selection bias. Older individuals with substance use disorders may be more willing to volunteer for intervention research studies, which may explain why our sample that was recruited from this pool included few MSM aged under 25 years. The inclusion of only those who own a mobile phone may also have excluded lower income communities who do not have reliable access to mobile devices. Future research comparing EMA studies that do or do not provide participants with mobile devices may be warranted. Finally, this study focused solely on text message–based EMA. With the rising popularity and ubiquity of smartphones, smartphone-based apps may offer opportunities to address several of the design challenges presented in this paper.

Conclusions

Our findings provide additional insight into the potential of EMA to collect substance use data from racially/ethnically diverse MSM. This study presents EMA as a feasible and acceptable approach that may help mitigate challenges in research conducted on stigmatized behaviors among racial and ethnic minority SUMSM. A user-centered and personalized survey design, the prioritization of privacy, and the impact of participants' self-reflection beyond the study were important facilitators of EMA completion among our participants. Future EMA studies among racial and ethnic minority SUMSM should endeavor to retain these study elements to achieve high acceptability and compliance. Important barriers identified in our sample, such as the lack of a mechanism to correct errors and the failure to contextualize questions about sensitive topics, should be addressed to improve the acceptability of EMA approaches in this marginalized population. Ultimately, efforts to refine EMA as a study tool will help ensure equitable benefit from emerging technologies and reduce digital divides across communities disproportionately impacted by HIV and substance use.

Acknowledgments

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under award number R25DA028567. The authors would like to thank the study participants for their time.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[PDF File (Adobe PDF File), 367 KB - mhealth_v8i4e15282_app1.pdf]

References

1. Centers for Disease Control and Prevention (CDC). 2016. Substance Use Among Gay and Bisexual Men URL: https://www.cdc.gov/msmhealth/substance-abuse.htm [accessed 2019-06-27] [WebCite Cache ID 78jr14XC3]
2. Centers for Disease Control and Infection (CDC). 2019. HIV Infection Risk, Prevention, and Testing Behaviors Among Men Who Have Sex With Men: National HIV Behavioral Surveillance, 23 US Cities URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-22.pdf [accessed 2019-06-27] [WebCite Cache ID 78jrCdRWm]
3. Finlayson TJ, Le B, Smith A, Bowles K, Cribbin M, Miles I, Centers for Disease ControlPrevention (CDC). HIV risk, prevention, and testing behaviors among men who have sex with men--National HIV Behavioral Surveillance System, 21 US cities, United States, 2008. MMWR Surveill Summ 2011 Oct 28;60(14):1-34 [FREE Full text] [Medline: 22031280]
4. Vosburgh HW, Mansergh G, Sullivan PS, Purcell DW. A review of the literature on event-level substance use and sexual risk behavior among men who have sex with men. AIDS Behav 2012 Aug;16(6):1394-1410. [doi: 10.1007/s10461-011-0131-8] [Medline: 22323004]
5. Carrico AW, Zepf R, Meanley S, Batchelder A, Stall R. Critical review: when the party is over: a systematic review of behavioral interventions for substance-using men who have sex with men. J Acquir Immune Defic Syndr 2016 Nov 1;73(3):299-306 [FREE Full text] [doi: 10.1097/QAI.0000000000001102] [Medline: 27258233]
6. Woolf-King SE, Rice TM, Truong HM, Woods WJ, Jerome RC, Carrico AW. Substance use and HIV risk behavior among men who have sex with men: the role of sexual compulsivity. J Urban Health 2013 Oct;90(5):948-952 [FREE Full text] [doi: 10.1007/s11524-013-9820-0] [Medline: 23974946]
7. Santos G, Rowe C, Hern J, Walker JE, Ali A, Ornelaz M, et al. Prevalence and correlates of hazardous alcohol consumption and binge drinking among men who have sex with men (MSM) in San Francisco. PLoS One 2018;13(8):e0202170 [FREE Full text] [doi: 10.1371/journal.pone.0202170] [Medline: 30118495]
8. Brooks R, Rotheram-Borus MJ, Bing EG, Ayala G, Henry CL. HIV and AIDS among men of color who have sex with men and men of color who have sex with men and women: an epidemiological profile. AIDS Educ Prev 2003 Feb;15(1 Suppl A):1-6. [doi: 10.1521/aeap.15.1.5.1.23607] [Medline: 12630595]
9. Maulsby C, Millett G, Lindsey K, Kelley R, Johnson K, Montoya D, et al. HIV among Black men who have sex with men (MSM) in the United States: a review of the literature. AIDS Behav 2014 Jan;18(1):10-25. [doi: 10.1007/s10461-013-0476-2] [Medline: 23620241]
10. Hall HI, Byers RH, Ling Q, Espinoza L. Racial/ethnic and age disparities in HIV prevalence and disease progression among men who have sex with men in the United States. Am J Public Health 2007 Jun;97(6):1060-1066. [doi: 10.2105/AJPH.2006.087551] [Medline: 17463370]
11. Centers for Disease Control and Prevention (CDC). 2019. HIV and African American Gay and Bisexual Men URL: https://www.cdc.gov/hiv/group/msm/bmsm.html [accessed 2019-06-27] [WebCite Cache ID 78jsJhy1E]

12. White D, Rosenberg ES, Cooper HL, del Rio C, Sanchez TH, Salazar LF, et al. Racial differences in the validity of self-reported drug use among men who have sex with men in Atlanta, GA. Drug Alcohol Depend 2014 May 1;138:146-153 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.02.025] [Medline: 24629628]
13. Irwin TW, Morgenstern J. Drug-use patterns among men who have sex with men presenting for alcohol treatment: differences in ethnic and sexual identity. J Urban Health 2005 Mar;82(1 Suppl 1):i127-i133 [FREE Full text] [doi: 10.1093/jurban/jti032] [Medline: 15738312]
14. Rendon A, Livingston M, Suzuki S, Hill W, Walters S. What's the agreement between self-reported and biochemical verification of drug use? A look at permanent supportive housing residents. Addict Behav 2017 Jul;70:90-96. [doi: 10.1016/j.addbeh.2017.02.011] [Medline: 28214742]
15. Napper LE, Fisher DG, Johnson ME, Wood MM. The reliability and validity of drug users' self reports of amphetamine use among primarily heroin and cocaine users. Addict Behav 2010 Apr;35(4):350-354 [FREE Full text] [doi: 10.1016/j.addbeh.2009.12.006] [Medline: 20053503]
16. Richardson J, Fendrich M, Johnson TP. Neighborhood effects on drug reporting. Addiction 2003 Dec;98(12):1705-1711. [doi: 10.1111/j.1360-0443.2003.00561.x] [Medline: 14651502]
17. Johnson TP, Bowman PJ. Cross-cultural sources of measurement error in substance use surveys. Subst Use Misuse 2003 Aug;38(10):1447-1490. [doi: 10.1081/ja-120023394] [Medline: 14509547]
18. Mitchell O, Caudy MS. Examining racial disparities in drug arrests. Justice Q 2015;32(2):288-313. [doi: 10.1080/07418825.2012.761721]
19. Li DH, Janulis P, Mustanski B. Predictors of correspondence between self-reported substance use and urinalysis screening among a racially diverse cohort of young men who have sex with men and transgender women. Addict Behav 2019 Jan;88:6-14 [FREE Full text] [doi: 10.1016/j.addbeh.2018.08.004] [Medline: 30099289]
20. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
21. 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]
22. Pew Research Center. 2019 Jun 12. Demographics of Mobile Device Ownership and Adoption in the United States URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-06-27] [WebCite Cache ID 78jtAinge]
23. 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]
24. Roth AM, Felsher M, Reed M, Goldshear JL, Truong Q, Garfein RS, et al. Potential benefits of using ecological momentary assessment to study high-risk polydrug use. Mhealth 2017;3:46 [FREE Full text] [doi: 10.21037/mhealth.2017.10.01] [Medline: 29184898]
25. Hopper JW, Su Z, Looby AR, Ryan ET, Penetar DM, Palmer CM, et al. Incidence and patterns of polydrug use and craving for ecstasy in regular ecstasy users: an ecological momentary assessment study. Drug Alcohol Depend 2006 Dec 1;85(3):221-235. [doi: 10.1016/j.drugalcdep.2006.04.012] [Medline: 16730923]
26. Reback CJ, Rünger D, Fletcher JB, Swendeman D. Ecological momentary assessments for self-monitoring and counseling to optimize methamphetamine treatment and sexual risk reduction outcomes among gay and bisexual men. J Subst Abuse Treat 2018 Sep;92:17-26 [FREE Full text] [doi: 10.1016/j.jsat.2018.06.005] [Medline: 30032940]
27. Wray TB, Kahler CW, Monti PM. Using ecological momentary assessment (EMA) to study sex events among very high-risk men who have sex with men (MSM). AIDS Behav 2016 Oct;20(10):2231-2242 [FREE Full text] [doi: 10.1007/s10461-015-1272-y] [Medline: 26746212]
28. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
29. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e67 [FREE Full text] [doi: 10.2196/mhealth.4344] [Medline: 26085078]
30. Smiley SL, Elmasry H, Webb Hooper M, Niaura RS, Hamilton AB, Milburn NG. Feasibility of ecological momentary assessment of daily sexting and substance use among young adult African American gay and bisexual men: A pilot study. JMIR Res Protoc 2017 Feb 2;6(2):e9 [FREE Full text] [doi: 10.2196/resprot.6520] [Medline: 28153816]
31. Duncan DT, Kapadia F, Kirchner TR, Goedel WC, Brady WJ, Halkitis PN. Acceptability of ecological momentary assessment among young men who have sex with men. J LGBT Youth 2017;14(4):436-444 [FREE Full text] [doi: 10.1080/19361653.2017.1365038] [Medline: 29988981]
32. Turner CM, Coffin P, Santos D, Huffaker S, Matheson T, Euren J, et al. Race/ethnicity, education, and age are associated with engagement in ecological momentary assessment text messaging among substance-using MSM in San Francisco. J Subst Abuse Treat 2017 Apr;75:43-48 [FREE Full text] [doi: 10.1016/j.jsat.2017.01.007] [Medline: 28237053]
33. Rowe C, Hern J, DeMartini A, Jennings D, Sommers M, Walker J, et al. Concordance of text message ecological momentary assessment and retrospective survey data among substance-using men who have sex with men: a secondary analysis of a

randomized controlled trial. JMIR Mhealth Uhealth 2016 May 26;4(2):e44 [FREE Full text] [doi: 10.2196/mhealth.5368] [Medline: 27230545]
34. Linas BS, Genz A, Westergaard RP, Chang LW, Bollinger RC, Latkin C, et al. Ecological momentary assessment of illicit drug use compared to biological and self-reported methods. JMIR Mhealth Uhealth 2016 Mar 15;4(1):e27 [FREE Full text] [doi: 10.2196/mhealth.4470] [Medline: 26980400]
35. Murphy DA, Hser Y, Huang D, Brecht M, Herbeck DM. Self-report of longitudinal substance use: a comparison of the UCLA natural history interview and the addiction severity index. J Drug Issues 2010 Mar;40(2):495-516 [FREE Full text] [doi: 10.1177/002204261004000210] [Medline: 21212856]
36. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9):1408-1416 [FREE Full text]
37. Crabtree BF, Miller WL. Doing Qualitative Research. Second Edition. Thousand Oaks, CA: Sage Publications; 1999.
38. Reback CJ, Fletcher JB, Shoptaw S, Mansergh G. Exposure to theory-driven text messages is associated with HIV risk reduction among methamphetamine-using men who have sex with men. AIDS Behav 2015 Jun;19(Suppl 2):130-141. [doi: 10.1007/s10461-014-0985-7] [Medline: 25563501]
39. Reback CJ, Grant DL, Fletcher JB, Branson CM, Shoptaw S, Bowers JR, et al. Text messaging reduces HIV risk behaviors among methamphetamine-using men who have sex with men. AIDS Behav 2012 Oct;16(7):1993-2002 [FREE Full text] [doi: 10.1007/s10461-012-0200-7] [Medline: 22610370]
40. Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
41. Hegazi A, Lee MJ, Whittaker W, Green S, Simms R, Cutts R, et al. Chemsex and the city: sexualised substance use in gay bisexual and other men who have sex with men attending sexual health clinics. Int J STD AIDS 2017 Mar;28(4):362-366. [doi: 10.1177/0956462416651229] [Medline: 27178067]
42. Maxwell S, Shahmanesh M, Gafos M. Chemsex behaviours among men who have sex with men: a systematic review of the literature. Int J Drug Policy 2019 Jan;63:74-89. [doi: 10.1016/j.drugpo.2018.11.014] [Medline: 30513473]
43. Pollard A, Nadarzynski T, Llewellyn C. Syndemics of stigma, minority-stress, maladaptive coping, risk environments and littoral spaces among men who have sex with men using chemsex. Cult Health Sex 2018 Apr;20(4):411-427. [doi: 10.1080/13691058.2017.1350751] [Medline: 28741417]
44. Carrillo H, Hoffman A. From MSM to heteroflexibilities: Non-exclusive straight male identities and their implications for HIV prevention and health promotion. Glob Public Health 2016;11(7-8):923-936. [doi: 10.1080/17441692.2015.1134272] [Medline: 26783732]
45. Smithson J. Using and analysing focus groups: limitations and possibilities. Int J Soc Res Methodol 2000;3(2):103-119. [doi: 10.1080/136455700405172]

Abbreviations

EMA: ecological momentary assessment
FG: focus group
MSM: men who have sex with men
SUMSM: substance-using men who have sex with men

Edited by G Eysenbach; submitted 27.06.19; peer-reviewed by J Andriesen, P Janulis, F Calvo; comments to author 21.10.19; revised version received 25.01.20; accepted 04.02.20; published 29.04.20.

Please cite as:
Hsiang E, Offer C, Prescott M, Rodriguez A, Behar E, Matheson T, Santa Maria D, Santos GM
Bridging the Digital Divide Among Racial and Ethnic Minority Men Who Have Sex With Men to Reduce Substance Use and HIV Risk: Mixed Methods Feasibility Study
JMIR Mhealth Uhealth 2020;8(4):e15282
URL: http://mhealth.jmir.org/2020/4/e15282/
doi:10.2196/15282
PMID:32347807

Original Paper

COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria

Iniobong Ekong[1*], MBBS, MPH, CGEIT; Emeka Chukwu[2*], BEng, MSc, CGEIT; Martha Chukwu[3*], BSc, MPA

[1]Department of Health Planning, Research and Statistics, FCT Health and Human Services Secretariat, Abuja, Nigeria

[2]Department of Computer Information System, Faculty of Information & Communication Technology, University of Malta, Msida, Malta

[3]Ragnar Nurkse Department of Innovation and Governance, School of Business and Governance, Tallinn University of Technology, Tallinn, Estonia

[*]all authors contributed equally

Corresponding Author:
Emeka Chukwu, BEng, MSc, CGEIT
Department of Computer Information System
Faculty of Information & Communication Technology
University of Malta
PG room A24, Level 0
Msida, MSD 2080
Malta
Phone: 356 99330888
Email: nnaemeka_ec@hotmail.com

Abstract

Background: The coronavirus disease (COVID-19) pandemic is the biggest global economic and health challenge of the century. Its effect and impact are still evolving, with deaths estimated to reach 40 million if unchecked. One effective and complementary strategy to slow the spread and reduce the impact is to trace the primary and secondary contacts of confirmed COVID-19 cases using contact tracing technology.

Objective: The objective of this paper is to survey strategies for digital contact tracing for the COVID-19 pandemic and to present how using mobile positioning data conforms with Nigeria's data privacy regulations.

Methods: We conducted an exploratory review of current measures for COVID-19 contact tracing implemented around the world. We then analyzed how countries are using mobile positioning data technology to reduce the spread of COVID-19. We made recommendations on how Nigeria can adopt this approach while adhering to the guidelines provided by the National Data Protection Regulation (NDPR).

Results: Despite the potential of digital contact tracing, it always conflicts with patient data privacy regulations. We found that Nigeria's response complies with the NDPR, and that it is possible to leverage call detail records to complement current strategies within the NDPR.

Conclusions: Our study shows that mobile position data contact tracing is important for epidemic control as long as it conforms to relevant data privacy regulations. Implementation guidelines will limit data misuse.

(JMIR Mhealth Uhealth 2020;8(4):e19139) doi:10.2196/19139

KEYWORDS

COVID-19; contact tracing; Nigeria's National Data Protection Regulation; General Data Protection Regulation; GDPR; coronavirus; surveillance; mHealth; eHealth; digital health

Introduction

The coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. This infectious respiratory disease was first detected in Wuhan City, China, in December 2019. It was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020, and has currently infected over two million people worldwide and has killed over 150,000 people. Globally, responses have been swift and in full influenza pandemic control

mode [2]. Travel and movement restrictions to curtail spread both within and across cities are in force. Many cities around the world are in lockdown or lock-in mode. Some have issued dusk-to-dawn curfews. In other scenarios, large gatherings have either been banned or discouraged. Estimates suggest that this pandemic can claim the lives of as many as 40 million people globally [3]. The Spanish flu, which lasted between 1918 and 1920 in some places, has been estimated to have cost the lives of 21-50 million people globally [4]. Evidence suggests that influenzas can mainly be spread through large clusters [5]. The WHO global influenza preparedness plan presents guidelines for the management and control of influenza and other disease [6]. Nigeria, one of the countries that adopts WHO guidelines, has over 493 cases of COVID-19 as of April 17, 2020, with 17 mortalities. This is a substantial increase since the index case was reported on February 27, 2020. To better manage the spread, Nigeria's federal government has declared a lockdown in key affected states (ie, Lagos, Ogun, and the Federal Capital Territory). The lockdown was in addition to several mitigating actions by state governments, ranging from a ban on social gatherings to dusk-to-dawn curfews. During the lockdown, schools, markets, churches, mosques, banks, offices, parks, motor parks, and airports remain closed, often for a 14-day period.

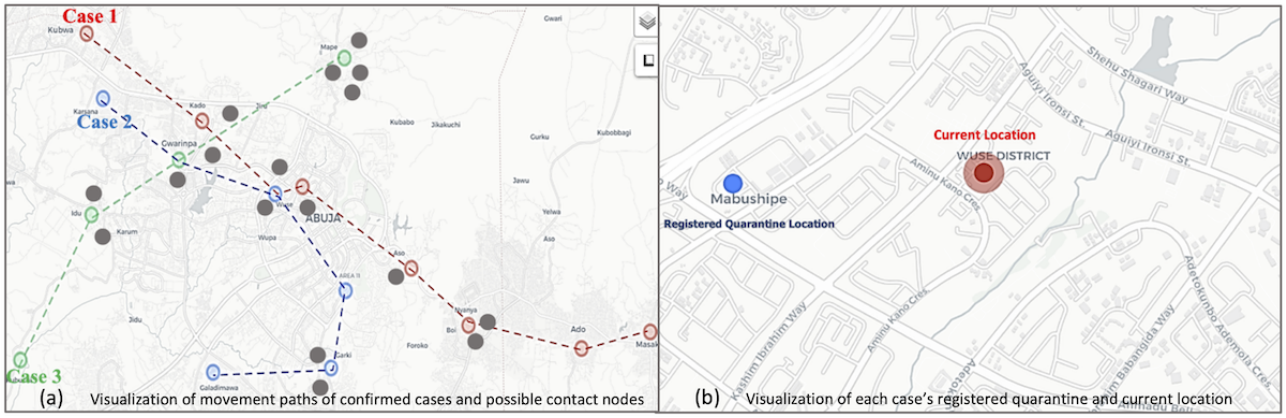
The Nigeria Centre for Disease Control (NCDC) reported that it is currently conducting contact tracing of over 9000 contacts of confirmed cases in an attempt to effectively contain the spread of the disease, in line with the recommended measures for pandemic response [7,8]. These measures include antiviral, vaccine, and nonpharmaceutical measures such as case isolation, household quarantine, school or workplace closure, and travel restrictions. Given the scale of the COVID-19 pandemic, nonpharmaceutical actions appear to be the only practical and logical option in the absence of any known antiviral drug or vaccine. Resources are stretched even in countries with advanced health care systems, as seen in Italy, the United Kingdom, China, and the United States [9,10].

Although the NCDC's approach has been commended for its compliance with WHO guidelines for large-scale containment and contact tracing, there remain options that may yet be explored [11]. Given the inadequacy of testing kits, it is believed that the number of confirmed cases may be far lower than the actual number of cases in Nigeria and most African countries. This is fueling speculations of a real, catastrophic-level pandemic if isolation, containment, quarantine, and contact-tracing mechanisms are not urgently implemented. In a country with an already weak health care system occasioned by poor health investment choices, managing such an outbreak will become impossible.

There is, therefore, a need to develop and adopt new strategies, particularly digitally-enabled strategies, to facilitate a more extensive, accurate, seamless, and timely response in line with the high frequency of new infections among contacts of confirmed cases (ie, the secondary infection rate) [12]. The adoption of digital solutions in Nigeria has been focused on electronic forms for contact data collection and visualization for follow-up [13]. Digital technologies can do more than be a tool for field data collection or serve as an outbreak investigation platform. Data on households and general population movement patterns can be extracted through digital technologies [14]. Farrahi et al [15] showed that over a 9-month period, 72 participants made 10,992 phone calls and 9432 SMS records representing communication flow; additionally, these participants made 1,973,547 Bluetooth interactions representing physical proximity movements. When extrapolated for three cases in Abuja, the capital city of Nigeria, their movement can result in thousands of interactions. In the case of an infection, these three cases can initiate an exponential number of contacts through these interactions, as seen in Figure 1A. The registered quarantine address can be visualized on the map and movement of quarantine subjects can be monitored with notifications enabled (Figure 1B).

Our paper reviews global practices in the use of mobile positioning data to achieve a more targeted and efficient approach at contact tracing and disease surveillance, especially in the context of the COVID-19 pandemic. We discuss how this approach is possible within regulatory confines. We also recommend a novel strategy for coordinating agencies to leverage mobile positioning data, and how to ensure patient privacy is preserved.

Figure 1. Visualization of Movement paths of cases and quarantine location.



Methods

The COVID-19 pandemic is emerging and only three months old with little scholarly work to justify a systematic search, review, and analysis approach. We conducted an exploratory (nonsystematic) internet search for technology approaches and responses to COVID-19. Results from global and national agencies responsible for infection prevention and control were analyzed to ascertain how they currently use technology. We also reviewed how these use cases fit within the regulatory framework for contact tracing and isolation. A similar internet search methodology was adopted for Nigeria's response and its use of digital tools for contact tracing.

Results

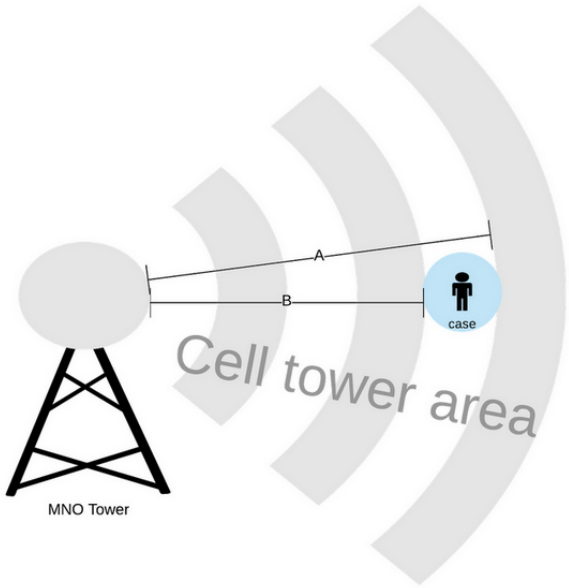
Our search yielded results based on emerging trends and the use of digital technologies by countries around the world to respond to COVID-19. We first present global perspectives and response strategies on the use of mobile position data during previous and current pandemics. We then present Nigeria's approach.

Mobile Position Data: How It Works

The GSM Association puts the total number of mobile subscribers at 5 billion unique subscribers and 7 billion connected devices [16]. Nigeria has 184 million active mobile subscriber lines [17]. Mobile telecommunications subscriber communication and movement data was used for contact tracing during the Ebola outbreak [14]. Many countries are currently using mobile data for a more rapid response to the COVID-19 pandemic [18,19]. There has been a 90% increase in the number of countries implementing digital tracking measures and a 100% increase in reports of censorship [20]. These approaches range from the use of anonymized aggregate data to monitor the general mobility of people and track the mobile phones of confirmed cases to tracking suspected patients and their contacts. In some cases, these approaches were individualized and mandatory while, in others, they were aggregated and anonymized. In all cases, there were collaborations between the government, mobile network operators (MNOs), and other data controllers such as technology companies and financial services providers.

At any time, each mobile subscriber is connected to a segment of the MNO base station tower. For simplicity, we have presented a cell tower and a subscriber in Figure 2. We used letters A and B to illustrate the farthest and shortest distance of the subscriber from the base station tower based on power throughput and internal cell tower position triangulation. The difference between A and B, representing the diameter of a user's device, which is a proxy for the user's location, often ranges between 50 and 300 meters, but depends other factors [14].

Figure 2. Location of a subject with respect to a mobile network operator (MNO) cell tower.



Global Strategies

Table 1 details some of the strategies governments around the world are adopting to track and isolate COVID-19 patients and their contacts or for lockdown/lock-in enforcement. In the United States, US $500 million of the US $2 trillion economic stimulus bill recently signed into law has been allocated to the US Centers for Disease Control and Prevention to launch a new surveillance and data collection system to monitor the spread of COVID-19 [21]. This move is a first for the United States since stringent patient data privacy and security regulations have hampered the adoption of contact tracing as a countermeasure for epidemic control in the past [22]. Similarly, the state of Massachusetts has announced the launch of what it calls the "first contact tracing" call center with 1000 virtual assistants to call and trace contacts of COVID-19-positive persons [23].

The European Union's General Data Protection Regulation (GDPR) is being tested on a large scale. Within the regulation, a patient can decide not to disclose who they have been in contact with or legally resist being traced [24]. Evidence has emerged that Germany, Austria, and Italy are using aggregated

call detail records (CDRs) to enforce lockdown and stay-at-home policies [25]. As this is an evolving challenge and European countries such as Italy and France are amongst the worst affected, changes to the GDPR regulations are expected and anticipated.

In China, the government worked with telecommunications companies to track and contact people who had traveled through Hubei province during the early days of the disease outbreak. Location data was shared with China's National Health Commission and other agencies, enabling them to retrospectively simulate the location of confirmed cases and their contacts, who were then issued warnings via social media [26]. Information has also emerged that the Chinese government may have leveraged its large network of sensors and surveillance cameras supported by an artificial intelligence–powered facial recognition and recommender system in its response to the COVID-19 outbreak [27]. This success may not be unconnected with the often criticized and loose patient data privacy and security regulation in China.

It was, however, observed that the extent of compliance with international and country-level regulations regarding data privacy considerations in deploying this digital technology varied from country to country.

Table 1. Strategies planned or adopted by countries for the use of mobile positioning data in response to the COVID-19 pandemic.

Country	Strategy planned or adopted
United States [21]	The state of Massachusetts announced the launch of its first contact tracing call center to be manned by 1000 virtual assistants [23]. The US federal government announced a US $500 million package for COVID-19 surveillance for the CDC [21].
China [22,28]	A mandatory smartphone app “Health code” that leverages a mesh network for infected persons contact tracing and notification.
Italy, Germany, and Austria [25]	Telecommunications providers allow for the sharing of location data with health authorities to check whether people are remaining at home. The data is aggregated and anonymous, mapping concentrations rather than individuals to respect Europe’s privacy laws.
South Korea [29]	The government created a map of cell phone data provided by telecommunications and credit card companies. The map was made public, so everyone could track their level of exposure.
Israel [19]	The government is using GMS call detail records in addition to patient mobile phone position data to locate contacts and trace their movement patterns.
Iran [30]	Iranian authorities developed a mobile app with government endorsement for COVID-19 self-diagnosis checks. It, however, also discretely collects user’s location data.
Singapore [18,21]	Singapore is using a mobile app that uses a Bluetooth-based mesh network to detect people's proximity to those who have been exposed to COVID-19 and warns them to get tested if they come in close contact.

The Nigerian Strategy

Human travel patterns and mobility can be assessed using available mobile phone data, and its application can be useful in disease epidemiology [31]. Panigutti et al [31] also revealed the adequacy of mobile phone data for tracking infectious disease spread, particularly in heavily populated and highly interconnected communities.

Border restrictions, internal travel restrictions, and school closures or total lockdown are reasonable but have minimal impact compared to effective case isolation or quarantine, which have been shown to have a significant impact if properly conducted [2]. This is particularly important in Nigeria’s case, where total compliance to these strategies cannot be guaranteed. Therefore, data on case isolation and quarantine should be a significant priority in Nigeria. Moreover, data is useful in modeling disease transmission. Specifically, collecting and analyzing data on transmission in different social contexts is highly effective in mapping intervention strategies since the impact of case isolation and quarantine depends on reducing contact between unaffected individuals and the index and other cases while they are ill [2].

In order for the NCDC to effectively conduct the current large-scale contact tracing of over 9000 contacts of confirmed cases, use of digital technology is inevitable. The number of contacts may even be more than this number considering the frequency of new infections. Currently, there are several digital contact data capture solutions, including the *Surveillance, Outbreak Response Management and Analysis System* (SORMAS). These solutions require a field epidemiologist or their representative to visit every contact.

Discussion

Principal Findings

Evidence suggests that contact tracing and data protection can go together [32]. Significant progress is being made with current strategies. As promising as they may seem, data privacy concerns remain a major impediment; it is necessary to find a balance between deploying the technology and maintaining data safety and patient privacy. Existing patient privacy regulations are currently being tested. Some countries have attempted to relax existing stringent regulations that protect patient privacy to allow for greater access; others have worked around them. According to Woods [21], many of the new digital technology approaches appear inevitable and legitimate, given the unprecedented high frequency of the COVID-19 infection spread. Many countries have now also invoked speedy legislative processes to give legitimacy to their workarounds and deployments.

In Israel, for instance, the cabinet has passed an emergency law to use mobile data for tracking people infected with COVID-19, trace their contacts, and identify individuals for quarantine [19]. This law was passed overnight, bypassing parliamentary approval. In the United States, privacy advocates are proposing stringent procedures to keep personal information safe, including deletion, once the data are no longer in use to prevent abuse by law enforcement agents [21].

The National Data Protection Regulation (NDPR) of Nigeria was promulgated in 2019 [33]. Amongst other stipulations, the regulation outlines the guiding principles for data processing in Section 5. These principles consider data processing unlawful if there is no consent by the individual data subjects (in this case, the confirmed persons), if it is inaccurate with prejudice to human dignity and not protected against cybercrime, and if it is stored beyond a reasonable period of time. However, despite these guiding principles, Section 6, part 2.0, subsection 2.2 (e) of the document lists the conditions for lawful data processing and states that:

> *Processing is necessary for the performance of a task carried out in the public interest or the exercise of official public mandate vested in the controller.*

The data controller in the case of mobile positioning data is the MNO, the entity that determines the purposes for and the manner in which network subscriber phone data is processed or is to be processed. Section 11 of the regulation states that data processing by a third party (eg, a public authority such as the Federal Ministry of Health, the NCDC, or anybody engaged in processing the location data such as a technology company) shall be governed by a written contract with the data controller. Interestingly, though the NDPR protects the privacy of personal mobile location data, it also provides exceptions for the use of such data that override public interest, such as the current COVID-19 outbreak.

Recommendations

Mobile phone location data can be effectively utilized in Nigeria for COVID-19 response. The government can leverage existing mobile technology resources and infrastructure available in-country by working with MNOs and technology firms to optimize the ongoing contact tracing and surveillance of over 9000 known contacts of confirmed cases. This collaboration should be guided by the NDPR in order to protect and safeguard individuals' data, prevent a breach of data privacy rights as well as inappropriate use and abuse by law enforcement agencies beyond the period of contact tracing and surveillance.

In practice, however, the first step should involve anonymized mobile subscriber data in line with good data governance policy. In the spirit of goodwill, informed consent of confirmed cases should be appropriately obtained once they are diagnosed, whenever possible. The use of the public interest exception should be a last resort. A simplified guideline for these processes for adhering to the NDPR should be written and made transparently available for data custodians, requesting bodies, data handlers, and the patient or contact.

A third-party agreement should also be formally signed between parties interfacing with patient data in any way. A typical use case sensitive to data privacy concerns is the use of information only regarding visits to public facilities, including public transportation systems, parks, churches, mosques, or malls, by COVID-19-positive individuals, as described by Ohmukai et al [34]. The use of CDRs has been proven to be effective in detecting outbreak clusters, followed by the use of other frontline data collection tools for mitigating impact and containment [14]. A key limitation of using CDRs from MNOs is that for basic 2G (second generation) phone users, the location will rely on mobile network phone mast location triangulation only. This approach alone has a proximity accuracy of 50-300 meters. This accuracy level is not sufficient to identify persons who have been in contact with a COVID-19 patient since the WHO contact definition prescribes two meters [7]. The use of telecommunication CDRs should complement other strategies for effective results.

The immediate action after a successful contact trace is communicating the expected course of action to citizens of an infected community cluster. A simple, user-friendly interface using Unstructured Supplementary Service Data will help improve information requests and management for low-income but literate users. Interactive voice response technology will be suitable and appropriate for awareness response for low-literate users in their local language.

Conclusions

Mobile positioning data can significantly improve the capacity and scope of timely outbreak response and will help governments as well as other responders in Nigeria. When implemented early [15], there are opportunities to leverage positioning data to break the chains of disease transmission in community clusters. It can improve the efficiency of currently used field data collection and outbreak investigation platforms when used in synergy.

While mobile positioning data can be used within the current regulation, guidelines for data handlers must include measures to curtail misuse and unauthorized access. Future research should design and implement models for mobile position contact tracing.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2020. Naming the coronavirus disease (COVID-19) and the virus that causes it URL: https://tinyurl.com/t82w9ka [accessed 2020-04-01]

2. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature 2006 Jul 27;442(7101):448-452. [doi: 10.1038/nature04795] [Medline: 16642006]
3. van Elsland SL, O'Hare R. Imperial College London. 2020 Mar 26. Coronavirus pandemic could have caused 40 million deaths if left unchecked URL: https://www.imperial.ac.uk/news/196496/coronavirus-pandemic-could-have-caused-40/ [accessed 2020-04-03]
4. Johnson NPAS, Mueller J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. Bull Hist Med 2002;76(1):105-115. [doi: 10.1353/bhm.2002.0022] [Medline: 11875246]
5. Ferguson NM, Fraser C, Donnelly CA, Ghani AC, Anderson RM. Public health. Public health risk from the avian H5N1 influenza epidemic. Science 2004 May 14;304(5673):968-969. [doi: 10.1126/science.1096898] [Medline: 15143265]
6. World Health Organization. 2005. WHO Global Influenza Preparedness Plan URL: https://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_5.pdf [accessed 2020-03-29]
7. World Health Organization. 2020 Apr 03. Coronavirus disease 2019 (COVID-19): Situation Report - 74 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200403-sitrep-74-covid-19-mp.pdf?sfvrsn=4e043d03_12 [accessed 2020-03-01]
8. World Health Organization Writing Group, Bell D, Nicoll A, Fukuda K, Horby P, Monto A, et al. Non-pharmaceutical interventions for pandemic influenza, national and community measures. Emerg Infect Dis 2006 Jan;12(1):88-94 [FREE Full text] [doi: 10.3201/eid1201.051371] [Medline: 16494723]
9. Cascella M, Rajnik M, Cuomo A, Dulebohn S, Di Napoli R. Features, Evaluation and Treatment Coronavirus (COVID-19). Treasure Island, FL: StatPearls Publishing; 2020.
10. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. N Engl J Med 2020 Mar 18. [doi: 10.1056/nejmoa2001282]
11. World Health Organization. 2020 Mar 19. Considerations for quarantine of individuals in the context of containment for coronavirus disease (COVID-19) URL: https://tinyurl.com/y8smo4pe [accessed 2020-03-25]
12. World Health Organization. 2020. Household transmission investigation protocol for coronavirus disease 2019 (COVID-19) URL: https://www.who.int/publications-detail/household-transmission-investigation-protocol-for-2019-novel-coronavirus-(2019-ncov)-infection [accessed 2020-04-01]
13. Dimagi.com. 2020. Digital Solution for COVID-19 Response URL: https://www.dimagi.com/covid-19/ [accessed 2020-04-01]
14. Shibasaki R. International Telecommunication Union. 2017. Call detail record (CDR) analysis: Sierra Leone URL: https://www.itu.int/en/ITU-D/Emergency-Telecommunications/Documents/2017/Reports/SL/D012A0000CA3301PDFE.pdf [accessed 2020-03-20]
15. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. PLoS One 2014 May 1;9(5):e95133 [FREE Full text] [doi: 10.1371/journal.pone.0095133] [Medline: 24787614]
16. GSMA Intelligence. 2017. Number of Mobile Subscribers Worldwide Hits 5 Billion URL: https://www.gsma.com/newsroom/press-release/number-mobile-subscribers-worldwide-hits-5-billion/ [accessed 2020-04-01]
17. Nigerian Communications Commission. 2017. URL: https://www.ncc.gov.ng/ [accessed 2017-02-17]
18. Choudhury SR. CNBC. 2020 Mar 25. Singapore says it will make its contact tracing tech freely available to developers URL: https://www.cnbc.com/2020/03/25/coronavirus-singapore-to-make-contact-tracing-tech-open-source.html [accessed 2020-03-04]
19. Lomas N. Tech Crunch. 2020 Mar 18. Israel passes emergency law to use mobile data for COVID-19 contact tracing URL: https://techcrunch.com/2020/03/18/israel-passes-emergency-law-to-use-mobile-data-for-covid-19-contact-tracing/ [accessed 2020-03-18]
20. Cozzens T. GPS World. 2020 Mar 26. 19 countries track mobile location to fight COVID-19 URL: https://tinyurl.com/y8kpuvcv [accessed 2020-03-29]
21. Woods A. New York Post. 2020 Mar 26. CDC to launch new surveillance system to track coronavirus spread URL: https://nypost.com/2020/03/26/cdc-to-launch-new-surveillance-system-to-track-coronavirus-spread/ [accessed 2020-04-01]
22. Rafalski EM. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Encyclopedia of Health Services Research 1996. [doi: 10.4135/9781412971942.n180]
23. Carraggi M. Patch. 2020 Apr 03. MA To Launch First Coronavirus Contact Tracing Program In US URL: https://patch.com/massachusetts/boston/ma-launch-first-coronavirus-contact-tracing-program-u-s [accessed 2020-04-03]
24. Information Commissioner's Office. Guide to the General Data Protection Regulation (GDPR). Information Commissioner's Office 2018 [FREE Full text] [doi: 10.1211/pj.2017.20203048]
25. Pollina E, Busvine D. Reuters. 2020 Mar 18. European mobile operators share data for coronavirus fight URL: https://tinyurl.com/yadpu9gm [accessed 2020-04-01]
26. Grind K, McMilan R, Mathews AW. The Wall Street Journal. 2020 Mar 17. To Track Virus, Governments Weigh Surveillance Tools That Push Privacy Limits, Geolocation and facial-recognition systems can locate vectors of infections, but they also gather highly personal data URL: https://tinyurl.com/v99rex4 [accessed 2020-04-02]
27. Yuan S. Aljazeera China. 2020 Mar 01. How China is using AI and big data to fight the coronavirus URL: https://www.aljazeera.com/news/2020/03/china-ai-big-data-combat-coronavirus-outbreak-200301063901951.html [accessed 2020-03-20]

28. Zhang O. Associated Press / Time. 2020 Apr 02. Inside China's Smartphone 'Health Code' System Ruling Post-Coronavirus Life URL: https://time.com/5814724/china-health-code-smartphones-coronavirus/ [accessed 2020-03-10]
29. Lyons K. The Verge. 2020 Mar 23. Governments around the world are increasingly using location data to manage the coronavirus URL: https://tinyurl.com/y75xuwyz [accessed 2020-03-29]
30. Gilbert D. Vice. 2020 Mar 14. Iran Launched an App That Claimed to Diagnose Coronavirus. Instead, It Collected Location Data on Millions of People URL: https://tinyurl.com/yddt2p38 [accessed 2020-03-27]
31. Panigutti C, Tizzoni M, Bajardi P, Smoreda Z, Colizza V. Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models. R Soc Open Sci 2017 May 17;4(5):160950 [FREE Full text] [doi: 10.1098/rsos.160950] [Medline: 28572990]
32. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
33. National Information Technology Development Agency. 2019. Nigeria General Data Protection Regulation URL: https://nitda.gov.ng/ [accessed 2020-04-02]
34. Ohmukai I, Yamamoto Y, Ito M, Okumura T. arXiv.org. 2020. Tracing patients' PLOD with mobile phones: Mitigation of epidemic risks through patients' locational open data URL: http://arxiv.org/abs/2003.06199 [accessed 2020-03-27]

Abbreviations

2G: second generation
CDC: US Centers for Disease Control and Prevention
CDR: call detail record
COVID-19: coronavirus disease
GDPR: General Data Protection Regulation
MNO: mobile network operator
NCDC: Nigeria Centre for Disease Control
NDPR: National Data Protection Regulation
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SORMAS: Surveillance, Outbreak Response Management and Analysis System
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.04.20; peer-reviewed by T Yasaka, GE Iyawa, D Pförringer; comments to author 16.04.20; revised version received 17.04.20; accepted 19.04.20; published 27.04.20.

Please cite as:
Ekong I, Chukwu E, Chukwu M
COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria
JMIR Mhealth Uhealth 2020;8(4):e19139
URL: http://mhealth.jmir.org/2020/4/e19139/
doi:10.2196/19139
PMID:32310817

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org