JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 7 (2019), Issue 9 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Reviews

Original Papers

Corrigenda and Addenda

Review

From Paper to Digital Applications of the Pain Drawing: Systematic Review of Methodological Milestones

Nour Shaballout[1*], MD; Till-Ansgar Neubert[1*]; Shellie Boudreau[2], MSc; Florian Beissner[1], Dr phil nat

[1]Somatosensory and Autonomic Therapy Research, Institute for Neuroradiology, Hannover Medical School, Hannover, Germany

[2]Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[*]these authors contributed equally

Corresponding Author:
Florian Beissner, Dr phil nat
Somatosensory and Autonomic Therapy Research
Institute for Neuroradiology
Hannover Medical School
Carl-Neuberg-Straße 1
Hannover, 30625
Germany
Phone: 49 5115350 ext 8413
Email: beissner.florian@mh-hannover.de

Abstract

Background: In a pain drawing (PD), the patient shades or marks painful areas on an illustration of the human body. This simple yet powerful tool captures essential aspects of the subjective pain experience, such as localization, intensity, and distribution of pain, and enables the extraction of meaningful information, such as pain area, widespreadness, and segmental pattern. Starting as a simple pen-on-paper tool, PDs are now sophisticated digital health applications paving the way for many new and exciting basic translational and clinical applications.

Objective: Grasping the full potential of digital PDs and laying the groundwork for future medical PD apps requires an understanding of the methodological developments that have shaped our current understanding of uses and design. This review presents methodological milestones in the development of both pen-on-paper and digital PDs, thereby offering insight into future possibilities created by the transition from paper to digital.

Methods: We conducted a systematic literature search covering *PD acquisition*, *conception of PDs*, *PD analysis*, and *PD visualization.*

Results: The literature search yielded 435 potentially relevant papers, from which 53 methodological milestones were identified. These milestones include, for example, the grid method to quantify pain area, the pain-frequency maps, and the use of artificial neural networks to facilitate diagnosis.

Conclusions: Digital technologies have had a significant influence on the evolution of PDs, whereas their versatility is leading to ever new applications in the field of medical apps and beyond. In this process, however, there is a clear need for better standardization and a re-evaluation of methodological and technical limitations that no longer apply today.

(JMIR Mhealth Uhealth 2019;7(9):e14569) doi:10.2196/14569

KEYWORDS

pain drawing; digital pain drawing; pain chart; pain map; pain body map; pain diagram; ehealth; medical app

Introduction

Pain is a multifaceted subjective experience that poses unique challenges for objective assessment [1]. To date, many qualitative and quantitative assessments of pain rely solely on self-reporting. Perhaps, the simplest method involves pointing to the painful area or the use of words to describe the location and, if known, the quality of the pain. However, pointing and use of words often lack clarity and are challenging to quantify. A more objective tool for capturing pain location and even quality, amongst other aspects of the subjective pain experience, is a pain drawing (PD). When using a traditional PD, the patient marks or shades the location of pain and related symptoms on an outline of the human body or parts thereof [2,3]. This form

of communication allows physicians to capture the intensity, localization, and distribution characteristics of a patient's pain experience and extract meaningful and quantifiable information, such as pain area, intensity, and widespreadness.

Starting in 1949 as a simple pen-on-paper tool [2], PDs evolved into electronic form by the 1990s [4-9]. This evolution is now paving the way for new and exciting, basic translational and clinical applications. However, to grasp the full potential of digital PDs, it is necessary to understand and learn from the historical evolution of PDs and the methodological developments that have shaped our current understanding of uses and design. Furthermore, a consideration of the available digital technologies to date necessitates a re-evaluation of methodological and technical applications of PDs.

Previous reviews addressed specific aspects of PDs, such as iconography [3], reproducibility and reliability [10], association with psychological factors [11], and suitability for psychological screening [12-14]. This comprehensive review serves to assimilate the innovations and methodological milestones over the last 70 years that have advanced and shaped clinical and scientific application. Knowing these milestones is essential for the design of future PD applications in the context of mobile health. A further aim of the literature review was to uncover and reveal potentially overlooked and forgotten milestones related to *PD acquisition*, *conception of PDs*, *PD analysis*, and *PD visualization.*

For this review, *PD acquisition milestones* are defined as changes in the way we collect PDs in the clinic or research setting. *Conceptual milestones* represent advancements in our understanding of what information PDs can capture and the value this information provides. *PD analysis milestones* are new methods or approaches developed for extracting clinically relevant information, whereas *PD visualization milestones* are innovative designs or techniques for conveying meaning and presenting the information captured by the PD.

Methods

Given that PDs are also known as pain charts [2], pain maps [15], pain body maps [7], or pain diagrams [16], these terms in their singular and plural forms formed the basis of the search. To account for new advancements, we further added the term *digital* to identify this new form of PDs specifically. A systematic literature search in PubMed [17] using these terms was performed as of September 16, 2018. From this time, it yielded 512 results, and an additional 24 publications were further identified from reference lists of the initial search results and additional Web searches on Google scholar [18].

In a first pass, abstracts and (if necessary) text bodies of all publications were screened by authors NS and TN to check for the following exclusion criteria: (1) PDs were not based on body templates or parts thereof, (2) PDs were not made by patients or test subjects but by physicians or investigators, (3) study results were not obtained in adults, (4) the publication was a review article, and (5) the publication was not in English. This resulted in the exclusion of 101 publications. All remaining 435 papers were considered for further review, and the abstract and text body were screened by authors NS and TN in a second pass to identify papers that first disseminated potential methodological milestones.

In a third and final pass, papers constituting potential methodological milestones were reviewed by authors NS, TN, and FB. All potential milestones were followed up by performing a literature search to confirm provenance and rule out previous description by other papers. The final list of all milestone papers was reviewed and accepted by all authors. A flowchart of these procedures is shown in Figure 1 and the milestones have been illustrated in Figures 2, 4, 6, and 7.

Figure 1. Flow-chart of the literature search. Of the 53 methodological milestones identified by our search, 19 described pain drawing (PD) acquisition milestones, 18 conceptual milestones, 31 for PD analysis milestones, and 4 PD visualization milestones.

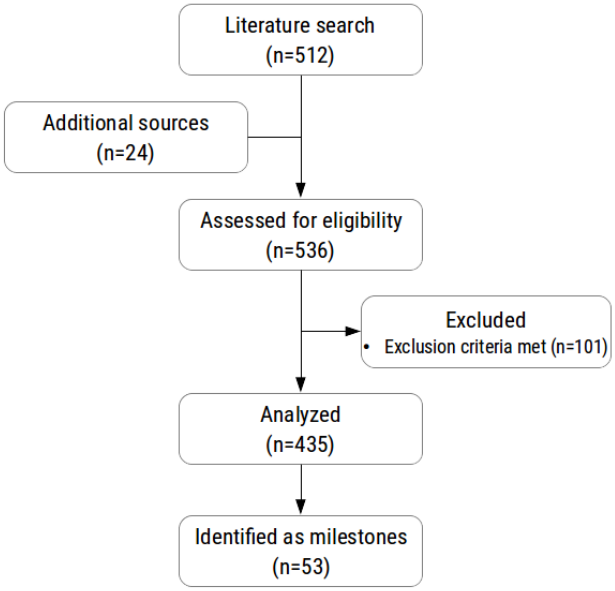
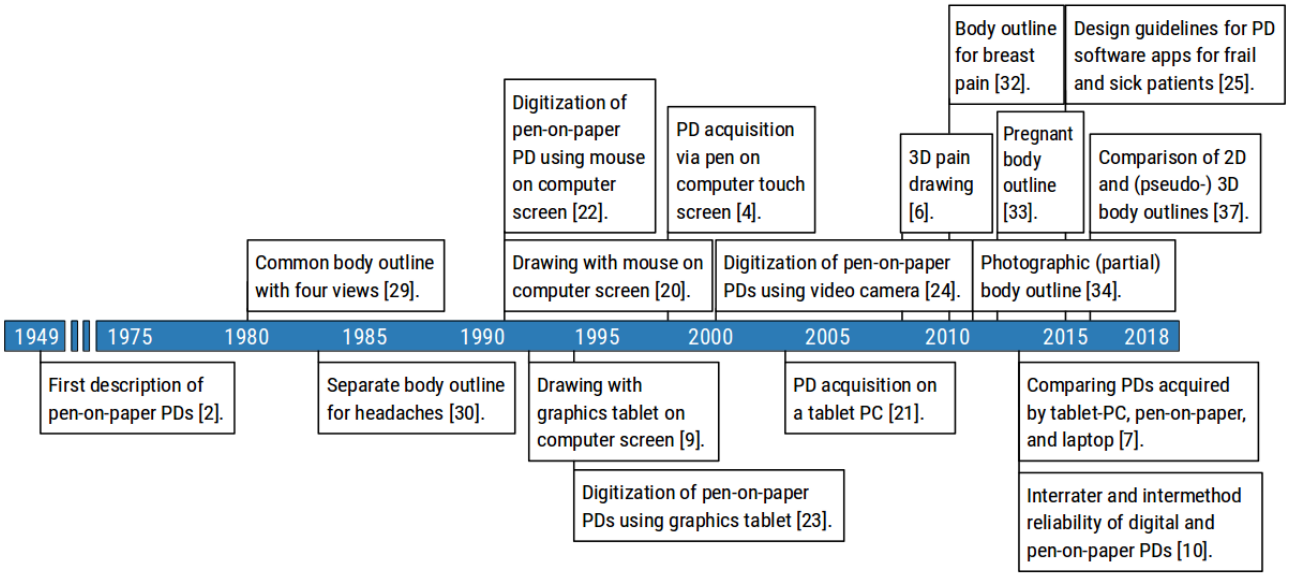


Figure 2. Methodological milestones in the area of PD acquisition. The acquisition methods for acquiring PDs over the last 2 decades appear to mirror the commercialization of digital technologies. PD: pain drawing; PC: personal computer; 3D: three-dimensional; 2D: two-dimensional.



Results

Brief Overview

We identified 53 milestone papers, of which 19 described PD acquisition milestones, 18 conceptual milestones, 31 PD analysis milestones, and 4 PD visualization milestones (see Multimedia Appendix 1).

The following sections discuss, together with other relevant scientific findings supporting or expanding, the significance of the identified milestones. Some of the milestones belong in more than one category and therefore are discussed where appropriate.

Pain Drawing Acquisition Milestones

PD acquisition milestones can be separated into 2 main topic clusters: PD data collection and digitization, and body templates.

Data Collection and Digitization

Albrecht Dürer's Renaissance drawing *The Sick Dürer* may be the first recorded account of a PD [19]. The first modern PDs, however, were pen-on-paper drawings where the patient used a pencil to "mark in on the charts wherever he experiences pain" [2]. Owing to its simplicity and ubiquity, the pen-on-paper acquisition method is probably the most common to date and will likely continue until supporting digital technologies become more widely adopted.

In 1991, Mann et al were first to acquire PDs directly on a computer [20] to assess the potential of using artificial neural networks (ANNs) for diagnosing low back pain disorders. However, pixel-based counts were first performed by North et al using a graphics tablet (Kaola) connected to an ordinary IBM personal computer (PC) to explore paresthesias evoked by implanted neurological stimulators [9]. For similar purposes, Aló et al acquired PDs with the aid of a pen-based interactive computer screen using a Windows-based software program (PainDoc, Quest-ANS Inc) [4]. A few years later, the same concept was replicated on a tablet PC (PenCentra 200, Fujitsu, Inc) [21] and tested as a clinical tool in a randomized control trial for automatic adjustments of a spinal cord stimulator.

The introduction of digitizing pads, electronic cameras, and image scanners facilitated computer-aided analyses of PDs by transforming pen-on-paper drawings into a digital format. Here, Mann et al were the first to digitize patients' pen-on-paper PDs by manually redrawing them on a computer screen using a mouse [22]. Bryner [23] in 1994 used a digitizing pad together with a custom-made software program written in Visual Basic (Microsoft) to create a digital PD (62,102 pixels) to quantify and compare the sensitivity for assessing pain extent between manual grid-based approaches and pixel-based counts. A unique approach to digitize the PD involved using a camera to capture an image of pen-on-paper PD to utilize ANNs [24]. Subsequent advancements from this point forward are difficult to assess as details on the exact digitization process are often insufficiently described in the published methods.

There are several studies focusing on assessing the clinical utility of PDs, and now, with the introduction of digital PDs, the usability and reliability of these platforms needed validation. Southerst et al were the first to assess and show good-to-excellent interrater and intermethod reliability of digitally acquired information about pain distribution to that obtained using pen-on-paper PDs [10]. In the same year, Jaatun et al assessed user interactions with paper, computer, and PC tablet drawings and reported a preference for tablet PDs, as based on patient opinions [7]. Two years later, the same authors published the first set of guidelines for designing PD software interfaces for patients with physical or cognitive impairments [25]. In 2016, the first quantitative comparison between paper and a digital platform (tablet PC) to collect PDs showed a high level of consistency and agreement [9].

In summary, the transitions for acquiring PDs over the last two decades appear to mirror the commercialization of digital technologies such as graphics tablets, touch screens, and custom-made computer programs. However, the driving forces

for implementing these new technologies are to facilitate or automate methods for treating, diagnosing, or managing pain. Collectively, the assessment and results of these acquisition milestones suggest that clinicians and researchers may choose either medium for acquiring PDs and can expect to see more digital PD technology in the future.

Body Templates

The reliability and accuracy of PD data collection methods for acquiring pain area, extent, and distribution are also highly dependent on the body template (or manikin). The literature search revealed several different body templates, as shown in Figure 3 [26,27]. Indeed, the body template is the central component of every PD. The features portrayed in the body template may influence a patient's ability to identify with the body and impact the quality of the PD.

The works of Palmer used an outline as body template, see Figure 3, that had already been in use for at least 50 years in the medical literature [2]. The exact origin of the body template outline is unclear. However, the body template was used in the seminal work by Henry Head on referred pain in visceral disease [28] in which the author made the drawings himself. Almost a century later, Margoles (1980) wrote a letter to the editor-in-chief of PAIN, requesting the use of a standard body template for the purpose of harmonizing PDs for accurate comparisons [29]. The proposed body template outline consisted of an anterior, posterior, and lateral views as well as the soles of the feet. However, this request was not widely adopted as evident by numerous versions of body templates in the scientific literature and clinical settings.

A body outline with additional views of the right and left head and upper and lower jaws for better depiction of orofacial pain, such as headaches and toothaches, was introduced by Toomey et al in 1983 [30]. Udén and Landin recognized the demand for sex-specific body outlines [31] and introduced the first body outline for female patients. This was naturally followed by the first body outline for patients with breast pain [32] and the first pregnant body outline [33]. Most recent are body templates depicting a realistic actual self [34,35]. The first realistic depiction used for basic and clinical research purposes is a photograph body templates and three-dimensional (3D) and pseudo-3D body templates. The first 3D body template for PDs was introduced by Ghinea et al in 2008 [6]. More realistic 3D and pseudo-3D body representations were later developed by other groups [5,36,37]. Comparing two-dimensional (2D) male and female as well as matching 2D and (pseudo-)3D body templates Egsgaard et al found that a majority of patients preferred sex-specific body templates and recommended 2D and (pseudo-)3D body templates to be used according to patient's preferences [37].

Conceptual Milestones

Conceptual milestones are ideas or results that have advanced our fundamental understanding of what PDs represent and what we can achieve by implementing them. A historical timeline review of these milestones is outlined in Figure 4. We identified 3 clusters of conceptual milestones: elements, generalizations, and sex-specific aspects of the PD.

Figure 3. The body template is a crucial ingredient of every pain drawing (PD) and should be chosen carefully as it may influence how much a patient is able to identify with the depicted body and thus impact the quality of the PD. (A) Body outline used by Palmer and many other early publications (modified after [28]). (B) Female body template (modified after [26]). (C) Hannover Body Template, a free body template with dermatome data (under CC BY 4.0) [87]. (D) Body template for frail and very sick patients (under CC BY-NC-ND 4.0) [27]. (E) Partial body template for the depiction of headaches (under CC BY 3.0) [57]. (F) Pregnant body template (under CC BY 3.0) [33]. (G) Pseudo-three-dimensional body template [5].

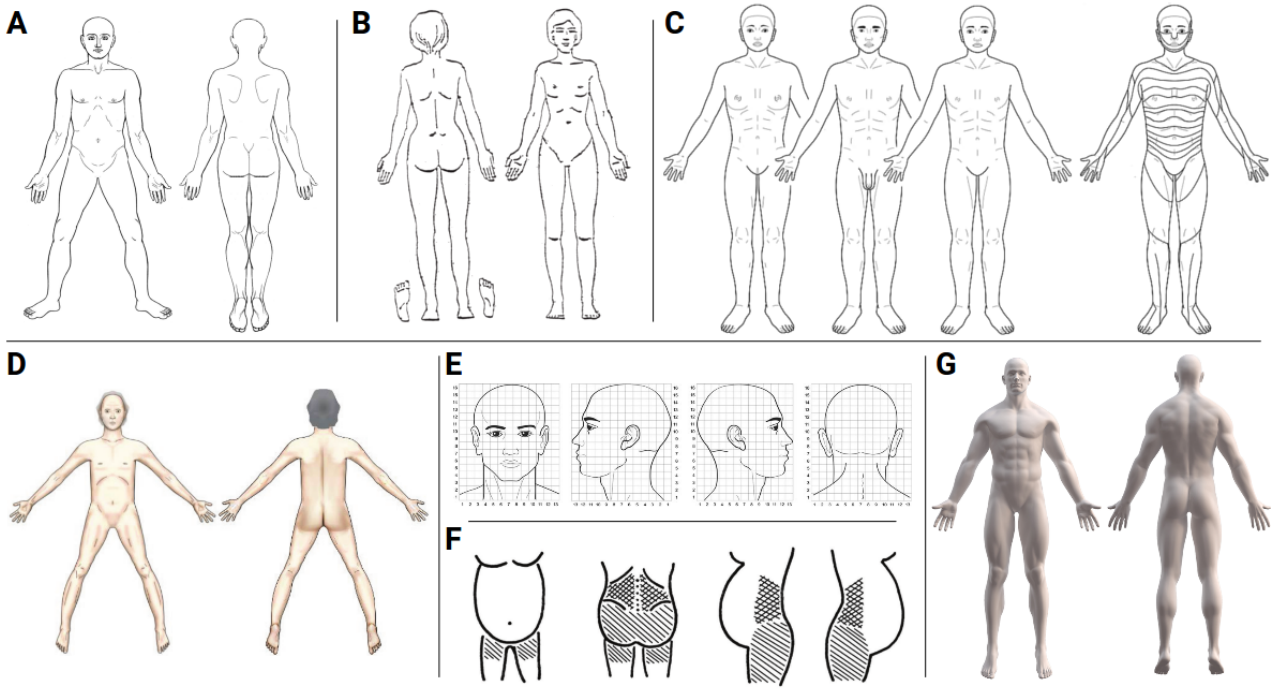
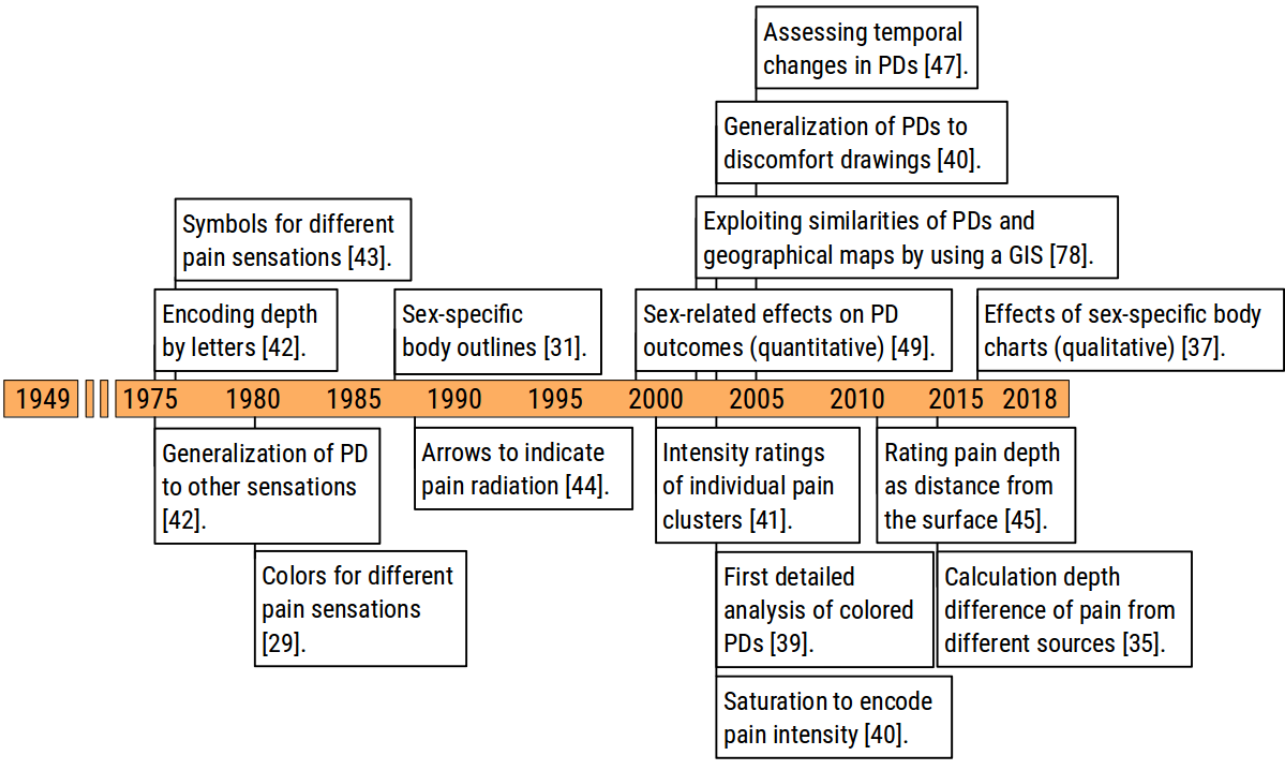


Figure 4. Methodological milestones in the area of conceptual pain drawing developments. PD: pain drawing; GIS: geographical information system.



Elements of the Pain Drawing

The largest of 3 conceptual milestone clusters addresses the typical building blocks or elements of PDs.

The idea of using color to represent different pain qualities or sensations, as illustrated in Figure 5, is well known but the exact origin less easily traced. Palmer wrote that after patients completed their PDs, he asked them about their type of pain and added this information to the PD himself [2]. The first clear documentation of using colors was presented by Margoles in the 1980s [29,38] and adopted by many in subsequent years. However, the first detailed analysis of the colored PDs emerged in 2003, when Masferrer et al reported "that colored pain drawings are no less useful than the black and white approach" [39]. Although pencils enable the encoding of pain intensity by shades of grey or color (eg, darker shades representing stronger pain), the applicability was not systematically exploited until Bertilson et al included the following instructions: "Shadow all pain/discomfort [...], shadow darker where there has been more discomfort" [40]. Before that, Türp et al introduced pain intensity ratings of individual clusters in a study exploring how generic pain intensity ratings are influenced by pain clusters in different parts of the body [41].

In line with the idea of color or shading encoding, the use of symbols for different types of pain, as illustrated in Figure 5, has been more widely adopted since its inception in 1975 [42,43]. Possible reasons may be that reproduction of color figures were inaccessible and costly, and at that time, medical publications rarely contained color. Symbol-based PDs, on the contrary, could be easily photocopied, interpreted, and presented in black-and-white or grayscale images.

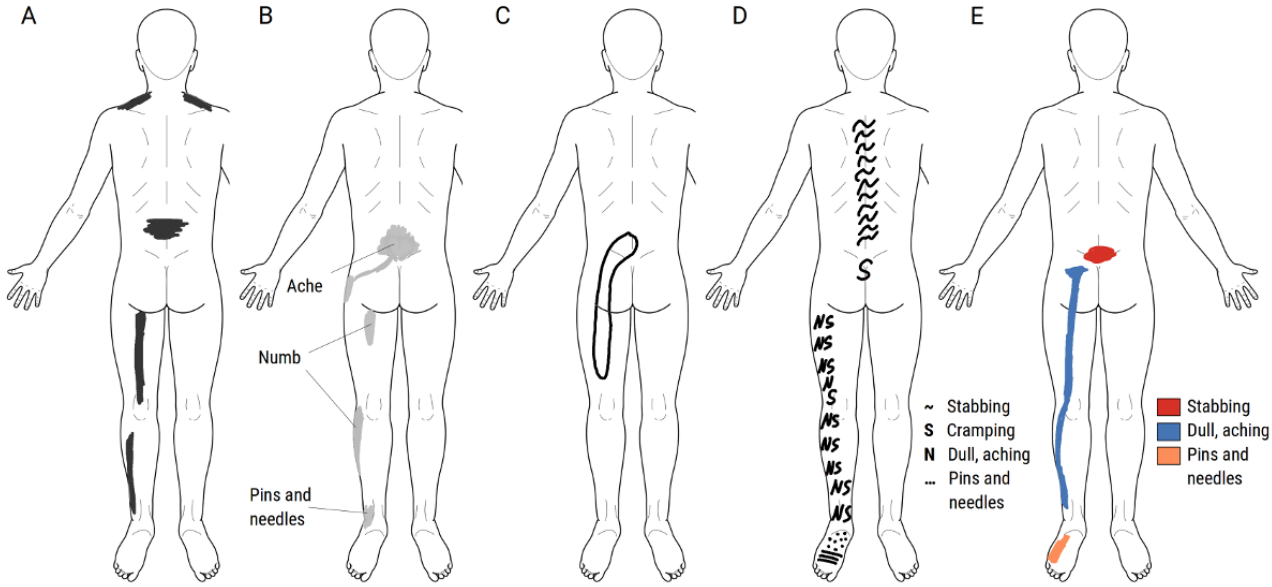
A new element added to the PD in 1988 was the systematic use of arrows to indicate radiating pain. Hildebrandt et al specifically instructed their patients to use arrows to document the area of the pain and the radiation extent [44].

Generalizations of the Pain Drawing

The second cluster of conceptual milestones are generalizations made to the *classic PD* over the decades. These include the addition of depth and time dimensions and inclusion of other symptoms and sensations.

One of the major shortcomings of early PDs was their inability to depict the depth of the patient's pain. Melzack in 1975 used the letter E for external and I for internal to encode depth [42], which is similar to that used by Margoles, who had his patients use a D to distinguish deep from superficial pain [29]. However, the *letter* method only allows for a rough estimation of the actual symptom depth. The first quantitative approach to rating pain depth in PDs was presented more than 20 years later by Jamison et al, who added a transverse section to their 3D manikin, where patients could put a mark to quantify depth as the distance from the surface [45]. Tucker et al expanded this approach in 2014 to calculate depth differences of pain elicited by stimulation of different muscle [35].

Figure 5. Schematic illustration of common methods for encoding pain location and sensation type in a pain drawing (PD). (A) Marking painful regions. (B) Tracing the outline of painful regions. (C) Marking painful regions and adding annotations. (D) Filling painful regions with predefined symbols. (E) Marking painful regions in predefined colors. All drawings were recreated using the app SymptomMapper [80], developed at Somatosensory and Autonomic Therapy Research, Hannover Medical School, and all methods are currently in use in both digital and pen-on-paper PDs.



In addition to depth, the concept of PDs was also expanded in the third and fourth dimension: The first 3D PDs were acquired by Ghinea et al in 2008 [6]. With their software, patients could manipulate the position of the manikin in all directions before marking their painful region on its body surface. Furthermore, 6 years later, the same group used virtual reality to visualize 3D PD to their patients [46]. Tracking the temporal dynamic of back pain patterns was first accomplished by Gibson and Frank [47]. In a feasibility study, they collected 2-hourly PDs and visual analogue ratings from users of electric wheelchairs finding that pain increased throughout the day in all users and was worst in the neck, back, and buttocks region.

Another PD generalization concerns the inclusion of other sensations, such as paresthesias by Melzack in 1975 [42]. More recently, Bertilson et al introduced the discomfort drawing, where patients mark all areas with pain and any other sort of discomforts such as nonpainful but unpleasant buzzing, tingling, or aching sensations. After completing the drawing, patients write the sort of discomfort next to the drawing using their own words [40].

Sex-Specific Aspects

Gender and sex-specific aspects in PDs were largely ignored until Udén and Landin introduced their gender-specific body outlines [31,48]. Before then, all body outlines were either male or relatively androgynous with prominent male appearance. The first report of gender differences in PD outcomes showed that women with neck-shoulder pain tend to draw larger areas and that their PDs are more symmetric than that of men [49]. This finding was later confirmed in a study exploring sex differences in musculoskeletal pain [50]. In the following years, specific body outlines for breast cancer survivors [32] and pregnancy-related lumbopelvic pain [33] emerged.

Finally, in 2016 and almost 30 years after the Udén and Landin paper, Egsgaard et al published the first investigation on qualitative effects of gender-specific body charts reporting that patients believed sex-specific body charts facilitate the communication of pain [37].

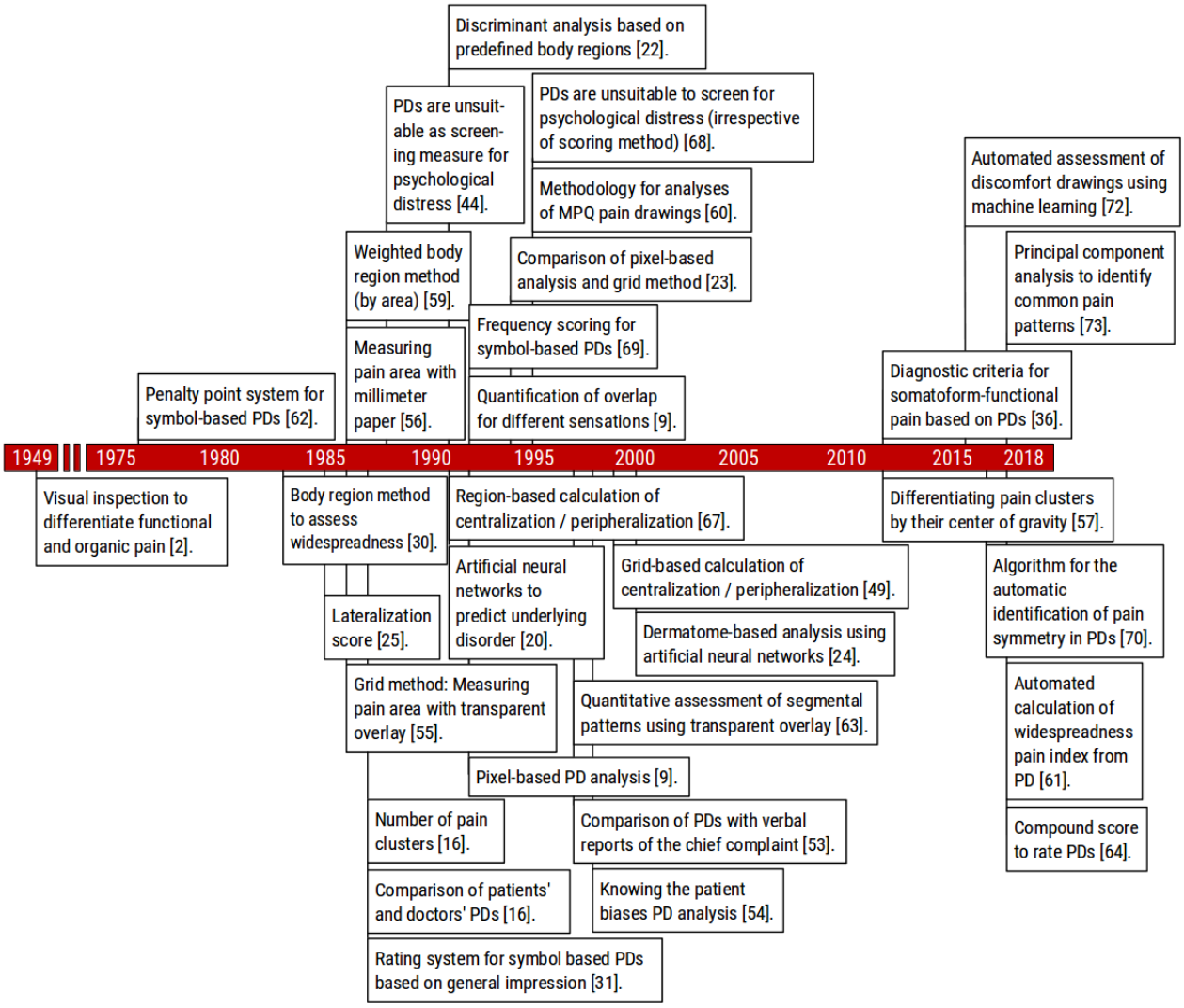
Pain Drawing Analysis Milestones

Like all expressions of pain, whether verbal or graphic, PDs need to convey meaningful and useful insight [3]. As this interpretation is not always straightforward, there is a need for interpretation aids that help clinicians draw the right conclusions from a particular PD. As illustrated in Figure 6, the aim to reduce subjectivity in PD analysis sparked a large number of PD-derived measures, rating systems, and diagnostic criteria.

Pain Drawing Reporting Style

Although the review is restricted to patient-made PDs, countless examples of PDs made by doctors appear throughout the medical literature [3]. Thus, the question of what differentiates the 2 types of PDs is an important one. In 1987, Cummings and Routan published the first study directly comparing patients' and doctors' PDs [15]. Interestingly, the authors presumed but did not confirm doctors' drawings to be more accurate as they were based upon a physical examination, and thus, patients' PDs should be compared with them. Similar studies were later published by other groups [51,52].

Figure 6. A timeline of methodological milestones contributing to advancements in PD analysis methods showing the golden age from 1985 to around 2000 as well as a renewal of interest in the 2010s. PD: pain drawing; MPQ: McGill pain questionnaire; IPQI: Integrated Pain Quantification Index.



Another analysis milestone was achieved by comparing patients' PDs with their verbal reports of the chief pain complaint. The results showed that patients' verbal descriptions of their chief complaint to a dentist frequently failed to capture and communicate pain located outside of the face region [53]. In this case, the additional pain—may be an important sign of temporomandibular disorders—was captured by using PDs, according to the authors, a prerequisite for initiating adequate treatment [53].

A central topic to the integrity of PDs is the bias of the observer. Here, Reigo et al recommend that the communication of information to be documented in a PD be performed in a blinded fashion. By comparing interobserver agreement in blinded and unblinded doctors, Reigo et al showed that clinical knowledge of the patient appears to introduce a strong bias [54].

Measures Derived from Pain Drawings

Measures derived from patients' PDs can be broadly divided into those that incorporate topographic measures (ie, anatomical knowledge) and those that do not, henceforth called simple measures.

Simple Measures

The most common and relatively simple measures obtained from PDs are pain area and extent. Pain area can be defined as the total area marked in a PD, whereas pain extent refers to how many different regions of the body are affected by the pain.

Pain area, as based on pixel counts, is easy to quantify in digitized PDs. However, a similar measure for pen-on-paper drawings requires additional interpretation tools, such as a grid system. Quantitative approaches for the assessment of pain area began in 1986. Gatchel et al applied a transparent overlay onto pen-on-paper drawings that displayed a grid system [55], whereas Fordyce et al used a transparent overlay of millimeter paper to quantify the pain area [56]. These 2 approaches added the area of pain as a new quantitative measure to the (short) list of independent pain assessment outcomes. One year later, Cummings and Routan introduced the number of *distinct pains*, that is, pain clusters as a new measure in their comparison of doctors' and patients' PDs [15]. North et al made the first pixel-based analysis of PDs in 1992 [9], where they quantified pain area by pixel count. The same study also introduced the quantification of overlap of different pain sensations, a method used by the authors to determine optimal stimulation settings

for their spinal cord stimulators. Two years later, Bryner [23] directly compared the grid system and pixel-based calculation methods, showing that the grid system overestimates pain area. The comparison made clear that the grid system introduces error, and the authors encouraged the adoption of pixel-based measurements. Finally, in 2012, Alonso-Blanco et al showed that the center of gravity is a means to localize pain clusters and, for example, differentiate between referred muscle pain, as demonstrated in myofascial pain and fibromyalgia syndrome [57].

Topographic Measures

The most commonly used topographic features of the body in PD analysis are predetermined body regions [58-61] and dermatomes [24,62-64].

In 1983, 3 years before Gatchel et al's and Fordyce et al's introduction of quantitative pain area assessment, Toomey et al published a body region method for assessing the total number of pain sites. This method consisted of counting the number of painful body regions using a predefined set of 32 scorable sites located over 7 body areas (head and neck, jaw, chest, abdomen, back, arms and hands, and legs and feet) [30]. It is important to note that the results obtained with this method are not so much a measure of pain area than of pain extent (or widespreadness). The method is very similar to that used for calculating the widespread pain index (WPI), an essential part of the 2010 and 2011 diagnostic criteria for fibromyalgia [65,66]. The WPI was a development independent of PD rating methods per se; however, Shaballout et al created an electronic PD automatically calculating WPI by masking the digital PD with a template of the 19 body regions and counting the nonempty regions [61].

A highly cited modification of the body region method came from Margolis et al in 1986 [59]. The authors published a new body outline with a different set of 45 scorable sites reflecting boundaries of anatomical landmarks. By assigning weights to each of the 45 scorable sites equal to the percentage of body surface, Margolis et al further developed the method to reflect a mixture of pain extent and widespreadness. By comparing their approach to a penalty point method introduced by Ransford et al (see below), the authors found that this method accounted for 56% of the variance in penalty point ratings. These findings suggest that the amount of body surface in pain may hold predictive power for screening patients with psychological distress or dysfunction.

In 1995, Escalante et al applied the body region method to analyze the PD in the McGill pain questionnaire (MPQ) [60]. As no rating system had been available before, they provided a set of scorable body sites adapted to the McGill body outline that could be printed on transparencies and opened the door for epidemiologic PD analyses of the MPQ.

Assessing the dermatomal distribution of pain has been a central part of PD analysis from the very beginning. However, whether or not a given pain pattern looked anatomically meaningful [62] is a debatable question, quantitative methods for pain pattern analysis were not used until 1997. Türp et al applied transparency-based dermatomal maps to distinguish local, regional, and widespread pain based on the exact dermatomal distribution [63]. More recently, Wallace et al introduced dermatome distance measure, defined as the number of segments between the most cranial and the most caudal painful dermatomes, to assess the severity of chronic pain disorders. Together with pain area, intensity, and persistence, the dermatome distance is part of the proposed Integrated Pain Quantification Index (IPQI), a one-dimensional pain score for representing the complex, multidimensional pain experience [64].

Other topographic measures derived from PDs are lateralization and peripheralization indices. Here, Margolis et al [58] were the first to calculate pain lateralization in chronic pain patients. On the basis of a body region approach, left-sided scores are subtracted from right-sided scores, so that positive and negative results indicate a right or left-sided lateralization, respectively [58]. Centralization or peripheralization, that is, a change in the distal-most extent of referred pain toward the lumbar midline or further away from it can be calculated similarly. Donelson et al differentially weighted body regions of the lower body by their distance from the lumbar region; thus, pain in more peripheral areas led to a higher score [67]. Finally, Toomingas used a grid-based approach to calculate the central-peripheral distribution of pain as the mean distance from the central line in a study characterizing neck, shoulder, and upper back pain among the general working population [49].

Rating Systems and Diagnostic Criteria

In the study by Palmer in 1949, interpretation of PDs was based on visual inspection, thus, relying solely on a doctors' experience [2]. Later, researchers applying the visual inspection method have emphasized certain aspects such as dermatomal patterns and symmetry as these criteria were originally proposed by Palmer as a method to differentiate between functional and organic pain. The first semiquantitative rating system to help differentiate between functional and organic pain came from Ransford et al in 1976 [62]. The group aimed to distinguish *organic* low back pain from what today would be called somatization disorder. The rating system assigned penalty points for elements of a PD, such as poor anatomic localization, drawings showing *expansion* or *magnification* of pain (eg, markings outside the outline), explanatory notes, circles or arrows to indicate particularly painful areas, or a tendency toward total body pain. The rating system was widely applied as well as criticized by many groups who were unable to replicate the original results. For instance, Hildebrandt et al showed that PDs as a screening measure for psychological distress were unreliable [44], and by comparing different scoring methods, Parker et al concluded that none of the methods was able to identify distressed patients or differentiate between organic and nonorganic pain patterns [68]. Since the original publication of Ransford's penalty point method, different modifications have been developed. In 1987, Udén et al noted that many of their patients were circling painful areas, adding explanatory notes, and making markings outside the body outline despite showing otherwise organic pain that responded to treatment. They developed a less quantitative approach based on *general impression* [31]. In addition, 5 years later, Sivik et al published a modification of Ransford's method replacing

some of the more subjective elements by a frequency scoring approach based on the following numbers: different pain types, markings in total, markings outside the body outline, markings with poor anatomical localization, and own markings [69].

More recently, Egloff et al developed diagnostic criteria for somatoform or functional pain by applying strictly quantitative methods of picture analysis [36]. Similar to Udén et al, they found that circle marks and marks outside the body outline are not specific for somatoform pain. In general, PDs with a higher number of marks, typically with symmetric patterns and the presence of long marks (lines), were identified as having a somatoform-functional origin [36]. Most recently, an algorithm for objective classification of symmetric pain patterns for electronic PD was developed and tested in patients with a common knee pain condition known as patellofemoral pain (PFP) [70].

Data Mining and Machine Learning Approaches

New possibilities for PD analysis opened up with the introduction of electronic PDs and computer-based analysis programs. Both developments move the potential of PD closer to becoming a tool capable of identifying the underlying cause of specific pain patterns reliably. The first step in this direction was by applying artificial intelligence in the form of ANNs to analyze PDs from patients with low back pain [20]. In the same year, the authors reached another milestone by using discriminant analysis based on predefined body regions to classify PDs into 1 of 5 lumbar spine disorders, with an accuracy of 46.2% (chance level: 20%) [22]. Surprisingly, this was only slightly lower than human expert raters, who reached 51% correct classifications [71]. Several years later, Sanders et al proposed a low back pain triage (degree of urgency) software application. They showed that training an ANN with dermatomal patterns resulted in significantly better classification than when training was performed using simple grid-based PD data [24]. More recently, Zhang et al developed a decision support system using machine learning to automatically assign diagnostic labels to PDs (discomfort drawings) [72]. The latest milestone applied principal component analysis (PCA) and k-means clustering to a relatively large cohort of patients with PFP and revealed 3 mutually independent pain distribution patterns [73]. Although the PCA study focused on knee pain, the results further support the utility of using ANN for diagnostic applications, whereas the results stemming from other machine learning methods suggest that these methods in combination may help identify and clarify underlying drivers of many (painful) diseases and syndromes.

Pain Drawing Visualization Milestones

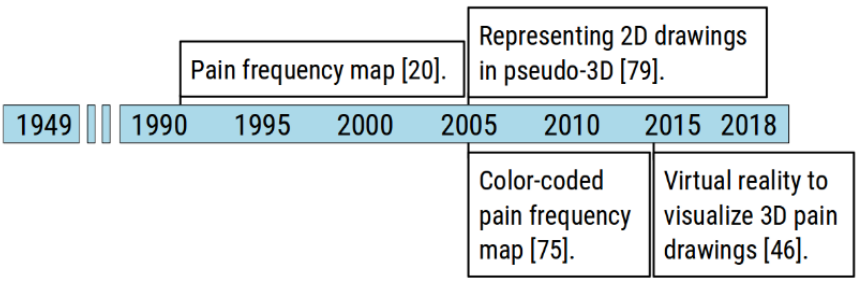
PDs are an instrument to visualize, document, and explore otherwise difficult descriptions of the pain experience. The last of the 4 main areas of methodological advancement is, therefore, a summary of milestones for visualizing data captured by using PDs and the results derived from them (Figure 7).

One of the most common ways of presenting statistical group results in PD studies is pain frequency maps. These maps show the distribution of pain for a select group (eg, patients), and the first was published in 1991 by Mann and Brown [20]. The pain frequency map used points for each pain mark from each PD and gave a rough visual impression of the most common pain locations and distribution in patients with spinal stenosis, herniated disc, and other underlying disorders. Other early and very different maps came from the groups of Türp et al, Slipman et al, and Svensson et al, who used bar charts [74], grayscale grids [75], and overlays of the tracings of each person's pain map [76] to represent pain frequency. Slipman et al later published the first color-coded pain-frequency map showing the dermatomal distribution of referred pain evoked by stimulating individual cervical discs [77]. In the PCA study by Boudreau et al, the pain frequency map represented the raw PD and was then filtered to differentiate more clearly the most common shape of each pain distribution pattern on and around the knee [73].

An interesting analogy exists between PDs and geographic maps as reported by Ghinea et al, who in 2002 suggested that geographic information systems are a suitable technical solution for storing and analyzing digitized PDs as well [78].

Finally, several groups developed ways to improve visualization of PDs, notably Hwang et al, who represented results of 2D drawings on a pseudo-3D body template [79] and Spyridonis et al, who used virtual reality to visualize 3D PDs to their patients [46].

Figure 7. Timeline of methodological milestones in the area of pain drawing visualization. 3D: three-dimensional; 2D: two-dimensional.



Discussion

Principal Findings

In this review, we have compiled a historical timeline detailing several methods for analyzing and visualizing data captured using PDs as well as conceptual steps for improving the applicability of PDs for basic and clinical research. A majority of the milestones revolve around PD analysis and interpretation. The systematic literature review revealed continuous developments along different lines of progression, namely *PD acquisition*, *conception of PDs*, *PD analysis*, and *PD*

visualization. In combination, these developments result in a more sophisticated PD since the original introduction. The future of PDs will depend on the utilization and adoption of the information into research and clinical settings. Advances in visualization of the information acquired by PDs may help facilitate this process, as this appears to be the most recent line of progression emerging in PD history.

A Toolbox for Clinic and Science

The conceptual developments of PDs mainly focused on improving the body templates to capture a deeper understanding of the pain experience and to better match the individual. These improvements revolve around core elements of the PD and include sex-specific body templates and encodings of intensity or the quality of pain. Altogether, these core elements can be viewed as a toolbox offering researchers and clinicians a number of options. Some of these core elements were already proposed by Palmer's groundbreaking publication [2], whereas others represent recent additions. A primary core element is the choice of a body template on which the drawing is to be made and of which a growing number of different versions exist (Figure 3). Encoding of sensations is a core element that manifests in many different versions, such as different symbols [43] or colors [29,39], expressing pain intensity by saturation [40], rating pain intensity for individual clusters [41], and indicating pain radiation by arrows [44]. Further supplemental PD elements are methods and measures to assess the depth of pain either by the *distance to body center* method [35,45] or by a simple binary rating using the letters E for external and I for internal [42]. An intermediate approach is to let patients choose among descriptions *on the skin*, *beneath the skin*, *muscle*, *organ*, and *bone* [80].

Overall, these core elements can be combined when tailoring a PD approach to a particular clinical or scientific need. To date, it is unclear which version of the core elements, body templates, and encoding of sensations is best. At this time, there is no dominant template or method, and this will contribute to a lack of standardization. Furthermore, PDs can be used to assess more than just pain, for example, discomfort drawings [40], general symptom drawings [80] or sensation drawings as evidenced by the recent application of PDs in studies on emotions [81-83], the placebo effect [84], or acupuncture [85,86]. This means the encoding of sensations may continue to develop in this area as the applicability of PDs expands.

How to Analyze a Pain Drawing

There are a significant number of PD analysis milestones. Many of them originating from pen-and-paper methods and then progressing into opportunities created by the introduction of PCs, tablets, and smartphones. The methods range from how to calculate pain area, extent, and widespreadness to encoding the PD as a grid system for ANN training. The analysis techniques are becoming sophisticated and require multidisciplinary teams extending from the clinic to mathematics and computer science. The PDs are similar to any other image that would be utilized in computer vision, such as those capable of identifying and discriminating between dog and cat. Methodological advances in PD analysis have resulted in several methods for quantifying the 3 main aspects of pain captured by the PD: pain intensity, localization, and distribution. Quantitative information on these aspects is well suited to complement other pain assessments such as questionnaires or analogue scales. Analytical methods include simple measures, such as pain area [55,56] or the number of clusters [15], topographic measures, such as segmental involvement [62,63] or widespreadness [61], and compound measures, such as the IPQI [64], that combine simple and topographic measures in a single one-dimensional score. More PD-derived measures will likely be developed, and their usefulness tested with the broader application of digital image analysis tools.

Need for Standardization

When reviewing and assembling the milestones, a common observation was a general lack of standards for using PDs. This concerns almost all critical methodological aspects. As a result, comparison of different study results is often complicated and sometimes impossible.

One of the main problems that has been pointed out as early as 1980 [29] is the lack of a standard body template for PD studies. Although many may argue that an arm is an arm irrespective of the exact body template used, we need to ask ourselves how much a PD body template biases the results by showing a muscular or skinny arm. Furthermore, many body templates differ in posture, and only some include lateral views. We acknowledge that there are special purposes such as assessing pain in pregnancy that necessitate the development of new body templates (see our discussion of milestones). For all other applications, however, it would be highly desirable that the community adopts a common body template or restricts itself to a minimal number of different templates. The question, however, is which one? To date, high-quality templates should be usable for both pen-and-paper and digital PDs [5,23], show all relevant body regions [29,30], be sex-specific [31], and ideally come with information on dermatomes [63] and other topographic regions [30,59]. As copyright is a common obstacle, the body template and dermatome schema provided by Neubert et al [61,87] has been made available under an open license, making the template free to use and accessible (asking permission is not required) [87]. We encourage publishers and the community to follow this example and release the body templates or start using templates that are openly available. Naturally, this also includes the emerging area of 3D or pseudo-3D body templates, where some realistic templates are available for research purposes [88,89] or even under an open license [90,91].

The second area in need of standardization is the instructions given to a patient before creating a PD. In our opinion, this aspect has not received enough attention in the literature, which has led to much confusion. One famous example is the drawing of symbols outside of the body template, the circling of painful areas, and the adding of explanatory notes by the patient, which some researchers consider signs of a somatization disorder [62,69], whereas others see them as perfectly reasonable ways to express pain [31]. In our experience, commonly used instructions such as "Describe your pain on the pain drawing" or "Mark your symptoms on this figure" lack the specificity necessary to achieve a consistent drawing style in patients.

Therefore, efforts should be made to determine the optimal set of instructions. One possibility may be to convey instructions in a graphical form [80,92]. Furthermore, when replicating a study from another group, researchers should use the same PD instructions used by that group. In the clinic, a joint discussion of the completed PD by the patient and the physician can help to avoid misunderstandings [2,93].

Limitations

Our approach to review the literature has several limitations. First, our selection of *PD acquisition*, *conception of PDs*, *PD analysis*, and *PD visualization as* categories to document advancements and individual milestones may inadvertently bias reporting. Second, the literature search primarily sourced papers assessing and evaluating PDs as primary instruments of measure and thus may miss publications outlining advancements as secondary or exploratory PD outcomes. Moreover, publications outlining acquisition and analysis methods of PDs were often insufficient to fully compare and differentiate different approaches. Thus, we may have also overlooked papers that have introduced a particular advancement but not explicitly mentioned it. Third, our literature search identified relevant publications by the search terms appearing in the title or abstract. Thus, any publications on PDs not mentioning these terms in the title or abstract have not been included. Furthermore, internet-based searches, in general, are prone to neglect publications from the preinternet era that are not digitized. We have tried to mitigate this limitation by checking the reference lists of all milestone papers before 1990, but we may still have overlooked important contributions.

How Digital Technology May Shape the Future of Pain Drawings

Digital technologies have had a significant influence on the evolution of PD methodology and will continue to do so. Although the first digital PDs in the early 90s [9,20] were still much more cumbersome than filling out pen-and-paper counterparts, digital PD acquisition has come a long way. Today, touch screens and digital pens are replacing keyboards and the computer mouse for data entry, whereas modern software apps work essentially like digital pen-and-paper platforms [5,7,87,94]. Furthermore, the latest available systems actively guide their users through the drawing process and thus improve PD quality. For example, guides help patients to conform to a particular drawing style (eg, by automatically filling closed shapes or restricting the drawable area to the body template). Some digital systems are even capable of calculating PD-derived measures in real time to aid diagnosis. These capabilities were initially developed to reduce research time and allow for more accurate assessments of the drawn pain area(s). However, the automatic calculations and visualization techniques provide a glimpse of new possibilities for integrating such information into busy workflows (clinic or research alike). Thus, it is an exciting prospect that such systems will become widely disseminated in the near future, and their clinical and scientific potential will be realized.

An area that has also significantly gained from digital technologies is PD analysis. Here, computer-based methods offer the possibility to analyze pain patterns in ever greater detail. Although digitally acquired PDs are advantageous for such analyses, they are by no means a prerequisite, as evidenced by numerous papers applying elaborate analyses to scanned or otherwise digitized pen-and-paper PDs [24,32,64,78,79]. Digital image processing also allows for digitization and analysis of extensive collections of pen-and-paper PDs (eg, as seen in the study by Wallace et al [64]), treasures which may be buried in medical archives).

However, the new technological possibilities also raise the question if specific popular methods, such as symbol-based PDs, grid-based methods [34,55], the counting of clusters [15] and scoring of body sites [30,59], are still in step with the times. They all reflect the limitations of the predigital age when photocopies were black and white, journal articles contained no color figures, and all PD scoring had to be done by hand. As a result, symbol-based PDs became popular as they could be reproduced in black and white and analyzed quickly by counting the different symbols, even though they were less intuitive for patients and had their spatial resolution limited by the size of the symbols, thus affecting all further analyses. In the age of digital PDs, however, these limitations no longer exist. Thus, instead of clinging to established yet outdated methods, we should embrace the new possibilities of the digital age. Patients can draw their pain patterns in full color at a resolution similar to that of pen-and-paper drawings and add information on intensity, depth, or any other relevant attribute in an iterative way. The final PDs can be archived and shared in full color either electronically or in printed form. Instead of reducing data size to a level that can be handled without a computer, measures derived from digital PDs can utilize all available data down to the pixel level and apply mathematical transformations of high complexity [64,73].

However, even the most informative measures, such as pain area, can never replace the PD itself. Although these simple measures are crucial for quantification and statistical analyses, there is an advantage by allowing our inherent ability to process and recognize patterns visually, thus individual examples of actual PDs should also be included in all future publications. Indeed, the ability to efficiently communicate pain is a primary advantage of PDs.

Conclusions

The PD as a clinical and research tool has undergone significant methodological development in the last 70 years and will continue to do so in the future. PDs capture many aspects of the subjective pain experience and have applications well beyond the pain specialty. Recent technological advancements, together with the versatility of the PD, have led to renewed interest in the past decade. Thanks to the transition from pen on paper to digital, we may soon see the dawn of a golden age of PDs.

Acknowledgments

The authors thank German Academic Exchange Service (DAAD) and the Horst Görtz Foundation for financial support. SAB is affiliated with the Center for Neuroplasticity and Pain, which is supported by the Danish National Research Foundation (DNRF121). These funding sources had no involvement in the design, data collection, data analysis, and manuscript preparation and/or publication decisions.

Conflicts of Interest

SAB has company holdings in Aglance Solutions ApS, which licenses a software app, Navigate Pain, used to collect electronic PDs. The remaining authors report no conflicts of interest.

Multimedia Appendix 1
List of methodological milestones.
[PDF File (Adobe PDF File), 127KB - mhealth_v7i9e14569_app1.pdf]

References

1. Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999 Apr;79(2):231-252. [doi: 10.1016/s0039-6109(05)70381-9] [Medline: 10352653]
2. Palmer H. Pain charts; a description of a technique whereby functional pain may be diagnosed from organic pain. N Z Med J 1949 Apr;48(264):187-213. [Medline: 18138692]
3. Schott GD. The cartography of pain: the evolving contribution of pain maps. Eur J Pain 2010 Sep;14(8):784-791. [doi: 10.1016/j.ejpain.2009.12.005] [Medline: 20079670]
4. Alo KM, Yland MJ, Kramer DL, Charnov JH, Redko V. Computer assisted and patient interactive programming of dual octrode spinal cord stimulation in the treatment of chronic pain. Neuromodulation 1998 Jan;1(1):30-45. [doi: 10.1111/j.1525-1403.1998.tb00028.x] [Medline: 22150884]
5. Boudreau SA, Badsberg S, Christensen SW, Egsgaard LL. Digital pain drawings: assessing touch-screen technology and 3D body schemas. Clin J Pain 2016 Feb;32(2):139-145. [doi: 10.1097/AJP.0000000000000230] [Medline: 25756558]
6. Ghinea G, Spyridonis F, Serif T, Frank AO. 3-D pain drawings-mobile data collection using a PDA. IEEE Trans Inf Technol Biomed 2008 Jan;12(1):27-33. [doi: 10.1109/TITB.2007.903266] [Medline: 18270034]
7. Jaatun EA, Haugen DF, Dahl Y, Kofod-Petersen A. Proceed with caution: transition from paper to computerized pain body maps. Procedia Comput Sci 2013;21:398-406. [doi: 10.1016/j.procs.2013.09.052]
8. Jamison RN, Fanciullo GJ, Baird JC. Computerized dynamic assessment of pain: comparison of chronic pain patients and healthy controls. Pain Med 2004 Jun;5(2):168-177. [doi: 10.1111/j.1526-4637.2004.04032.x] [Medline: 15209971]
9. North RB, Nigrin DJ, Fowler KR, Szymanski RE, Piantadosi S. Automated 'pain drawing' analysis by computer-controlled, patient-interactive neurological stimulation system. Pain 1992 Jul;50(1):51-57. [doi: 10.1016/0304-3959(92)90111-n] [Medline: 1381071]
10. Southerst D, Stupar M, Côté P, Mior S, Stern P. The reliability of measuring pain distribution and location using body pain diagrams in patients with acute whiplash-associated disorders. J Manipulative Physiol Ther 2013 Sep;36(7):395-402. [doi: 10.1016/j.jmpt.2013.05.023] [Medline: 23891480]
11. Reis F, Guimarães F, Nogueira LC, Meziat-Filho N, Sanchez TA, Wideman T. Association between pain drawing and psychological factors in musculoskeletal chronic pain: a systematic review. Physiother Theory Pract 2019 Jun;35(6):533-542. [doi: 10.1080/09593985.2018.1455122] [Medline: 29659315]
12. Bertozzi L, Rosso A, Romeo A, Villafañe JH, Guccione AA, Pillastrini P, et al. The accuracy of pain drawing in identifying psychological distress in low back pain-systematic review and meta-analysis of diagnostic studies. J Phys Ther Sci 2015 Oct;27(10):3319-3324 [FREE Full text] [doi: 10.1589/jpts.27.3319] [Medline: 26644701]
13. Carnes D, Ashby D, Underwood M. A systematic review of pain drawing literature: should pain drawings be used for psychologic screening? Clin J Pain 2006 Jun;22(5):449-457. [doi: 10.1097/01.ajp.0000208245.41122.ac] [Medline: 16772800]
14. Hartzell MM, Liegey-Dougall A, Kishino ND, Gatchel RJ. Utility of pain drawings rated for non-organic pain in chronic low back pain populations: a qualitative systematic review. J Appl Biobehav Res 2016 Sep 7;21(3):162-187. [doi: 10.1111/jabr.12048]
15. Cummings GS, Routan JL. Accuracy of the unassisted pain drawings by patients with chronic pain. J Orthop Sports Phys Ther 1987;8(8):391-396. [doi: 10.2519/jospt.1987.8.8.391] [Medline: 18797035]
16. Fortin JD, Aprill CN, Ponthieux B, Pier J. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique. Part II: clinical evaluation. Spine (Phila Pa 1976) 1994 Jul 1;19(13):1483-1489. [Medline: 7939979]
17. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/pubmed/ [accessed 2019-05-02] [WebCite Cache ID 7845pjlCc]
18. Google Scholar. URL: https://scholar.google.com [accessed 2019-05-02] [WebCite Cache ID 784623MiL]

19. Schott GD. The Sick Dürer--a Renaissance prototype pain map. Br Med J 2004 Dec 18;329(7480):1492 [FREE Full text] [doi: 10.1136/bmj.329.7480.1492] [Medline: 15604197]
20. Mann NH, Brown MD. Artificial intelligence in the diagnosis of low back pain. Orthop Clin North Am 1991 Apr;22(2):303-314. [Medline: 1826553]
21. North RB, Calkins S, Campbell DS, Sieracki JM, Piantadosi S, Daly MJ, et al. Automated, patient-interactive, spinal cord stimulator adjustment: a randomized controlled trial. Neurosurgery 2003 Mar;52(3):572-80; discussion 579. [doi: 10.1227/01.neu.0000047818.99414.fb] [Medline: 12590681]
22. Mann NH, Brown MD, Enger I. Statistical diagnosis of lumbar spine disorders using computerized patient pain drawings. Comput Biol Med 1991;21(6):383-397. [doi: 10.1016/0010-4825(91)90040-G] [Medline: 1838724]
23. Bryner P. Extent measurement in localised low-back pain: a comparison of four methods. Pain 1994 Nov;59(2):281-285. [doi: 10.1016/0304-3959(94)90081-7] [Medline: 7892026]
24. Sanders NW, Mann II NH. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. Comput Biol Med 2000 Sep;30(5):287-298. [doi: 10.1016/S0010-4825(00)00013-5] [Medline: 10913774]
25. Jaatun EA, Haugen DF, Dahl Y, Kofod-Petersen A. Designing a reliable pain drawing tool: avoiding interaction flaws by better tailoring to patients' impairments. Pers Ubiquit Comput 2015 Jun 20;19(3-4):635-648. [doi: 10.1007/s00779-015-0850-3]
26. Bertilson BC, Brosjö E, Billing H, Strender L. Assessment of nerve involvement in the lumbar spine: agreement between magnetic resonance imaging, physical examination and pain drawing findings. BMC Musculoskelet Disord 2010 Sep 10;11:202 [FREE Full text] [doi: 10.1186/1471-2474-11-202] [Medline: 20831785]
27. Jaatun EA, Jaatun MG. Advanced healthcare services enabled by a computerized pain body map. Procedia Comput Sci 2016;98:251-258. [doi: 10.1016/j.procs.2016.09.040]
28. Head H. On disturbances of sensation with especial reference to the pain of visceral disease. Brain 1893;16(1-2):1-133. [doi: 10.1093/brain/16.1-2.1]
29. Margoles MS. Letter to the editor. Pain 1980;8(1):115-117. [doi: 10.1016/0304-3959(80)90094-9]
30. Toomey TC, Gover VF, Jones BN. Spatial distribution of pain: a descriptive characteristic of chronic pain. Pain 1983 Nov;17(3):289-300. [doi: 10.1016/0304-3959(83)90101-x] [Medline: 6657289]
31. Udén A, Landin LA. Pain drawing and myelography in sciatic pain. Clin Orthop Relat Res 1987 Mar(216):124-130. [Medline: 3815937]
32. Jud SM, Fasching PA, Maihöfner C, Heusinger K, Loehberg CR, Hatko R, et al. Pain perception and detailed visual pain mapping in breast cancer survivors. Breast Cancer Res Treat 2010 Jan;119(1):105-110. [doi: 10.1007/s10549-009-0485-z] [Medline: 19641989]
33. Pierce H, Homer CS, Dahlen HG, King J. Pregnancy-related lumbopelvic pain: listening to Australian women. Nurs Res Pract 2012;2012:387428 [FREE Full text] [doi: 10.1155/2012/387428] [Medline: 22685643]
34. Elson DW, Jones S, Caplan N, Stewart S, St Clair GA, Kader DF. The photographic knee pain map: locating knee pain with an instrument developed for diagnostic, communication and research purposes. Knee 2011 Dec;18(6):417-423. [doi: 10.1016/j.knee.2010.08.012] [Medline: 20850976]
35. Tucker KJ, Fels M, Walker SR, Hodges PW. Comparison of location, depth, quality, and intensity of experimentally induced pain in 6 low back muscles. Clin J Pain 2014 Sep;30(9):800-808. [doi: 10.1097/AJP.0000000000000026] [Medline: 25098553]
36. Egloff N, Cámara RJ, von Känel R, Klingler N, Marti E, Ferrari ML. Pain drawings in somatoform-functional pain. BMC Musculoskelet Disord 2012 Dec 20;13:257 [FREE Full text] [doi: 10.1186/1471-2474-13-257] [Medline: 23256679]
37. Egsgaard LL, Christensen TS, Petersen IM, Brønnum DS, Boudreau SA. Do gender-specific and high-resolution three dimensional body charts facilitate the communication of pain for women? A quantitative and qualitative study. JMIR Hum Factors 2016 Jul 20;3(2):e19 [FREE Full text] [doi: 10.2196/humanfactors.5693] [Medline: 27440737]
38. Margoles M. The pain chart: spatial properties of pain. In: Melzack R, editor. Pain Measurement and Assessment. New York: Raven Press; 1983:215-225.
39. Masferrer R, Prendergast V, Hagell P. Colored pain drawings: preliminary observations in a neurosurgical practice. Eur J Pain 2003;7(3):213-217. [doi: 10.1016/S1090-3801(02)00113-1] [Medline: 12725843]
40. Bertilson BC, Grunnesjö M, Strender L. Reliability of clinical tests in the assessment of patients with neck/shoulder problems-impact of history. Spine (Phila Pa 1976) 2003 Oct 1;28(19):2222-2231. [doi: 10.1097/01.BRS.0000089685.55629.2E] [Medline: 14520035]
41. Türp JC, Kowalski CJ, Stohler CS. Generic pain intensity scores are affected by painful comorbidity. J Orofac Pain 2000;14(1):47-51. [Medline: 11203737]
42. Melzack R. The McGill pain questionnaire: major properties and scoring methods. Pain 1975 Sep;1(3):277-299. [doi: 10.1016/0304-3959(75)90044-5] [Medline: 1235985]
43. Mooney V, Cairns D, Robertson J. A system for evaluating and treating chronic back disability. West J Med 1976 May;124(5):370-376 [FREE Full text] [Medline: 132014]

44. Hildebrandt J, Franz CE, Choroba-Mehnen B, Temme M. The use of pain drawings in screening for psychological involvement in complaints of low-back pain. Spine (Phila Pa 1976) 1988 Jun;13(6):681-685. [Medline: 2972073]
45. Jamison RN, Washington TA, Gulur P, Fanciullo GJ, Arscott JR, McHugo GJ, et al. Reliability of a preliminary 3-D pain mapping program. Pain Med 2011 Mar;12(3):344-351 [FREE Full text] [doi: 10.1111/j.1526-4637.2010.01049.x] [Medline: 21276186]
46. Spyridonis F, Hansen J, Grønli T, Ghinea G. PainDroid: an Android-based virtual reality application for pain assessment. Multimed Tools Appl 2013 Jan 10;72(1):191-206. [doi: 10.1007/s11042-013-1358-3]
47. Gibson J, Frank A. Pain experienced by electric-powered chair users: a pilot exploration using pain drawings. Physiother Res Int 2005;10(2):110-115. [doi: 10.1002/pri.31] [Medline: 16146328]
48. Udén A, Aström M, Bergenudd H. Pain drawings in chronic back pain. Spine (Phila Pa 1976) 1988 Apr;13(4):389-392. [doi: 10.1097/00007632-198804000-00002] [Medline: 2970123]
49. Toomingas A. Characteristics of pain drawings in the neck-shoulder region among the working population. Int Arch Occup Environ Health 1999 Mar;72(2):98-106. [doi: 10.1007/s004200050344] [Medline: 10197481]
50. George SZ, Bialosky JE, Wittmer VT, Robinson ME. Sex differences in pain drawing area for individuals with chronic musculoskeletal pain. J Orthop Sports Phys Ther 2007 Mar;37(3):115-121. [doi: 10.2519/jospt.2007.2399] [Medline: 17416126]
51. Bunyaratavej K, Montriwiwatnchai P, Siwanuwatn R, Khaoroptham S. Localizing value of pain distribution patterns in cervical spondylosis. Asian Spine J 2015 Apr;9(2):210-217 [FREE Full text] [doi: 10.4184/asj.2015.9.2.210] [Medline: 25901232]
52. Post WR, Fulkerson J. Knee pain diagrams: correlation with physical examination findings in patients with anterior knee pain. Arthroscopy 1994 Dec;10(6):618-623. [doi: 10.1016/S0749-8063(05)80058-1] [Medline: 7880352]
53. Türp JC, Kowalski CJ, Stohler CS. Temporomandibular disorders--pain outside the head and face is rarely acknowledged in the chief complaint. J Prosthet Dent 1997 Dec;78(6):592-595. [doi: 10.1016/s0022-3913(97)70010-6] [Medline: 9421788]
54. Reigo T, Tropp H, Timpka T. Pain drawing evaluation--the problem with the clinically biased surgeon. Intra- and interobserver agreement in 50 cases related to clinical bias. Acta Orthop Scand 1998 Aug;69(4):408-411. [doi: 10.3109/17453679808999057] [Medline: 9798452]
55. Gatchel RJ, Mayer TG, Capra P, Diamond P, Barnett J. Quantification of lumbar function. Part 6: the use of psychological measures in guiding physical functional restoration. Spine (Phila Pa 1976) 1986;11(1):36-42. [Medline: 2939568]
56. Fordyce WE, Brockway JA, Bergman JA, Spengler D. Acute back pain: a control-group comparison of behavioral vs traditional management methods. J Behav Med 1986 Apr;9(2):127-140. [doi: 10.1007/BF00848473] [Medline: 2940370]
57. Alonso-Blanco C, Fernández-de-Las-Peñas C, de-la-Llave-Rincón AI, Zarco-Moreno P, Galán-Del-Río F, Svensson P. Characteristics of referred muscle pain to the head from active trigger points in women with myofascial temporomandibular pain and fibromyalgia syndrome. J Headache Pain 2012 Nov;13(8):625-637 [FREE Full text] [doi: 10.1007/s10194-012-0477-y] [Medline: 22935970]
58. Margolis RB, Krause SJ, Tait RC. Lateralization of chronic pain. Pain 1985 Nov;23(3):289-293. [doi: 10.1016/0304-3959(85)90107-1] [Medline: 4069726]
59. Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. Pain 1986 Jan;24(1):57-65. [doi: 10.1016/0304-3959(86)90026-6] [Medline: 2937007]
60. Escalante A, Lichtenstein MJ, White K, Rios N, Hazuda HP. A method for scoring the pain map of the McGill pain questionnaire for use in epidemiologic studies. Aging (Milano) 1995 Oct;7(5):358-366. [Medline: 8719602]
61. 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]
62. Ransford AO, Cairns D, Mooney V. The pain drawing as an aid to the psychologic evaluation of patients with low-back pain. Spine 1976;1(2):127-134. [doi: 10.1097/00007632-197606000-00007]
63. Türp JC, Kowalski CJ, Stohler CS. Greater disability with increased pain involvement, pain intensity and depressive preoccupation. Eur J Pain 1997;1(4):271-277. [doi: 10.1016/S1090-3801(97)90036-7] [Medline: 15102392]
64. Wallace MS, North J, Grigsby EJ, Kapural L, Sanapati MR, Smith SG, et al. An integrated quantitative index for measuring chronic multisite pain: the multiple areas of pain (MAP) study. Pain Med 2018 Feb 21 (epub ahead of print). [doi: 10.1093/pm/pnx325] [Medline: 29474648]
65. Wolfe F, Clauw DJ, Fitzcharles M, Goldenberg DL, Häuser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR preliminary diagnostic criteria for fibromyalgia. J Rheumatol 2011 Jun;38(6):1113-1122. [doi: 10.3899/jrheum.100594] [Medline: 21285161]
66. Wolfe F, Clauw DJ, Fitzcharles M, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010 May;62(5):600-610 [FREE Full text] [doi: 10.1002/acr.20140] [Medline: 20461783]
67. Donelson R, Grant W, Kamps C, Medcalf R. Pain response to sagittal end-range spinal motion. A prospective, randomized, multicentered trial. Spine (Phila Pa 1976) 1991 Jun;16(6 Suppl):S206-S212. [doi: 10.1097/00007632-199106001-00006] [Medline: 1830700]

68. Parker H, Wood PL, Main CJ. The use of the pain drawing as a screening measure to predict psychological distress in chronic low back pain. Spine (Phila Pa 1976) 1995 Jan 15;20(2):236-243. [doi: 10.1097/00007632-199501150-00022] [Medline: 7716631]
69. Sivik TM, Gustafsson E, Olsson KK. Differential diagnosis of low-back pain patients: a simple quantification of the pain drawing. Nord J Psychiatry 2009 Jul 12;46(1):55-62. [doi: 10.3109/08039489209106183]
70. Boudreau SA, Kamavuako EN, Rathleff MS. Distribution and symmetrical patellofemoral pain patterns as revealed by high-resolution 3D body mapping: a cross-sectional study. BMC Musculoskelet Disord 2017 Apr 18;18(1):160 [FREE Full text] [doi: 10.1186/s12891-017-1521-5] [Medline: 28420379]
71. Mann NH, Brown MD, Enger I. Expert performance in low-back disorder recognition using patient pain drawings. J Spinal Disord 1992 Sep;5(3):254-259. [Medline: 1387821]
72. Zhang C, Kjellstrom H, Ek CH, Bertilson B. Diagnostic Prediction Using Discomfort Drawings with IBTM. In: Proceedings of the 1st Machine Learning for Healthcare Conference. 2016 Presented at: PMLR'16; August 19-20, 2016; Los Angeles, CA, USA p. 226-238 URL: http://proceedings.mlr.press/v56/Zhang16.pdf
73. Boudreau SA, Royo AC, Matthews M, Graven-Nielsen T, Kamavuako EN, Slabaugh G, et al. Distinct patterns of variation in the distribution of knee pain. Sci Rep 2018 Nov 8;8(1):16522 [FREE Full text] [doi: 10.1038/s41598-018-34950-2] [Medline: 30410031]
74. Türp JC, Kowalski CJ, O'Leary N, Stohler CS. Pain maps from facial pain patients indicate a broad pain geography. J Dent Res 1998 Jun;77(6):1465-1472. [doi: 10.1177/00220345980770061101] [Medline: 9649175]
75. Slipman CW, Plastaras CT, Palmitier RA, Huston CW, Sterenfeld EB. Symptom provocation of fluoroscopically guided cervical nerve root stimulation. Are dynatomal maps identical to dermatomal maps? Spine (Phila Pa 1976) 1998 Oct 15;23(20):2235-2242. [doi: 10.1097/00007632-199810150-00019] [Medline: 9802168]
76. Svensson P, Graven-Nielsen T, Arendt-Nielsen L. Mechanical hyperesthesia of human facial skin induced by tonic painful stimulation of jaw muscles. Pain 1998 Jan;74(1):93-100. [doi: 10.1016/s0304-3959(97)00156-5] [Medline: 9514565]
77. Slipman CW, Plastaras C, Patel R, Isaac Z, Chow D, Garvan C, et al. Provocative cervical discography symptom mapping. Spine J 2005;5(4):381-388. [doi: 10.1016/j.spinee.2004.11.012] [Medline: 15996607]
78. Ghinea G, Gill D, Frank A, de Souza LH. Using geographical information systems for management of back-pain data. J Manag Med 2002;16(2-3):219-237. [doi: 10.1108/02689230210434952] [Medline: 12211347]
79. Hwang M, Kang YK, Kim DH. Referred pain pattern of the pronator quadratus muscle. Pain 2005 Aug;116(3):238-242. [doi: 10.1016/j.pain.2005.04.015] [Medline: 15964683]
80. Neubert T, Dusch M, Karst M, Beissner F. Designing a tablet-based software app for mapping bodily symptoms: usability evaluation and reproducibility analysis. JMIR Mhealth Uhealth 2018 May 30;6(5):e127 [FREE Full text] [doi: 10.2196/mhealth.8409] [Medline: 29848470]
81. Nummenmaa L, Glerean E, Hari R, Hietanen JK. Bodily maps of emotions. Proc Natl Acad Sci U S A 2014 Jan 14;111(2):646-651 [FREE Full text] [doi: 10.1073/pnas.1321664111] [Medline: 24379370]
82. Nummenmaa L, Hari R, Hietanen JK, Glerean E. Maps of subjective feelings. Proc Natl Acad Sci U S A 2018 Dec 11;115(37):9198-9203 [FREE Full text] [doi: 10.1073/pnas.1807390115] [Medline: 30154159]
83. García-Magariño I, Chittaro L, Plaza I. Bodily sensation maps: exploring a new direction for detecting emotions from user self-reported data. Int J Hum-Comput St 2018 May;113:32-47. [doi: 10.1016/j.ijhcs.2018.01.010]
84. Beissner F, Brünner F, Fink M, Meissner K, Kaptchuk TJ, Napadow V. Placebo-induced somatic sensations: a multi-modal study of three different placebo interventions. PLoS One 2015;10(4):e0124808 [FREE Full text] [doi: 10.1371/journal.pone.0124808] [Medline: 25901350]
85. Beissner F, Marzolff I. Investigation of acupuncture sensation patterns under sensory deprivation using a geographic information system. Evid Based Complement Alternat Med 2012;2012:591304 [FREE Full text] [doi: 10.1155/2012/591304] [Medline: 23243458]
86. Jung W, Shim W, Lee T, Park H, Ryu Y, Beissner F, et al. More than DeQi: spatial patterns of acupuncture-induced bodily sensations. Front Neurosci 2016;10:462 [FREE Full text] [doi: 10.3389/fnins.2016.00462] [Medline: 27807402]
87. Neubert T, Beissner F. Figshare. 2019. Hannover Body Template URL: https://figshare.com/articles/Hannover_Body_Template/7637387/5 [accessed 2019-08-15]
88. Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ. SMPL: a skinned multi-person linear model. ACM Trans Graph 2015 Oct 26;34(6):1-16. [doi: 10.1145/2816795.2818013]
89. MPI SMPL. 2019. SMPL: A Skinned Multi-Person Linear Model URL: http://smpl.is.tue.mpg.de/ [accessed 2019-05-06] [WebCite Cache ID 78A8cKNsQ]
90. Fujieda K, Okubo K. A reusable anatomically segmented digital mannequin for public health communication. J Vis Commun Med 2016;39(1-2):18-26. [doi: 10.1080/17453054.2016.1182473] [Medline: 27311009]
91. Mitsuhashi N, Fujieda K, Tamura T, Kawamoto S, Takagi T, Okubo K. BodyParts3D: 3D structure database for anatomical concepts. Nucleic Acids Res 2009 Jan;37(Database Issue):D782-D785 [FREE Full text] [doi: 10.1093/nar/gkn613] [Medline: 18835852]
92. Poulsen E, Overgaard S, Vestergaard JT, Christensen HW, Hartvigsen J. Pain distribution in primary care patients with hip osteoarthritis. Fam Pract 2016 Dec;33(6):601-606. [doi: 10.1093/fampra/cmw071] [Medline: 27538424]

93. Gerhardt A, Eich W, Treede R, Tesarz J. Conditioned pain modulation in patients with nonspecific chronic back pain with chronic local pain, chronic widespread pain, and fibromyalgia. Pain 2017 Mar;158(3):430-439. [doi: 10.1097/j.pain.0000000000000777] [Medline: 27902566]
94. Barbero M, Moresi F, Leoni D, Gatti R, Egloff M, Falla D. Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis. Eur J Pain 2015 Sep;19(8):1129-1138. [doi: 10.1002/ejp.636] [Medline: 25565607]

Abbreviations

2D: two-dimensional
3D: three-dimensional
ANN: artificial neural network
IPQI: Integrated Pain Quantification Index
MPQ: McGill pain questionnaire
PC: personal computer
PCA: principal component analysis
PD: pain drawing
PFP: patellofemoral pain
WPI: widespread pain index

Edited by G Eysenbach; submitted 06.05.19; peer-reviewed by UV Albrecht, J Galvez-Olortegui; comments to author 29.05.19; revised version received 17.06.19; accepted 27.06.19; published 05.09.19.

Please cite as:
Shaballout N, Neubert TA, Boudreau S, Beissner F
From Paper to Digital Applications of the Pain Drawing: Systematic Review of Methodological Milestones
JMIR Mhealth Uhealth 2019;7(9):e14569
URL: https://mhealth.jmir.org/2019/9/e14569/
doi:10.2196/14569
PMID:31489841

Original Paper

Home Videos as a Cost-Effective Tool for the Diagnosis of Paroxysmal Events in Infants: Prospective Study

Lu-Lu Huang[1*], MD; Yang-Yang Wang[1*], MD; Li-Ying Liu[1], MD; Hong-Ping Tang[2], MD; Meng-Na Zhang[1], MD; Shu-Fang Ma[1], MD; Li-Ping Zou[1], PhD

[1]Chinese People's Liberation Army General Hospital, Beijing, China

[2]Chang-De First People's Hospital, Changde, China

[*]these authors contributed equally

Corresponding Author:
Li-Ping Zou, PhD
Chinese People's Liberation Army General Hospital
No 28 Fuxing Road, Haidian District
Beijing, 100853
China
Phone: 86 13911880469
Email: zouliping21@hotmail.com

Abstract

Background: The diagnosis of paroxysmal events in infants is often challenging. Reasons include the child's inability to express discomfort and the inability to record video electroencephalography at home. The prevalence of mobile phones, which can record videos, may be beneficial to these patients. In China, this advantage may be even more significant given the vast population and the uneven distribution of medical resources.

Objective: The aim of this study is to investigate the value of mobile phone videos in increasing the diagnostic accuracy and cost savings of paroxysmal events in infants.

Methods: Clinical data, including descriptions and home videos of episodes, from 12 patients with paroxysmal events were collected. The investigation was conducted in six centers during pediatric academic conferences. All 452 practitioners present were asked to make their diagnoses by just the descriptions of the events, and then remake their diagnoses after watching the corresponding home videos of the episodes. The doctor's information, including educational background, profession, working years, and working hospital level, was also recorded. The cost savings from accurate diagnoses were measured on the basis of using online consultation, which can also be done easily by mobile phone. All data were recorded in the form of questionnaires designed for this study.

Results: We collected 452 questionnaires, 301 of which met the criteria (66.6%) and were analyzed. The mean correct diagnoses with and without videos was 8.4 (SD 1.7) of 12 and 7.5 (SD 1.7) of 12, respectively. For epileptic seizures, mobile phone videos increased the mean accurate diagnoses by 3.9%; for nonepileptic events, it was 11.5% and both were statistically different (P=.006 for epileptic events; P<.001 for nonepileptic events). Pediatric neurologists with longer working years had higher diagnostic accuracy; whereas, their working hospital level and educational background made no difference. For patients with paroxysmal events, at least US $673.90 per capita and US $128 million nationwide could be saved annually, which is 12.02% of the total cost for correct diagnosis.

Conclusions: Home videos made on mobile phones are a cost-effective tool for the diagnosis of paroxysmal events in infants. They can facilitate the diagnosis of paroxysmal events in infants and thereby save costs. The best choice for infants with paroxysmal events on their initial visit is to record their events first and then show the video to a neurologist with longer working years through online consultation.

(JMIR Mhealth Uhealth 2019;7(9):e11229) doi:10.2196/11229

KEYWORDS

paroxysmal events; infant; home videos; online consultation

Introduction

Paroxysmal events in infants are characterized by sudden, mostly short-term, involuntary movements involving various parts of their body [1-3]. Some are epileptic seizures, whereas some are nonepileptic seizures resulting from immaturity of the central nervous system or other pathological or nonepileptic mechanisms [4,5]. Approximately 20% to 40% of patients in epilepsy referral centers are diagnosed with paroxysmal nonepileptic events [6]. Bye et al [3] reported that paroxysmal nonepileptic events are diagnosed in 43% of children who underwent video electroencephalography (VEEG) monitoring. However, there are still difficulties in accurately diagnosing paroxysmal nonepileptic events in this population [7,8]. First, because infants are unable to express their complaints; seizure descriptions are mostly made by their caregivers, who have never been trained to identify seizure types. Thus, the descriptions may be inaccurate or incomplete [9]. Second, without video recordings it is more difficult to make a diagnosis [10].

Video EEG, the gold standard in the differential diagnosis of paroxysmal events [11,12], plays a central role in clinical work [10,13-15]. With VEEG, the diagnostic accuracy rate can reach up to 88.0% [16], whereas that of ambulatory EEG is only 67.5% [17]. However, paroxysmal events usually occur at home with no aura, so a VEEG test monitoring the paroxysmal events is unavailable in most cases. Worse still, many centers in China do not own VEEG test facilities, so some sporadic events may never be recorded. In such cases, clinical acumen based on the description of the events is the only assistance clinicians have to make their diagnoses [16].

In China, the contradiction between the surging number of children and the lack of pediatricians, and between the concentrated distribution of high-level hospitals (mostly in developed areas) and widespread distribution of patients (many in remote areas) makes it difficult for many patients to get a timely diagnosis and follow-up treatment. Except for Beijing and Shanghai, the supply-and-demand ratios of pediatricians are less than 0.80, and there is a need for another 191,981 to 198,287 pediatricians [18]. Another harsh reality is that family of patients from remote and less-developed areas have to spend a lot of money on traffic and accommodation fees. All these have greatly exacerbated the issue of “difficulties and high costs of getting medical services” in China [19].

With the increasing popularity of mobile phones worldwide, recording the paroxysmal events has become easier. With mobile phones, patients can choose an online consultation instead of visiting a traditional outpatient clinic, which will greatly decrease their costs. There have been many studies on the value of home video and its contribution to increasing diagnostic accuracy [20,21]. However, none focused on the value of home video for paroxysmal events in infants. Also, none provided patients with guidance on what doctors they should visit after recording the events.

This study aims to identify the value of home videos for the diagnosis of paroxysmal events and its potential use for online consultation and providing guidance about which doctor to choose when patients or their caregivers have recorded events and chosen online consultation.

Methods

Study Design

The study is a prospective study with three steps: investigating the value of the home videos for diagnosis on the patients’ first visit, the cost savings for these patients with videos when choosing online consultation, and the type of doctor they should choose.

The trial was conducted following the international rules of good clinical practice. The study was approved by the General Hospital of the People’s Liberation Army. Informed consent was obtained from each patient’s parents. The parents would describe the paroxysmal events to the senior epileptologist, who would edit the description to meet the criteria of the clinical medical record and include any relevant personal or family history. Therefore, each patient’s description information was complete and would accurately simulate the patient’s consultation process in the outpatient clinic.

Video Selection

Twelve video recordings showing various paroxysmal events in infants were collected from the Chinese PLA General Hospital outpatient clinic from May 2015 to January 2016. The inclusion and exclusion criteria are listed in Textbox 1. Similar videos were removed until the shortest remained, but we ensured that each typical episode was presented. The flowchart for choosing the videos is presented in Figure 1. The VEEG reports of all 12 patients were also collected.

Textbox 1. Inclusion and exclusion criteria for the videos.

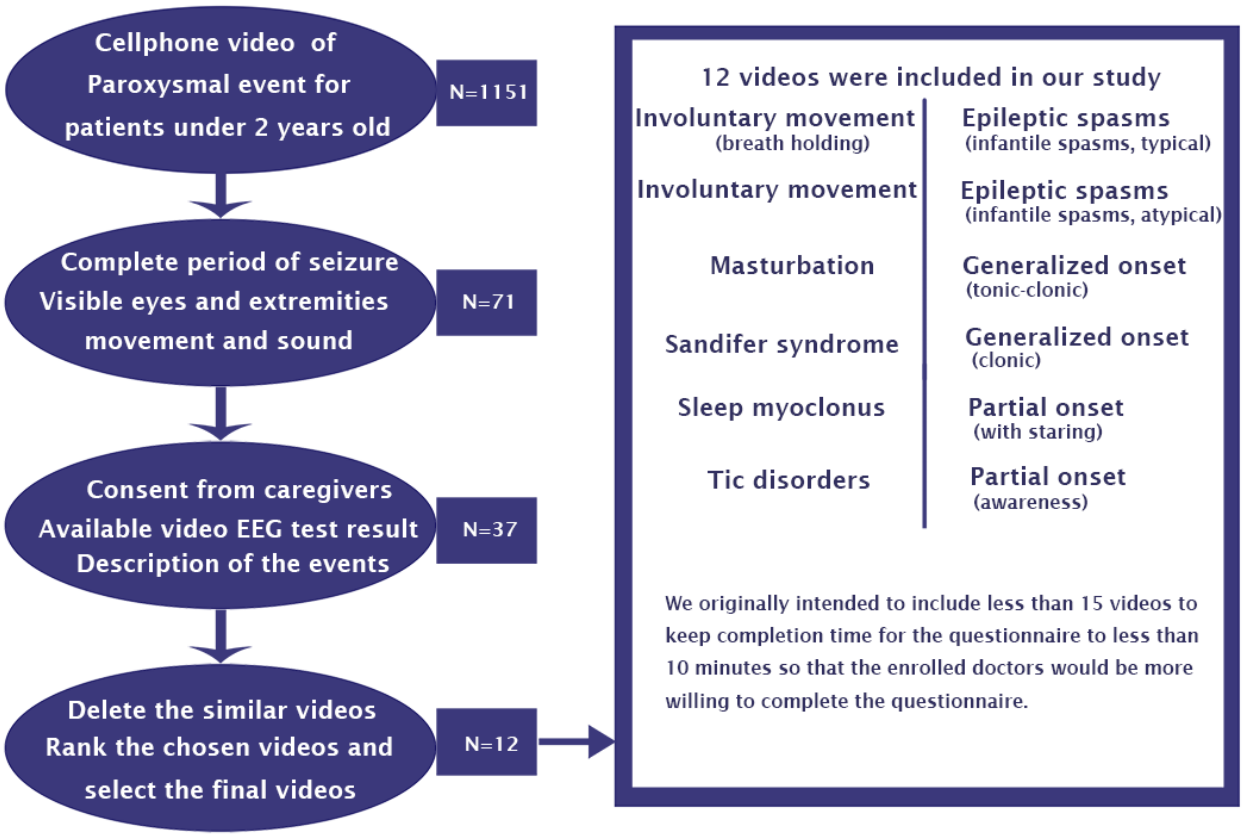
Inclusion criteria

1. The resolution of the videos is high enough to ensure that the patient’s facial features are visible;
2. All possible body movements of the patient were recorded; and
3. The sound in the videos is clear, and whether there is excessive ventilation can be distinguished.

Exclusion criteria

1. No consent has been achieved from the patients’ caregivers; and
2. The video is longer than 1 minute (may affect the efficiency of public playback).

Figure 1. Flowchart for selecting qualified videos.

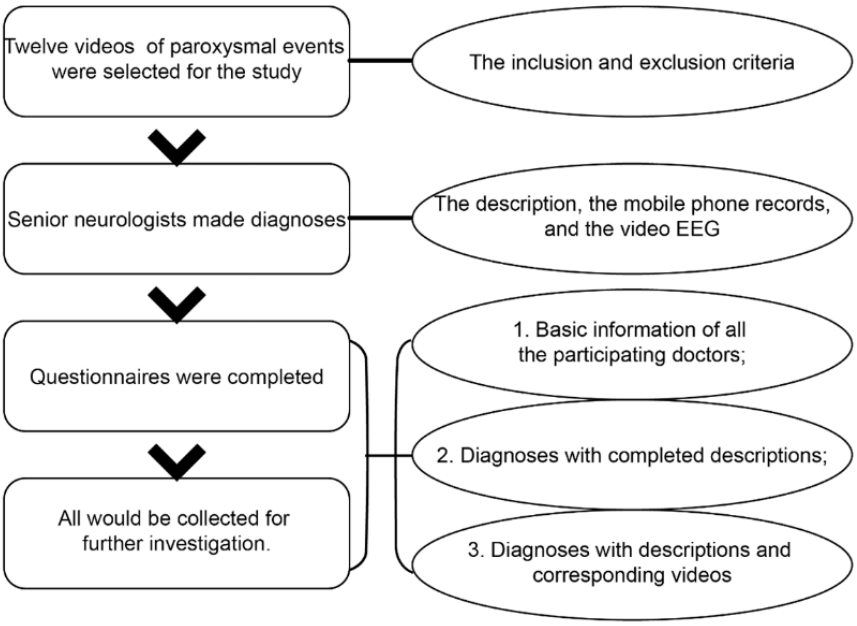


All corresponding descriptions, home videos, and VEEG reports were presented to two senior epileptologists blind to the study purpose, and they made diagnoses accordingly. Events were categorized as epileptic or nonepileptic: if epileptic, the specific seizure type was listed; if nonepileptic, a diagnosis to explain the paroxysmal events was given. When the diagnoses from the two epileptologists were not the same, a third epileptologist would review the data and provide the diagnoses. We did not encounter a situation in which all three reviewers could not achieve an agreement.

Data Collection

We conducted our investigation in six centers during the pediatric academic conferences. The person playing the videos was unaware of the diagnosis. Figure 2 shows the three steps of the study. A questionnaire was designed to simulate the process of a clinic consultation. The first part of the questionnaire was the basic information of the doctors, including their educational background, profession (pediatrician or pediatric neurologist), working hospital levels (first/secondary/tertiary hospital), and working years. This was because, in clinical work, the doctor's basic information is open to all patients visiting the outpatient clinic. The second part provided the doctors with the description of the episodes. This part simulated the process of collecting the medical record at the beginning of the patient's visit. The description provided all the information that the patient would provide to the clinician when there was not a video. The doctors were required to make their diagnoses and fill in the questionnaire. It should be noted that we originally required the doctor to identify the specific type of epilepsy when they considered it an epileptic seizure, but when we simulated the process in our hospital, we found that the general pediatricians could not clearly identify the types of epilepsy. Therefore, we deleted this part and only required the doctors to identify the episodes as epileptic or nonepileptic. The third part of the questionnaire presented the same descriptions as the second part, but the difference was that in the process of actually collecting these data, the doctors would simultaneously see the corresponding videos before making their diagnoses. Before the investigation, the enrolled doctors were informed of the purpose of the study; all data were anonymous. Only the completed and identifiable questionnaires were eligible for our study.

Figure 2. Flowchart of the study of the value of videos. EEG: electroencephalography.



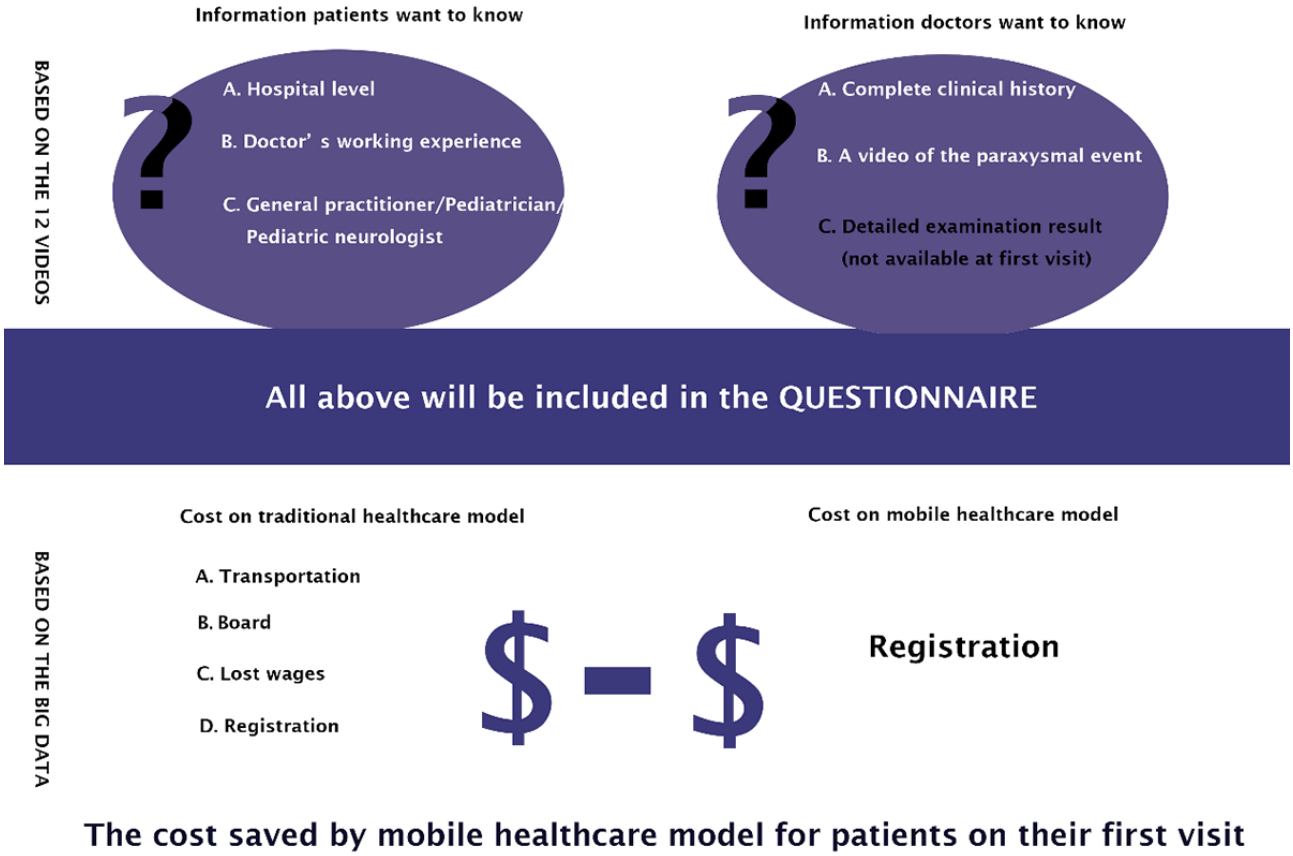
Data Analysis

We calculated the cost savings with the help of home videos on the basis of the contribution videos made on correct diagnoses. In this part, we used the model of online consultation instead of the traditional outpatient visit because this is the increasingly popular choice in China with the popularity of mobile phones [22]. Also, we believed that online consultation is a more reasonable choice considering the difficulties and high costs of getting medical services in China, especially for patients from remote areas [23-25]. The cost savings by online consultation and for patients with or without videos were obtained from published articles or government work reports, including transportation fees, accommodation fees, loss of wages, and registration fees. Figure 3 illustrates this simulation process and how the cost savings were calculated.

All data were recorded in Epidata and exported to SPSS version 24 for statistical analysis. Calculations for the correct diagnosis percentage for each episode were made, and the difference was compared by chi-square test. For the impact of the doctors' background on diagnosis, we gave a correct diagnosis a value of 1 and an incorrect diagnosis a value of 0, and then calculated each doctor's score with and without the help of home videos. We then compared the influence of doctors' profession, educational background, working hospital level, and working years on accurate diagnosis by multiple linear regression analysis. We further analyzed the scores of each physician on the diagnosis of epileptic seizures and nonepileptic seizures with or without videos to analyze the role of the doctor's background information in both cases. The statistical tests presented are two-sided; a *P* value less than .05 was considered statically significant.

All patients, if they consult online for their first visit, can save the cost of transportation, accommodation, and loss of wages. Because there has not been any standard for the charge of online consultation in China, we equated the fee to the traditional outpatient registration fee, so it could not be saved. There are also no data on the number of pediatric patients with epilepsy; therefore, we first estimated the annual incidence and the total number of new epilepsy patients through related literature [12,26-32]. Then, we calculated the total number of new patients annually who presented with a paroxysmal event and were diagnosed as a nonepileptic event according to the ratio of the epileptic events to nonepileptic events in all paroxysmal events in children [33,34]. The proportion of infant patients was not available, so when calculating the total cost savings of bringing videos and choosing online consultation, we extended the age to 9 years and calculated the total cost savings for patients younger than 9 years nationwide. In addition, if the patients did not obtain accurate diagnoses on their first visit, they would have to pay additional charges on a later visit. We will never know how many visits these patients needed before accurate diagnosis; therefore, we assumed that they would all achieve an accurate diagnosis on their second visit so the cost savings would be the minimum. We then compared the cost savings of bringing videos with that of just a description to study the value of videos. The formulas are presented in Multimedia Appendix 1.

Figure 3. The simulation process and the value of videos on cost savings.



Results

Mobile Phone Videos and Diagnosis Accuracy

A total of 452 questionnaires were collected, 301 of which met the criteria (66.6%) and were included in the study. Table 1 shows the demographic characteristics of the whole study sample.

The mean number of correct diagnoses through descriptions only and descriptions with home videos was 7.5 (SD 1.7) of 12 and 8.4 (SD 1.7) of 12, respectively. Details of the percentages of correct diagnoses are shown in Table 2.

Table 1. Demographic characteristics of the whole study sample (N=301).

Demographic items	Study sample
Age (months), mean (SD)	16.0 (37.1)
Gender, n (%)	
Male	79 (26.2)
Female	222 (73.8)
Hospital, n (%)	
Secondary level	93 (30.9)
Tertiary level	208 (69.1)
Duration of working, n (%)	
<10 years	149 (49.5)
≥10 years	152 (50.5)
Profession, n (%)	
Pediatrician	218 (72.4)
Pediatric neurologist	83 (27.6)

Table 2. Classification of paroxysmal events and percentage of doctors who correctly identified the episodes (N=301).

Video	Seizure classification	Correct identification, n (%)		*P* value
		Description only	Description with video	
Epileptic				
1	Partial	266 (88.4)	262 (87.0)	.78
3	Spasm	233 (77.4)	261 (86.7)	.048
5	Generalized	262 (87.0)	291 (96.7)	.04
7	Partial	274 (91.0)	267 (88.7)	.62
10	Generalized	260 (86.4)	262 (87.0)	.89
11	Spasm	232 (77.1)	260 (87.0)	.048
Nonepileptic				
2	Involuntary movements	95 (31.6)	178 (59.1)	<.001
4	Sleep myoclonus	204 (67.8)	230 (76.4)	.07
6	Tic disorders	152 (50.5)	198 (65.8)	.001
8	Sandifer syndrome	110 (36.5)	86 (28.6)	.09
9	Involuntary movements	85 (28.2)	168 (55.8)	<.001
12	Masturbation	83 (27.6)	119 (39.5)	.01

For epileptic events, the mean percentage of correct diagnoses with only a description available was 84.9% (SD 5.9%), and with a home video it was 88.8% (SD 3.9%). The difference was statistically significant (*P*<.001). For nonepileptic events, the mean percentage of correct diagnoses with only a description available was 39.9% (SD 15.9%), and with a home video it was 51.4% (SD 17.5%). The difference was also statistically significant (*P*<.001).

Cost Savings With Home Videos

The means of the transportation fees, accommodation fees, further examination fees, and loss of wages according to the published literature and governmental reports [26-31] were US $290.00 (SD $56.14), US $86.30 (SD $22.54), US $192.90 (SD $61.73), and US $104.70 (SD $90.02), respectively. The annual new pediatric patient (younger than 9 years) population is 197,945 in China. Online consultation could save US $1.28 million (8.22 million yuan) per year. For infants with paroxysmal events with videos on their first visit, at least 12.02% of the total cost can be saved.

Which Doctor to Choose

In the analyses of doctor's background information, we found that the number of working years was the key factor for a correct diagnosis. Whether or not videos were provided, the level of the doctor's hospital and educational background were irrelevant to the correctness of the diagnosis. When only parents' descriptions were available, profession was the only factor that affected the correctness of the diagnosis for epileptic seizures; for nonepileptic seizures, pediatric neurologists with relatively longer working years made more accurate diagnoses. When both descriptions and videos were available, profession and hospital level were the two factors that affected the correctness of the diagnosis for epileptic seizures; for nonepileptic seizures, working years and profession were related to the accuracy of the diagnosis. The statistical results are listed in Table 3.

Table 3. Variables of doctors' backgrounds and the effect on correct diagnoses.

Variable	Standardized coefficients, beta (95% CI)	*P* value
Description		
Epileptic		
Education level	.042 (−.078, .162)	.49
Hospital level	−.062 (−.186, .062)	.33
Working years	−.096 (−.208, .017)	.09
Profession	.125 (.005, .246)	.04
Nonepileptic		
Education level	.046 (−.073, .165)	.45
Hospital level	−.056 (−.178, .066)	.37
Working years	.194 (.083, .305)	.001
Profession	.122 (.003, .241)	.04
Total		
Education level	.067 (−.050, .185)	.26
Hospital level	−.088 (−.210, .033)	.15
Working years	.118 (.008, .228)	.26
Profession	.187 (.069, .304)	.002
Video		
Epileptic		
Education level	−.030 (−.149, .090)	.62
Hospital level	−.152 (−.275, −.029)	.02
Working years	.128 (.017, .239)	.02
Profession	.113 (−.006, .233)	.06
Nonepileptic		
Education level	.053 (−.066, .173)	.38
Hospital level	.008 (−.115, .131)	.90
Working years	.144 (.012, .033)	.01
Profession	.163 (.008, .043)	.008
Total		
Education level	.039 (−.077, .155)	.51
Hospital level	−.065 (−.184, .055)	.29
Working years	.204 (.096, .313)	<.001
Profession	.216 (.100, .332)	<.001

Discussion

It is better for patients to take videos of paroxysmal events for their first visit, whether they will be diagnosed with epilepsy or not. Videos contribute a lot for the differential diagnosis. In our study, the availability of home videos increased the mean correct diagnosis percentage by 3.9% for epileptic events and11.5% for nonepileptic events. Compared with descriptions alone, videos are better at reflecting all the information of paroxysmal events. In a previous study including 45 semiological signs that can be used to distinguish paroxysmal nonepileptic events from epileptic events, only six proved reliable, and eyewitness reports were unreliable [9]. Due to the lack of relevant knowledge, parents' descriptions may exaggerate some clinical symptoms, which will influence doctors' judgments. The videos recorded by the caregivers are relatively more objective and can avoid this situation.

If there are videos available, some patients will be diagnosed with nonepileptic seizures just by the video recording and there will be no need for them to perform more examinations. Although some will be diagnosed as epileptic seizures, the videos can also save their cost on examination because the video can aid interpretation of ambulatory EEG in approximately one-third of patients and they may no longer need to perform

a VEEG test [35]. Thus, videos can help doctors make correct diagnoses easier and earlier, and the benefit of early diagnosis can sometimes be huge because it may improve the prognosis of epileptic infants [36]. For a paroxysmal event, the doctors may be able to identify it as a nonepileptic through video easily and another 11.5% of the nonepileptic patients and their families will then bear a lighter mental and financial burden from subsequent diagnoses and treatment.

When home videos are available, online consultation is a better choice for their first visit. Previous studies have shown that the incidence of epileptic misdiagnosis in online diagnosis and the treatment process was no different from that of the traditional outpatient process [37]. Moreover, online diagnosis and treatment have a considerable advantage in integrating medical resources and reducing patients' costs [38-41]. For patients during their first visit, there have been no examination results available, especially EEG monitoring. They may spend a lot on transportation, accommodation, loss in wages, and registration only to find that doctors need more examination results and they need to record the events before making a final diagnosis. They have to return home or live near the hospital for several days before their appointment date for further examinations or to record the events. Many costs would have been saved if they had used the online consultation after recording the events. According to our study, online consultation could save US $673.90 per capita and US $1.28 million (8.22 million yuan) in total per year for these patients. If these patients recorded the paroxysmal events, at least 12.02% of the total cost for correct diagnosis could be saved.

Therefore, we recommend all parents of infant patients with paroxysmal events record the events and choose online consultation for their first visit. This method is more important for to nonepileptic patients because it not only reduces their economic burden, but also saves them from unnecessarily worrying about an epilepsy diagnosis.

After recording the events and choosing online consultation, the patients may question which doctor to visit. From our study, experienced doctors—especially experienced neurologists—rather than doctors in higher-level hospitals or with better-educated backgrounds, are the best candidates for patients on their first visit. For the first visit, the caregivers naturally think that the better-educated and experienced neurologists in the higher-level hospital are their first choice. However, the level of hospital and a doctor's educational background were not important according to our study. Nagy et al [42] have reported that the diagnostic accuracy of first-year medical students was even lower than that of patients' parents.

In this study, we included a special nonepileptic event, Sandifer syndrome, which is a type of gastroesophageal reflux associated with laryngospasm. In infants, it may be misdiagnosed as seizures because of the presence of limb posturing, abnormal eye movements, and even opisthotonus [43]. These events mostly occur in sleep and thus cause diagnostic difficulties [12]. In our study, instead of increasing the identification rates, the videos confused the clinicians, who thought it was an epileptic seizure. Therefore, VEEG is still necessary in some situations.

Our study excluded the videos with poor quality, but videos from patients' caregivers may not be of high enough quality in real clinical work, and this will influence the diagnosis [44]. Although we emphasize the importance of recording videos, the quality of videos should also be considered.

There are some limitations of our study. First, although we tried to ensure that the descriptions were complete, we cannot fully equate the descriptions to a conventional outpatient visit. Second, the cost savings from our study are just the minimum because we will never know, for specific patients, how many visits they may need before a correct diagnosis.

Home videos made on mobile phones are a cost-effective tool for the diagnosis of paroxysmal events in infants. They can facilitate correct diagnosis and thereby save their cost. Therefore, the best choice for infants with paroxysmal events on their initial visit is to record their events first and then show the video to a neurologist with longer working years through online consultation.

Acknowledgments

This case report was supported by The National Key Research and Development Program of China (2016YFC1000707).

The authors thank Qian Lu for assistance with data acquisition and study supervision.

Authors' Contributions

L-LH: analysis and interpretation of data, manuscript writing. Y-YW: acquisition and analysis of data, critical revision of the manuscript for grammar and structure. L-YL: critical revision of the manuscript for intellectual content. H-PT: acquisition of data, study supervision. M-NZ: acquisition of data, study supervision. S-FM: acquisition of data, study supervision. L-PZ: study concept and design, critical revision of the manuscript for intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Formulas for cost savings.
[PDF File (Adobe PDF File), 22KB - mhealth_v7i9e11229_app1.pdf]

References

1. Wasserman D, Herskovitz M. Epileptic vs psychogenic nonepileptic seizures: a video-based survey. Epilepsy Behav 2017 Dec;73:42-45. [doi: 10.1016/j.yebeh.2017.04.020] [Medline: 28605633]
2. Korff C, Nordli D. Paroxysmal events in infants: persistent eye closure makes seizures unlikely. Pediatrics 2005 Oct;116(4):e485-e486. [doi: 10.1542/peds.2005-0632] [Medline: 16199674]
3. Bye A, Kok D, Ferenschild F, Vles JS. Paroxysmal non-epileptic events in children: a retrospective study over a period of 10 years. J Paediatr Child Health 2000 Jun;36(3):244-248. [Medline: 10849225]
4. Facini C, Spagnoli C, Pisani F. Epileptic and non-epileptic paroxysmal motor phenomena in newborns. J Matern Fetal Neonatal Med 2016 Nov;29(22):3652-3659. [doi: 10.3109/14767058.2016.1140735] [Medline: 26918366]
5. Holmes G, Ben-Ari Y. The neurobiology and consequences of epilepsy in the developing brain. Pediatr Res 2001 Mar;49(3):320-325. [doi: 10.1203/00006450-200103000-00004] [Medline: 11228256]
6. Martin R, Burneo J, Prasad A, Powell T, Faught E, Knowlton R, et al. Frequency of epilepsy in patients with psychogenic seizures monitored by video-EEG. Neurology 2003 Dec 23;61(12):1791-1792. [doi: 10.1212/01.wnl.0000098890.13946.f5] [Medline: 14694050]
7. Uldall P, Alving J, Hansen L, Kibaek M, Buchholt J. The misdiagnosis of epilepsy in children admitted to a tertiary epilepsy centre with paroxysmal events. Arch Dis Child 2006 Mar;91(3):219-221 [FREE Full text] [doi: 10.1136/adc.2004.064477] [Medline: 16492886]
8. Beach R, Reading R. The importance of acknowledging clinical uncertainty in the diagnosis of epilepsy and non-epileptic events. Arch Dis Child 2005 Dec;90(12):1219-1222 [FREE Full text] [doi: 10.1136/adc.2004.065441] [Medline: 16131503]
9. Syed T, LaFrance WJ, Kahriman E, Hasan S, Rajasekaran V, Gulati D, et al. Can semiology predict psychogenic nonepileptic seizures? A prospective study. Ann Neurol 2011 Jun;69(6):997-1004. [doi: 10.1002/ana.22345] [Medline: 21437930]
10. Sohal A, Khan A, Hussain N. Prolonged video-EEG in identifying paroxysmal nonepileptic events in children with epilepsy: a useful tool. J Clin Neurophysiol 2014 Apr;31(2):149-151. [doi: 10.1097/WNP.0000000000000035] [Medline: 24691233]
11. Plioplys S, Asato M, Bursch B, Salpekar J, Shaw R, Caplan R. Multidisciplinary management of pediatric nonepileptic seizures. J Am Acad Child Adolesc Psychiatry 2007 Nov;46(11):1491-1495. [doi: 10.1097/chi.0b013e31814dab98] [Medline: 18049299]
12. Patel H, Dunn D, Austin J, Doss J, LaFrance WC, Plioplys S, et al. Psychogenic nonepileptic seizures (pseudoseizures). Pediatr Rev 2011 Jun;32(6):e66-e72. [doi: 10.1542/pir.32-6-e66] [Medline: 21632872]
13. Connolly M, Wong P, Karim Y, Smith S, Farrell K. Outpatient video-EEG monitoring in children. Epilepsia 1994;35(3):477-481. [Medline: 8026391]
14. Srikumar G, Bhatia M, Jain S, Maheshwari MC. Usefulness of short term video-EEG monitoring in children with frequent intractable episodes. Neurol India 2000 Mar;48(1):29-32 [FREE Full text] [Medline: 10751810]
15. Patel AD, Haridas B, Grinspan ZM, Stevens J. Utility of long-term video-EEG monitoring for children with staring. Epilepsy Behav 2017 Dec;68:186-191. [doi: 10.1016/j.yebeh.2017.01.002] [Medline: 28214777]
16. Malone A, Ryan C, Fitzgerald A, Burgoyne L, Connolly S, Boylan GB. Interobserver agreement in neonatal seizure identification. Epilepsia 2009 Sep;50(9):2097-2101 [FREE Full text] [doi: 10.1111/j.1528-1167.2009.02132.x] [Medline: 19490044]
17. Chen LS, Mitchell WG, Horton EJ, Carter Snead O. Clinical utility of video-EEG monitoring. Pediatr Neurol 1995 Apr;12(3):220-224. [doi: 10.1016/0887-8994(95)00021-7]
18. Song Q, Wang F, Song L, Zhuang N, Qiu J, Wang L, et al. The pediatrician demands and gaps under the universal two-child policy. Chinese J Health Policy 2016;9(2):65-70.
19. Tingting C. The grey correlation analysis of the structure of the total cost of health and the per capita medical cost in China. Chinese J Health Stat 2017 Jun;34(3):494-496.
20. Dash D, Sharma A, Yuvraj K, Renjith A, Mehta S, Vasantha P, et al. Can home video facilitate diagnosis of epilepsy type in a developing country? Epilepsy Res 2016 Dec;125:19-23. [doi: 10.1016/j.eplepsyres.2016.04.004] [Medline: 27328162]
21. Liu X, Wang R, Zhou D, Hong Z. Feasibility and acceptability of smartphone applications for seizure self-management in China: questionnaire study among people with epilepsy. Epilepsy Behav 2016 Feb;55:57-61. [doi: 10.1016/j.yebeh.2015.11.024] [Medline: 26745631]
22. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Dec 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
23. Zhou K, Zhang X, Ding Y, Wang D, Lu Z, Yu M. Inequality trends of health workforce in different stages of medical system reform (1985-2011) in China. Hum Resour Health 2015 Dec 08;13:94 [FREE Full text] [doi: 10.1186/s12960-015-0089-0] [Medline: 26645960]
24. Xie M, Zhang X. Health-care pricing in China. Lancet 2014 Oct;384(9952):1428. [doi: 10.1016/S0140-6736(14)61863-8]
25. Wang T, Zeng R. Addressing inequalities in China's health service. Lancet 2015 Oct;386(10002):1441. [doi: 10.1016/S0140-6736(15)00402-X]
26. Hui R, Zhipeng Y, Wenmin W, Lizi G, Yanling L. Epidemiological survey on epilepsy in Jinuo ethnic minority group in Yunnan Province. J Int Neurol Neurosur 2007;34(1):10-12.

27. Dengsheng W, Guozhong L, Guangyu M. Prevalence of epilepsy in Dongning-a rural area of China. J Apoplexy Nerv Dis 2002;8(3):159-161.
28. Taiping W, Meihua H, Fenxia L, Junxiu W, Yunping S, Tong F, et al. Two epilepsy epidemiological investigation reports in a county in Shanxi Province. Chinese J Integr Med Cardio/Cerebrovasc Dis 2009;7(1):111-112.
29. Hen-zhong Y, Ling-ling Z, Jian J, Jun X, Zuo-han L, Bin G, et al. Jian-zhong epidemiological investigation on epilepsy in children in rural area of Yangzhou. Jiangsu J Prevent Med 2002;13(1):4-5.
30. Dongxiang Q, Jianzhang L, Shufen Y, Haidong L, Ligui W, Zhanguo Q, et al. Epilepsy epidemiological investigation of rural population in Henan jiaozuo wuzhi county. Chinese J Pract Nerv Dis 2001;4(6):82-83.
31. Guisheng C, Dai X, Kong F, Xu Z, Zhao X, Zan X, et al. WHO/IEM/ILAE-Epilepsy Campaign-Epidemiological Survey on epilepsy in rural areas of Ningxia autonomous region. Chinese J Neuromed 2003;2(6):421-423.
32. Wiseman H, Reuber M. New insights into psychogenic nonepileptic seizures 2011-2014. Seizure 2015 Sep;29:69-80 [FREE Full text] [doi: 10.1016/j.seizure.2015.03.008] [Medline: 26076846]
33. Park E, Lee J, Lee BL, Lee M, Lee J. Paroxysmal nonepileptic events in pediatric patients. Epilepsy Behav 2015 Jul;48:83-87. [doi: 10.1016/j.yebeh.2015.05.029] [Medline: 26074345]
34. Aaberg K, Gunnes N, Bakken I, Lund Søraas C, Berntsen A, Magnus P, et al. Incidence and prevalence of childhood epilepsy: a nationwide cohort study. Pediatrics 2017 May;139(5):1 [FREE Full text] [doi: 10.1542/peds.2016-3908] [Medline: 28557750]
35. Goodwin E, Kandler R, Alix JJ. The value of home video with ambulatory EEG: a prospective service review. Seizure 2014 Jun;23(6):480-482 [FREE Full text] [doi: 10.1016/j.seizure.2014.02.008] [Medline: 24631016]
36. Ojeda J, Ivanez V, Gutierrez-Gutierrez G, Lopez GS, Miralles A. Utility of homemade video recordings in epilepsy. Epilepsy Curr 2012;12(1):4. [doi: 10.4172/2329-6895.1000311]
37. Patterson V, Samant S, Singh M, Jain P, Agavane V, Jain Y. Diagnosis of epileptic seizures by community health workers using a mobile app: a comparison with physicians and a neurologist. Seizure 2018 Feb;55:4-8 [FREE Full text] [doi: 10.1016/j.seizure.2017.12.006] [Medline: 29291457]
38. Glynn P, Eom S, Zelko F, Koh S. Feasibility of a mobile cognitive intervention in childhood absence epilepsy. Front Hum Neurosci 2016;10:575 [FREE Full text] [doi: 10.3389/fnhum.2016.00575] [Medline: 27895568]
39. Shegog R, Begley CE. Clinic-based mobile health decision support to enhance adult epilepsy self-management: an intervention mapping approach. Front Public Health 2017;5:256 [FREE Full text] [doi: 10.3389/fpubh.2017.00256] [Medline: 29043247]
40. Liu X, Wang R, Zhou D, Hong Z. Smartphone applications for seizure care and management in children and adolescents with epilepsy: feasibility and acceptability assessment among caregivers in China. Epilepsy Res 2016 Dec;127:1-5. [doi: 10.1016/j.eplepsyres.2016.08.002] [Medline: 27522560]
41. Duffy S, Lee TH. In-person health care as option B. N Engl J Med 2018 Jan 11;378(2):104-106. [doi: 10.1056/NEJMp1710735] [Medline: 29320653]
42. Nagy E, Major A, Farkas N, Hollódy K. Epileptic seizure or not? Proportion of correct judgement based only on a video recording of a paroxysmal event. Seizure 2017 Dec;53:26-30 [FREE Full text] [doi: 10.1016/j.seizure.2017.08.017] [Medline: 29101825]
43. Orenstein SR. An overview of reflux-associated disorders in infants: apnea, laryngospasm, and aspiration. Am J Med 2001 Dec 03;111 Suppl 8A:60S-63S. [doi: 10.1016/s0002-9343(01)00823-3] [Medline: 11749927]
44. Ojeda J, Ivanez V, Lopez S, Miralles A. Usefulness of homemade videos in epilepsy diagnosis. Ann Neurol 2010;68:S42.

Abbreviations

EEG: electroencephalography
VEEG: video electroencephalography

Edited by G Eysenbach; submitted 05.06.18; peer-reviewed by B Chaudhry, T Cruvinel; comments to author 06.10.18; revised version received 01.01.19; accepted 12.06.19; published 12.09.19.

Please cite as:
Huang LL, Wang YY, Liu LY, Tang HP, Zhang MN, Ma SF, Zou LP
Home Videos as a Cost-Effective Tool for the Diagnosis of Paroxysmal Events in Infants: Prospective Study
JMIR Mhealth Uhealth 2019;7(9):e11229
URL: https://mhealth.jmir.org/2019/9/e11229/
doi:10.2196/11229
PMID:31516128

Original Paper

Biofeedback-Assisted Resilience Training for Traumatic and Operational Stress: Preliminary Analysis of a Self-Delivered Digital Health Methodology

Paul N Kizakevich[1], MS, PE; Randall P Eckhoff[2], BS; Gregory F Lewis[3], PhD; Maria I Davila[4], PhD; Laurel L Hourani[2], PhD; Rebecca Watkins[2], BS; Belinda Weimer[2], MA; Tracy Wills[2], MS; Jessica K Morgan[2], PhD; Tim Morgan[2], BA; Sreelatha Meleth[2], PhD; Amanda Lewis[2]; Michelle C Krzyzanowski[2], PhD; Derek Ramirez[2], MS; Matthew Boyce[2], BS; Stephen D Litavecz[2], MBA; Marian E Lane[2], PhD; Laura B Strange[5], PhD, RN

[1]Bioinformatice Program, Research Computing Division, RTI International, Research Triangle Park, NC, United States

[2]RTI International, Research Triangle Park, NC, United States

[3]Kinsey Institute, Indiana Unniversity, Bloomington, IN, United States

[4]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Consultant, Hilton Head, SC, United States

Corresponding Author:
Paul N Kizakevich, MS, PE
Bioinformatice Program
Research Computing Division
RTI International
3040 Cornwallis Road
Research Triangle Park, NC, 27709
United States
Phone: 1 919 541 6639
Email: kiz@rti.org

Abstract

Background: Psychological resilience is critical to minimize the health effects of traumatic events. Trauma may induce a chronic state of hyperarousal, resulting in problems such as anxiety, insomnia, or posttraumatic stress disorder. Mind-body practices, such as relaxation breathing and mindfulness meditation, help to reduce arousal and may reduce the likelihood of such psychological distress. To better understand resilience-building practices, we are conducting the Biofeedback-Assisted Resilience Training (BART) study to evaluate whether the practice of slow, paced breathing with or without heart rate variability biofeedback can be effectively learned via a smartphone app to enhance psychological resilience.

Objective: Our objective was to conduct a limited, interim review of user interactions and study data on use of the BART resilience training app and demonstrate analyses of real-time sensor-streaming data.

Methods: We developed the BART app to provide paced breathing resilience training, with or without heart rate variability biofeedback, via a self-managed 6-week protocol. The app receives streaming data from a Bluetooth-linked heart rate sensor and displays heart rate variability biofeedback to indicate movement between calmer and stressful states. To evaluate the app, a population of military personnel, veterans, and civilian first responders used the app for 6 weeks of resilience training. We analyzed app usage and heart rate variability measures during rest, cognitive stress, and paced breathing. Currently released for the BART research study, the BART app is being used to collect self-reported survey and heart rate sensor data for comparative evaluation of paced breathing relaxation training with and without heart rate variability biofeedback.

Results: To date, we have analyzed the results of 328 participants who began using the BART app for 6 weeks of stress relaxation training via a self-managed protocol. Of these, 207 (63.1%) followed the app-directed procedures and completed the training regimen. Our review of adherence to protocol and app-calculated heart rate variability measures indicated that the BART app acquired high-quality data for evaluating self-managed stress relaxation training programs.

Conclusions: The BART app acquired high-quality data for studying changes in psychophysiological stress according to mind-body activity states, including conditions of rest, cognitive stress, and slow, paced breathing.

(JMIR Mhealth Uhealth 2019;7(9):e12590) doi:10.2196/12590

KEYWORDS

resilience, psychological; heart rate variability; Personal Health Informatics and Intervention Toolkit; PHIT; respiratory sinus arrhythmia; stress, psychological; relaxation therapy; biofeedback, psychology; well-being; mindfulness; digital health; mhealth

Introduction

Background

Psychological resilience—the ability to recover from a traumatic experience and return to mental well-being—is critical to minimize health effects, such as anxiety, substance abuse, sleep problems, or posttraumatic stress disorder (PTSD) [1-6]. Exposure to trauma may leave the autonomic system in a chronic state of hyperarousal [7]. Heart rate variability (HRV), a measure of beat-to-beat cardiac interval variation, reflects vagal parasympathetic tone and changes in autonomic status [8]. Studies have found an association between PTSD and reduced HRV thought to be related to sustained hyperarousal and anxiety [9-14]. Conversely, higher HRV indicates greater flexibility and ability to regulate emotional responses, linking stress response to both enhanced mental health and resilience.

Reduction of arousal during or shortly after trauma exposure may prevent or reduce the likelihood of psychological distress, including PTSD symptoms [15-17]. Mindfulness meditation and relaxation training have been associated with a reduction in hyperarousal [17], may increase HRV [18,19], and hold promise for PTSD treatment [18,20]. HRV biofeedback, providing real-time HRV monitoring during relaxation training, has been shown to improve depression, anxiety, PTSD, and stress symptoms [21]. When practiced consistently, HRV biofeedback can also increase HRV and may help alleviate PTSD symptoms [22,23]; however, others have reported mixed results [24,25], indicating the need for further research.

Objective

The Biofeedback-Assisted Resilience Training (BART) study is evaluating whether routine practice of slow, paced breathing with and without HRV biofeedback can enhance psychological resilience by facilitating an HRV rebound after a stressor task. To support the study, we developed the BART mobile app, enabling participants to practice relaxation training outside of a formal training environment. This paper describes the BART resilience training app, demonstrates HRV biofeedback, presents processes that may have wider applicability for mobile health research, and reports the interim results of app usage.

Methods

Study Population

The BART study is being conducted in a mixed population of military personnel, veterans, and civilian first responders at multiple sites across multiple states in the United States. We recruited participants from a convenience sample of Navy, Marine Corps, and Army Reserve units and National Guard armories from North Carolina, Georgia, and Virginia, and fire and police units in the Raleigh-Durham, North Carolina, area who volunteered to participate for a 60- to 90-minute onsite training session, practice their training at home, and complete a suite of survey assessments over the course of 1 year. Eligibility criteria included having a smartphone and knowing their password. We offered monetary incentives in addition to allowing them to keep their study-related heart rate (HR) monitor chest strap.

This study was approved by the University of North Carolina Institutional Review Board under an authorization agreement with the RTI International Committee for the Protection of Human Subjects; and the US Army Medical Research and Materiel Command, Office of Research Protections, Human Research Protection Office.

Study and App Design

Study Protocol

The BART study is comparing 4 resilience training regimens: paced breathing at 5 or 6 breaths per minute, each with or without HRV biofeedback. Participants are randomly assigned to 1 of these 4 regimens and asked to practice paced breathing at least 3 times a week for 6 weeks, and thereafter for 12 months (Table 1). Participants use a chest-belt HR sensor to acquire HRV measurements, including participants randomly allocated to no HRV biofeedback. While such participants cannot observe changes in HRV during training, continuous acquisition of HRV data in the background enables posttraining analyses of physiological responses to cognitive stress and paced breathing training.

The study begins with a setup day (day 0) on which individuals provide their consent to participate, install the app on their smartphone, complete baseline assessments, learn to use their HR monitor (Polar H7; Polar Electro, Bethpage, NY, USA), and practice 1 resilience training session, which includes a cognitive stress game. Special consideration is given to wearing the Polar HR sensor, linking the sensor to the BART app, and acquiring good-quality HR measurements. After this initial setup and instruction, participants execute all activities on their own for the duration of the study under the scheduling and direction of the BART app.

Table 1. Schedule of participant activities across the yearlong study duration.

Time	Resilience training	Resilience training with cognitive stress	Assessment	Incentive (US $)
Day 0	N/A[a]	Practice once	Baseline part 1	15
Days 0-3	N/A	N/A	Baseline part 2	5
Days 0-3	N/A	N/A	Baseline part 3	5
Weeks 1-6	Practice twice/week	Practice once/week	Weekly survey	10/week
Months 3, 6, and 9	Practice twice/week	Practice once/quarter	Quarterly survey	20/quarter
Month 12	Practice twice	Practice once	Final survey	20

[a]Not applicable.

BART app design was governed by the study protocol and schedule of participant activities (Table 1), the prescribed resilience training and stressor regimens, the mechanisms for helping participants to complete the study activities, and incentives to encourage adherence. A suite of self-report measures (eg, anxiety, posttraumatic stress, sleep quality, resilience) are taken at baseline, with a subset taken quarterly and at 12 months. Scheduling of activities is provided via the app, along with incentives to support adherence over the initial 6 weeks and throughout the 12 months. Owing to the geographic distribution of study recruitment sites, participants enter the study incrementally, thereby allowing a small study team to recruit, take consent, and provide initial training at various locations over an extended period. Consequently, each participant's protocol schedule is based on their personal study entry date.

App Development

Our previous work in predeployment stress inoculation training, HRV biofeedback, and mobile technologies for mindfulness-based stress reduction strongly influenced our design of the BART app [18,25,26]. Each of these studies involved stress relaxation training, a cognitive stressor, and HRV assessments. We reviewed our lessons learned from these studies to refine processes and incorporate new sensors and mobile technology in the BART app. Smartphone-delivered health and HRV biofeedback analyses of the prior Personal Health Intervention Toolkit for Duty research app [27] constituted the foundation for app development.

We implemented the BART app using the Personal Health Informatics and Intervention Toolkit (PHIT), a development framework geared to research-oriented mobile apps [28-30]. The PHIT framework eases app development for acquiring data, including self-report instruments, ecological momentary assessment diaries, cognitive tests, and game-like activities. For sensor data collection, PHIT supports intrinsic (eg, global positioning system, motion) and Bluetooth 4.0 data streams (eg, HR monitors). All data are tagged with study protocol, participant, date and time stamps, and other contextual information, then encrypted and stored locally in the app space. A virtual advisor provides a logic layer where analysis and planning take place. An activity manager schedules self-report and sensor data collection, intervention and training, alerts, incentive feedback, and behavior change according to the study protocol.

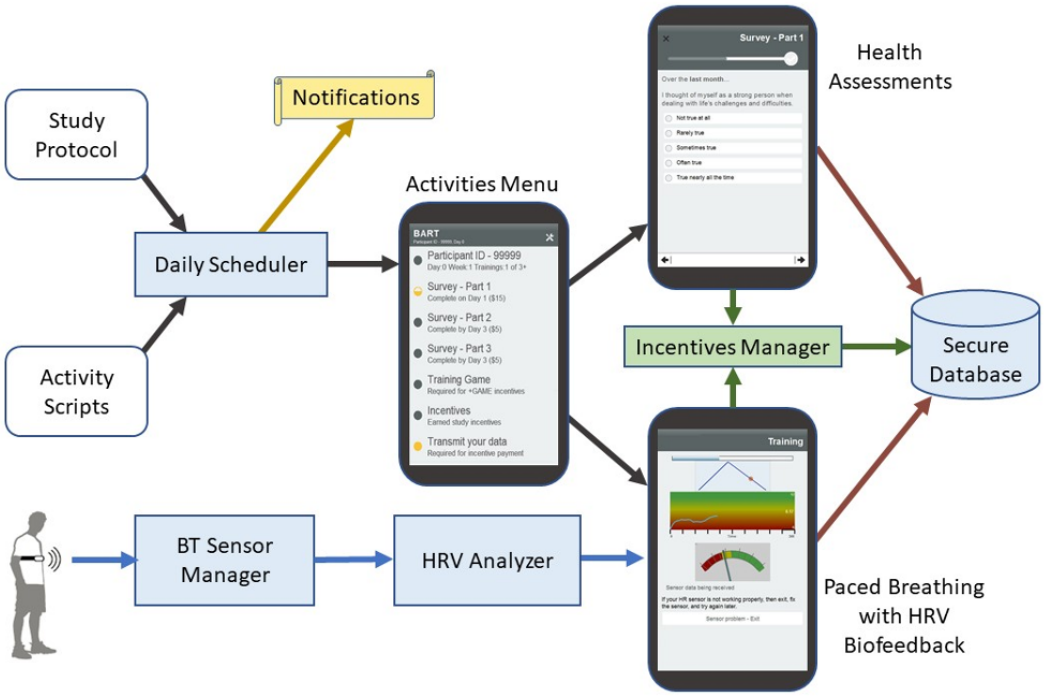
PHIT modules are implemented using XML and employ PHITScript to construct program logic and activate special app functions, such as collecting sensor data or scheduling notifications. Apps using the PHIT framework run locally without the need for an active internet connection. PHIT is based on Apache Flex (Apache Software Foundation) and AIR (Adobe Systems) technologies, which are both open source and widely used for mobile game development. A requirement for the BART study was that participants would use their own smartphones or tablets, necessitating app compatibility with both Android and iOS devices as provided by Adobe AIR.

All acquired and derived data are stored on the device in an encrypted SQLite (SQLite Consortium) database and periodically uploaded by the participant to a secure central data server. To eliminate financial burden, neither continuous internet access nor use of the participant's cellular data plan is required. Rather, data are uploaded whenever Wi-Fi internet access is available, and at the participant's direction and convenience, either via Wi-Fi or the participant's cellular data plan.

App Architecture

The BART operational schema (Figure 1) centers on a participant activities menu with various tasks such as health assessments, resilience training, and data uploads. The activities menu (Figure 1 and Figure 2a) is updated daily by a schedule manager according to protocol specifications and personal progress, using labeling and icon references as defined by PHITScript programming. Each day, activities are listed or removed, and a local notification is posted to the participant as a reminder to complete their activities. The menu may also be updated by changing the icon to note an incomplete activity, removing a completed activity from the menu, or tagging a listed activity with key user information—for example, advising on the number of resilience trainings remaining to meet incentive payment requirements for the current study week.

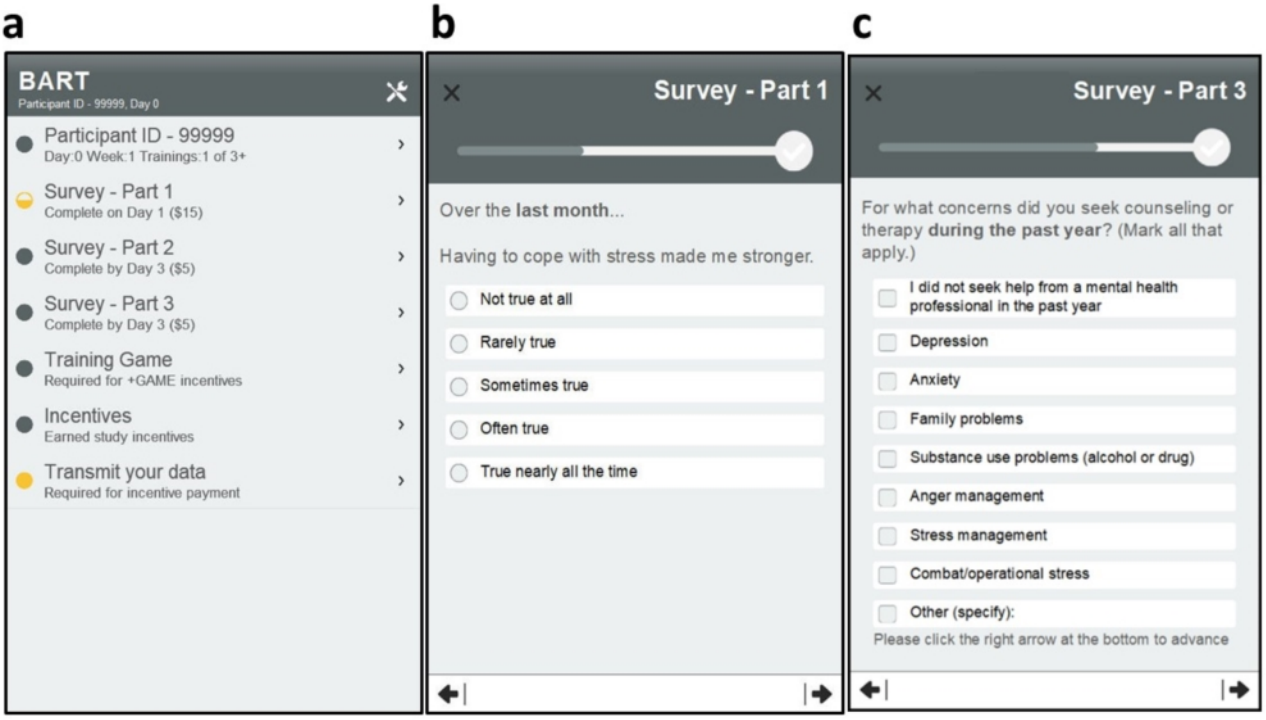
Figure 1. Overall architecture and major components of the Biofeedback-Assisted Resilience Training (BART) study mobile app. BT: Bluetooth; HRV: heart rate variability.



Primary outcome measures acquired via the app are resilience (Connor-Davidson Resilience Scale), coping measures (Brief Coping Scale, Perceived Stress Scale, and Posttraumatic Growth Inventory), and sleep problems (Sleep Disturbance Scale). Secondary outcomes are mental health (measured by the PTSD Checklist, 7-item Generalized Anxiety Disorder scale, and Center for Epidemiologic Studies Depression Scale), physical health (Short Form Health Survey), and alcohol use problems (Alcohol Use Disorders Identification Test). Covariates are combat and deployment, recent tobacco and caffeinated beverage use, age, education, use of other relaxation techniques, and interest in learning relaxation skills. These measures, along with demographic information, were aggregated into a set of brief survey instruments to be completed at baseline (surveys 1-3), weekly, quarterly, and at 12 months.

Health assessments are administered via brief self-report instruments, typically with a single question per screen (Figure 2b and c). As the user advances through each assessment, a graphic indicator informs progress toward completion. At completion of each self-report measure or resilience practice exercise, an incentives manager records the earned incentive to the database (Figure 1). The user is then advised to upload data or defer the upload to a more convenient time. The activities menu may also be updated. For activities with HR sensor data streams, a Bluetooth interface manager links the sensor and receives beat-by-beat HR information for HRV analysis. Once initiated, this process executes autonomously in the background while the participant performs resilience training. The raw HR and derived HRV measures are provided for feedback display and saved in the app database.

Figure 2. Biofeedback-Assisted Resilience Training (BART) app home screen activities menu and examples of health assessment survey questions. (a) Activities menu; (b,c) sample survey questions.



Heart Rate Variability

The BART project employs real-time HRV analysis to provide physiological biofeedback during resilience training. Beat-by-beat heart intervals, also called interbeat intervals, are acquired continuously during each training session from a Bluetooth Low Energy (Bluetooth Special Interest Group) HR monitor. The raw interbeat intervals are streamed in real time to an HRV analysis module and stored in the app database to allow for subsequent offline quality review and analysis.

Three variants of HRV measures are determined using the Porges-Bohrer HRV analysis methodology [31,32]: respiratory sinus arrhythmia (RSA), low-frequency HRV, and wideband HRV. RSA reflects parasympathetic vagal activity for expected spontaneous breathing rates, whereas low-frequency HRV is thought to reflect sympathetic activity, as well as other cardiovascular regulatory systems. The wideband measure ensures that very low breathing rates during paced breathing exercises are properly measured. Multimedia Appendix 1 provides HRV data processing details.

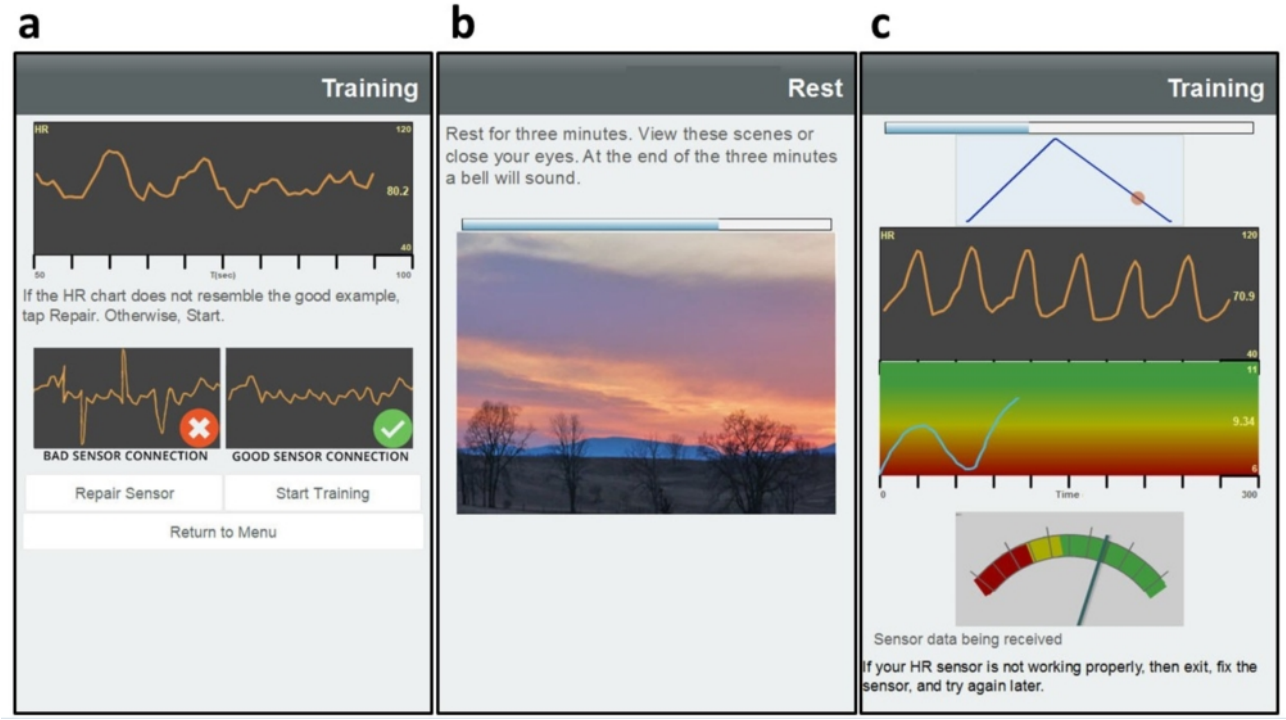
Resilience Training

We asked participants to practice resilience training 3 times each week for 6 weeks, a 2-step process comprising a 3-minute resting segment and a 5-minute resilience training segment. Each participant was randomly allocated after consent to receive 1 of 4 resilience training regimens: paced breathing at 5 or 6 breaths per minute, with or without HRV biofeedback. Before training, the participant is asked to be in a quiet place and put on the Polar H7 HR monitor. When the participant is ready, the HR monitor is activated and a beat-by-beat HR trend is displayed to check signal quality (Figure 3a). The participant reviews the HR trend and decides whether to proceed to resilience training or take measures, such as adjusting or moistening the chest strap sensor, to improve data quality. Resilience training begins with a 3-minute resting segment to relax the participant and establish baseline HRV measures. During this time, the participants may close their eyes or lightly focus on a series of peaceful landscapes that fade from one to another at 30 second intervals (Figure 3b). A narrator announces when each minute arrives to help reduce anxiety owing to waiting for the resting segment to finish.

For participants receiving resilience training *without biofeedback*, an animated ball is displayed as rising and falling upon a triangular graphic for paced breathing resilience training (Figure 3c top). Participants inhale as the ball rises and exhale while the ball falls, with ball movement set at 5 or 6 breaths per minute, with an inspiration to expiration ratio of 0.435 and an end-inspiration and end-expiration pause of 1.5 seconds. An audible tone with rising and falling pitch is played in synchrony with the rising and falling ball to allow for paced breathing with eyes closed.

For participants receiving resilience training *with biofeedback*, the animated ball and audible tones are rendered in similar fashion to that without biofeedback. Two modes of graphic biofeedback are provided: a trending HRV chart and a real-time dynamic HRV meter. The chart and meter are updated every 2 seconds against a color-coded background to show movement between calm (green) and stressful (red) psychophysiological states, reflecting higher and lower parasympathetic activation.

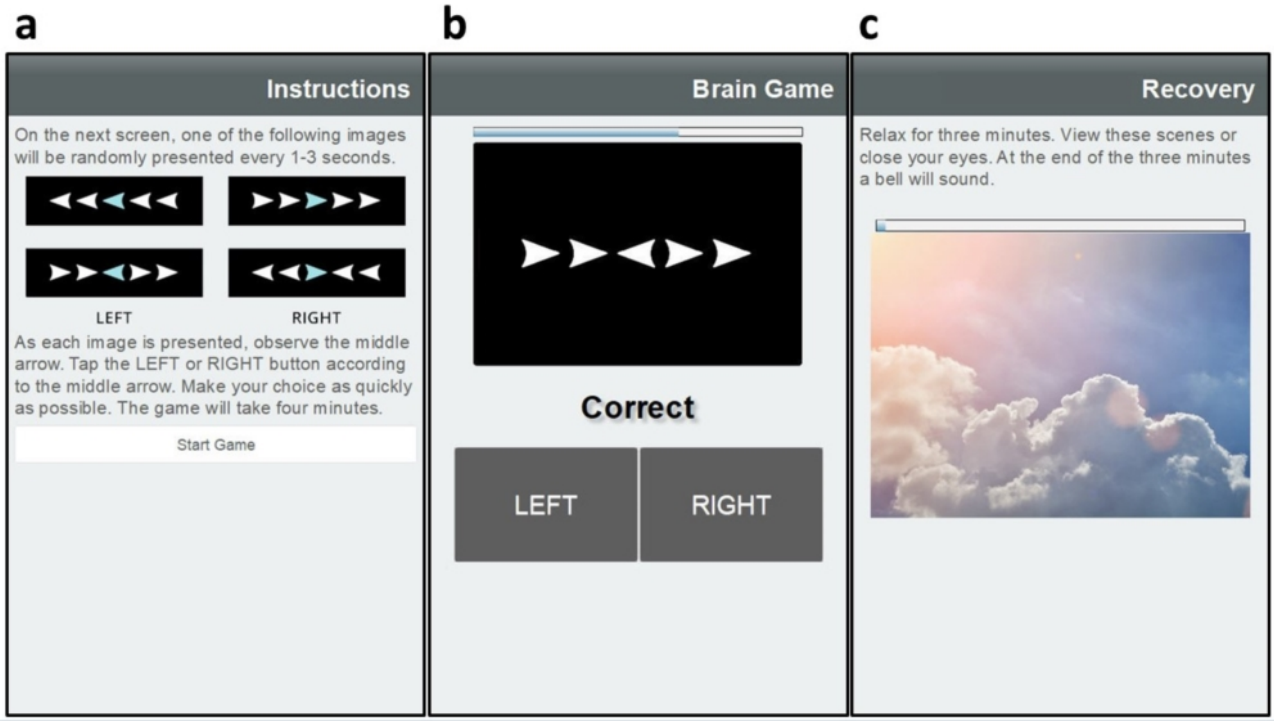
Figure 3. (a) Resilience training begins with verification of heart rate (HR) data quality to ensure high-quality heart rate variability (HRV) biofeedback. (b) Once the HR is checked, participants sit at rest for a 3-minute baseline period. (c) Then participants receive paced breathing via an audiovisual graphic animation (top), while the HR signal, HRV trend, and instantaneous HRV meter are displayed in real time (bottom).



On study days 0 and 1, and after 6 weeks of training, participants complete an enhanced training regimen called the Training Game. The Training Game builds on the basic resilience training exercise by incorporating the Eriksen flanker task [33], a game-like cognitive stress exercise designed to elicit psychophysiological stress. The Eriksen flanker task heightens psychological stress by requiring attention, providing anticipation, and imposing conflict in higher brain function. As before, HRV is measured throughout the Training Game, thereby allowing for objective assessment of resilience before and after 6 weeks of training.

The Training Game begins with a 3-minute rest, followed by instructions on performing the Eriksen flanker task (Figure 4a). When ready, the participant begins the Eriksen flanker task, which presents a series of stimulus screens comprising a field of arrows pointing to the left or right, with a central arrow that may be congruent or incongruent in direction with the 4 bounding arrows on either side (Figure 4b). The 4 bounding arrows are randomly rendered as pointing left or right, resulting in 4 available stimulus combinations. At a random interval ranging from 1 to 3 seconds, 1 of the 4 combinations of the central and flanking arrows is selected at random and presented for 400 milliseconds. Participants have 2.7 seconds to respond by tapping the left or right button to indicate the direction of the central arrow. The Eriksen flanker task continues presenting stimuli for 4 minutes, then the BART app advances to a 3-minute poststress recovery phase of sitting at rest (Figure 4c). After recovery, resilience training is provided as previously described with paced breathing at 5 or 6 breaths per minute, with or without HRV biofeedback.

Figure 4. The cognitive Training Game stressor exercise is preceded by (a) an instruction screen, followed by (b) 4 minutes of Eriksen flanker stimuli with user response, and completed with (c) 3 minutes of poststress recovery.



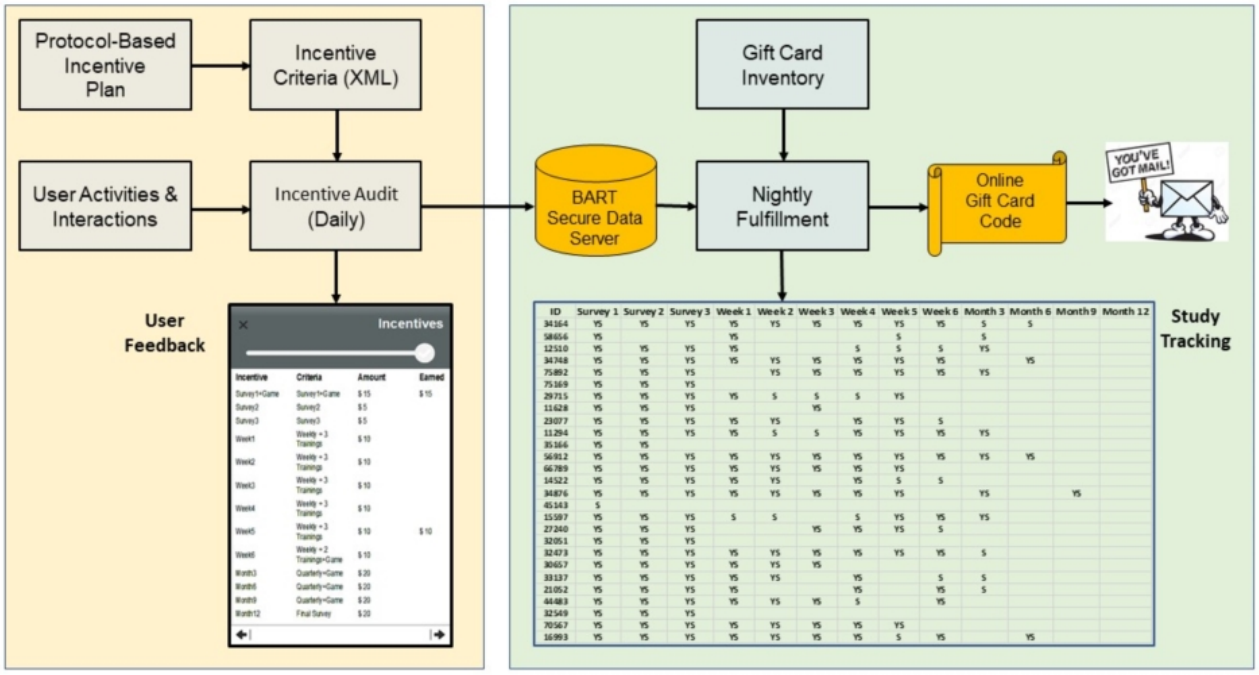
Protocol Adherence and Incentive Management

Adherence to study interventions and data collection is a common challenge in any research study, and especially when participants are asked to essentially perform the study on their own, albeit with app support. To support the frequency of resilience training (3 time per week), weekly assessments, and the yearlong period of follow-up assessments, the BART study design incorporated an incremental incentives approach to encourage participants to do scheduled activities and stay engaged across the duration. We were concerned, however, that manual monitoring of several hundred participants with time-shifted protocol schedules could be error prone and cause missed payments or awarding unearned payments. To mitigate such risk, we implemented incentive management to standardize both incentive qualification and automated distribution of incentive awards (Figure 5).

In concert with the BART protocol (Table 1), incentives are earned providing that the participant completes the requisite activities (Figure 5). Incentive criteria, along with labeling and monetary values, are coded in PHITScript and checked both daily and after the participant exits each activity module. Whenever the participant completes the criteria for a specific incentive, an incentive fulfillment request is stored to the local app database, and the in-app incentives table is updated to inform the user that an incentive has been earned and is pending award. When study data are transferred to the secure central data server, the incentive fulfillment request is noted for processing. Each night a procedure scans all participants for pending fulfillment requests, identifies unpaid incentives, emails a gift card code in the appropriate amount to participants, and tags the incentive payment as fulfilled.

To further support the incentive process, monitor payments, and validate the accuracy of the incentive management process, a report of incentives earned and paid is produced weekly so that study staff may check each participant's incentive record. This report not only aids in confirming payments, but also helps to resolve any problems that may have been experienced by participants and supports monitoring adherence across all participants.

Figure 5. Incentive management data flow and processes conducted within the Biofeedback-Assisted Resilience Training (BART) app (left) and in the secure backend data server (right) for monitoring and rewarding participant adherence.



Privacy and Security

Ensuring privacy and security of data and on-device analysis results is an absolute necessity for ethical reasons, to meet human studies requirements, and to support data quality in the conduct of self-managed mobile research protocols. Participants are provided a randomized participant identification (ID), which uniquely links all acquired data to that individual without any personally identifiable information. They also enter a self-defined secret 4-digit personal identification number (PIN) to prevent access by other individuals. When in use, the app screen deactivates after a set period (eg, 2 minutes) of no interaction, and current data and activity is hidden. The 4-digit PIN must then be entered to unlock the screen and allow the participant to continue.

Implementing a study using an app installed on the participants' devices requires app installation from a public app store. Since anyone might download and install the app, and possibly upload false data, we addressed this quality and security risk by including an app lockout requiring an unlock password. After participants consent to take part in the study, the unlock is revealed, the app installed, and the password entered to activate. This prevents extraneous persons from entering and corrupting the study.

All data are stored locally on the device in an encrypted SQLite database within the BART app, thereby permitting use without requiring a continuous internet connection. Data are stored using a 128-bit Advanced Encryption Standard algorithm with no personal identifying information. Data are periodically uploaded to a central secure data server whenever Wi-Fi internet access is available, thereby reducing use of the participants' cellular data plans. Data are transferred using the secure https protocol and stored in a secure SQL server database, which is accessible only to authorized persons via user ID and password authentication.

Data Analysis

To showcase how the BART app is being used and to present examples of the HRV measures during resilience training, we conducted a limited, interim review of user interactions and study data. Consequently, data presented here do not address the study hypotheses on the effectiveness of various training modes on building resilience. Analysis of training effectiveness on resilience and other outcome measurements will be addressed separately after completion of data collection.

We based data regarding app usage on participant rostering records and earned incentives reported. For each study activity (Table 1), we tallied a completed measure—that is, the number of participants who completed the activity and earned the corresponding incentive. Since each participant has a unique study calendar based on individual starting date, the activity schedule differs across participants. We therefore tallied the number of participants who were scheduled to perform each activity adjusted according to their individual start date (ranging from June 2017 to September 2018) until this analysis on October 1, 2018. Finally, we determined the ratio of completed to scheduled activities as a compliance measure for each required study activity. We calculated these analyses using Excel 2019 (Microsoft Corporation).

We reviewed the psychophysiological stress response during the cognitive stress and biofeedback training using the wideband biofeedback HRV measure across all segments (rest, stressor, recovery, and training). Our analysis was restricted to data taken during the first week of participation, before substantive resilience practice would yield any training effect.

We computed descriptive statistics for the subpopulation extracted for the HRV measurement examples, with categorical variables reported by frequencies and numeric variables by mean (SE). We analyzed grouped HRV data using a univariate general linear model and present the data graphically as mean (2 SE). We used unpaired *t* tests to evaluate changes in HRV for sequential training segments. We conducted statistical analyses using IBM SPSS Statistics 25 (IBM Corporation).

Results

Resilience Training and Protocol Compliance

Of the 328 enrolled participants to date, 207 (63.1%) adhered to the study training regimen of 3 resilience training sessions per week for at least six weeks. In total, 3136 training episodes had been performed in this subset across the first 180 days of each participant's involvement (studyDay; Figure 6). At first, compliance with the training regimen was excellent, with over 600 sessions conducted during the first week by the 207 participants who completed the 6-week training regimen. However, over the next several weeks, training compliance fell by almost one-third, and later to about one-half after a month. Following the 6-week training period, compliance was diminished far below 3 trainings per week. However, a small number of participants continued resilience training for at least six months, with several continuing for nearly a year (not shown on the plot in Figure 6).

We also examined adherence to completing self-report health and wellness surveys at baseline, weekly for 6 weeks, and quarterly for up to 12 months according to the study protocol (Table 1). Among the 328 enrolled participants, compliance with completing scheduled surveys decreased across the study duration (Figure 7). Although participants were asked to complete surveys 1 to 3 immediately after installing the app, many were short on time and indicated that they would do them later that day. However, 11.0% (36/328) did not even complete the initial survey. Each week, and quarter, as each data collection survey was scheduled, participants were reminded to complete the pending survey and receive their incentive via a smartphone notification. A total of 1760 incentives were earned and automatically awarded from June 2017 through September 2018. Despite this support by the BART app, completion rates fell to 50.0% (164/328) by week 2, then to 22.9% (75/328) at 3 months and 10.1% (33/328) at 6 months.

Figure 6. Total number of resilience training sessions across participants by study day.

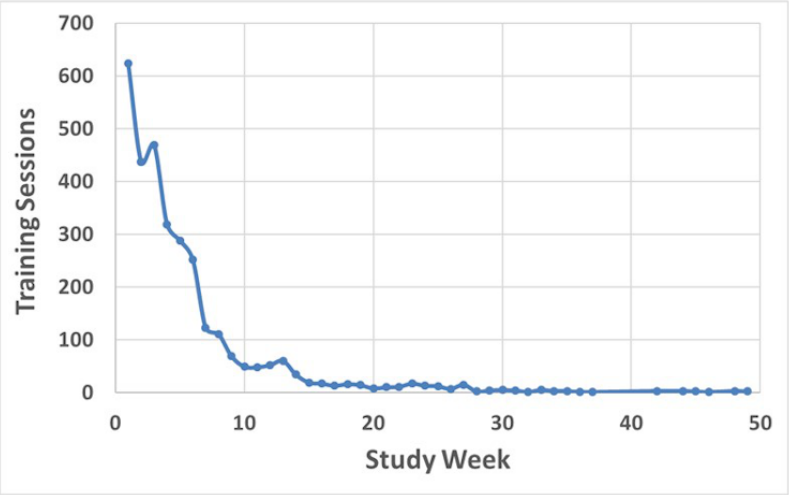
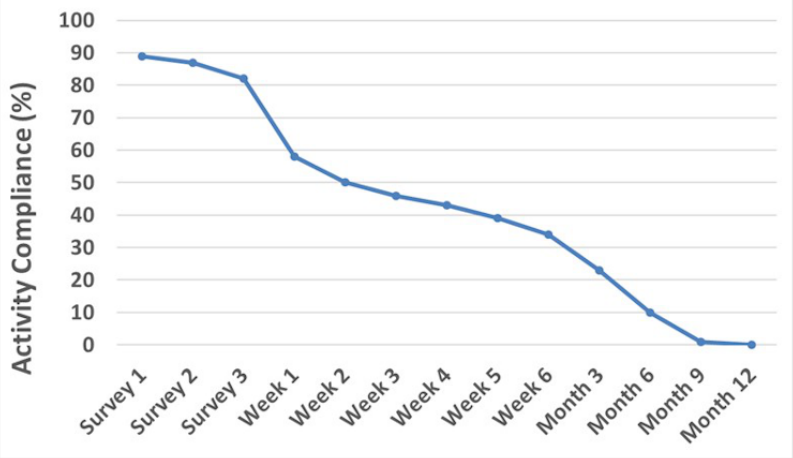


Figure 7. Compliance with scheduled study activities (listed in Table 1).

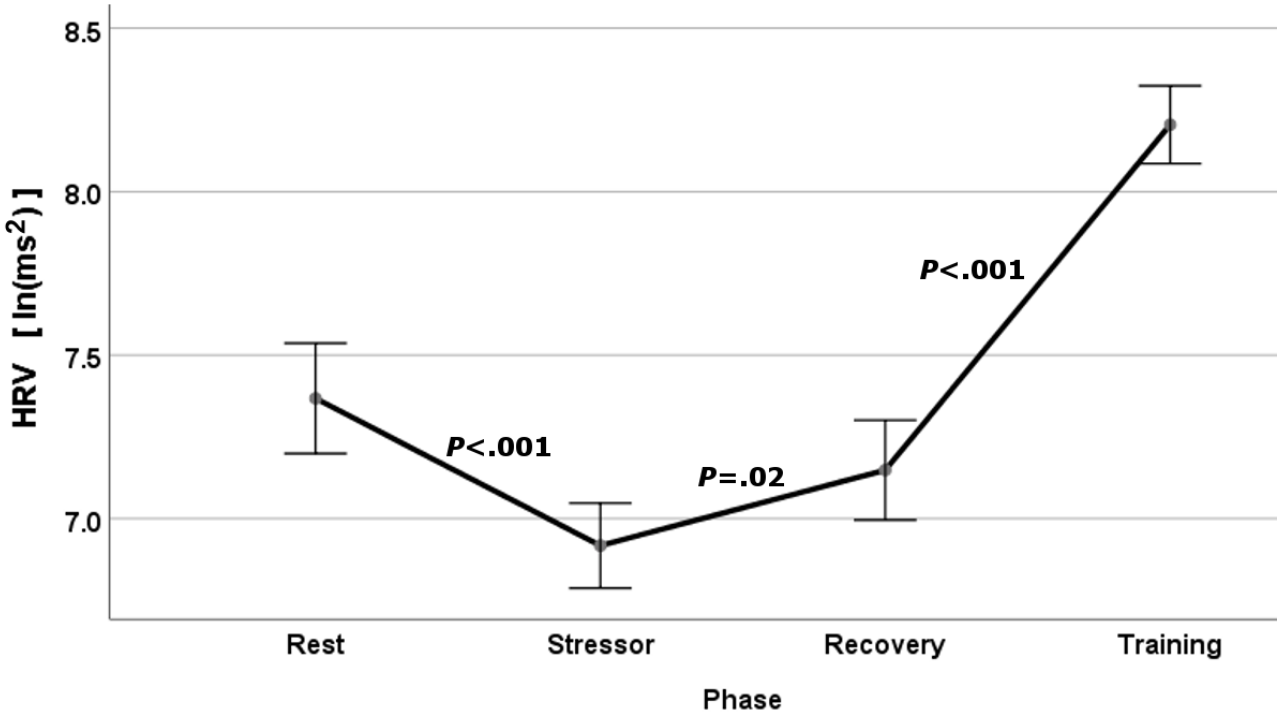


Heart Rate Variability

We included a subset of the dataset, comprising 49 men and women, aged 20 to 60 years (mean 36.7, SE 10.6), in the HRV review. Of the 49 participants in the subset, 23 (47%) were female and 26 (53%) were male. We excluded HRV values that were out of the expected range (HRV<0 or HRV>10) as outliers, as such data are likely due to interbeat interval artifacts. We included multiple HRV measures per individual segment, ranging from 305 to 613 measures.

We present grouped results for HRV, using the wideband biofeedback HRV measure, for each segment of the cognitive Training Game stressor exercise (Figure 8). As expected, HRV decreased from mean 7.37 (SE 1.77; n=305 measures) at rest to mean 6.92 (SE 1.41; n=515 measures) during the stressor phase, reflecting a reduction in parasympathetic activation during the Eriksen flanker stressor task (P<.001). Later during the poststress recovery segment, the HRV rebounded to mean 7.148 (SE 1.42; n=373 measures), approaching the prestress baseline (P=.02). During training, HRV increased very significantly to mean 8.205 (SE 1.39; n=613 measures), reflecting strong parasympathetic activation with slow paced breathing (P<.001).

Figure 8. Heart rate variability (HRV) at rest and during cognitive stress, recovery, and paced breathing. Error bars represent ±2 SE.



Discussion

Principal Findings

A large variety of mobile apps for stress reduction [27,34,35], mindfulness training [36,37], biofeedback [38], and HRV measurement [39] have emerged over the last decade, both for general use and as adjuncts to specific disease interventions [38,40]. Most merely provide narrative training support and practice reminders, with little evidence of efficacy [36]. Furthermore, most app implementations do not have concomitant self-report or physiological data gathering as is necessary for evaluating efficacy. As the BART app is built on the PHIT mobile health research platform [28], we are able to acquire research data throughout the 6-week training regimen, tag data according to training segments (ie, rest, stress, paced breathing), and acquire physiological HRV measurements to support hypothesis testing in our primary BART evaluation study. We expect, therefore, that using our self-delivered digital health methodology will improve the understanding of the efficacy and utility of mobile, self-directed mind-body interventions.

Physiological biofeedback during paced breathing resilience training and objective assessment of psychological arousal would not be possible without real-time continuous monitoring of HRV by the BART app. The HRV results during the Training Game exercise for the rest, cognitive stressor, recovery, and training segments (Figure 8) are consistent with previous results found in our predeployment stress inoculation studies [25,26], where we observed a significant decrease in HRV (RSA) during cognitive stress and a significant increase during relaxation breathing. A similar reduction in HRV with cognitive stress has been reported by other investigators during mental arithmetic [41] and random number generation [42] tasks. By assessing the vagal-mediated RSA throughout each resilience training episode, we can readily observe changes in arousal due to different psychophysical states (eg, rest, stress). Therefore, any potential improvement in base arousal or resilience to (cognitive) stress after the 6 weeks of resilience training should be readily demonstrated.

A limitation in this study is the use of cognitive stress as a surrogate for combat and operational stressors in this military population. Risk of death, exposure to combat or casualties, disconnection from loved ones, and working in extreme and

unusual environments are examples of trauma that our study population might experience. Such stressors cannot be readily mimicked, nor should they, as previously exposed participants could experience negative reactions to simulated exposures. Use of a controlled cognitive exercise provides an alternative, safe, and common context to assess stress reactivity for evaluation of relaxation training.

Along with the stated benefits, the BART study has yielded a variety of lessons for such self-directed app-based research. Using personal mobile apps not only to collect information, but also to manage protocol-based task scheduling, reminder notifications, and intervention activities, makes the study essentially self-administered by each participant. Unforeseen events, such as participant smartphone replacement, forgetting 4-digit security PINs, and assorted HR monitor failures, necessitated the implementation of technical support resources. We did this via website, telephone, and email interactions, with issue and resolution tracking using Jira Software v7.11.2 (Atlassian). Maintenance of personal interest, usability of sensors and devices, adherence to procedures, and timely technical support are critical in retaining participation for the study duration.

The nature of our study population (primarily military reserve units) imposed a requirement to recruit participants, then immediately install the BART app and provide initial training in a group setting, often with more than 20 individuals present. These large groups compromised our process to establish Bluetooth links between individual participant's HR monitor and smartphone in a multiparticipant environment. We addressed this by having participants configure their app in small subgroups, which eased the installation and setup process considerably.

Furthermore, while the selected HR monitors work quite well with exercise, obtaining a good HR signal was often difficult while the participant is sitting at rest (ie, not sweating). Multiple adjustments of the sensor strap and repeated configuration attempts were often required, and we suppose that continued problems of this sort likely contributed to participant dropout. Advances in wearable HR sensor technology, such as upper arm photoplethysmography, may make them easier to use and more reliable under resting conditions than a device designed for exercise. As such devices emerge with enough accuracy for HRV measurement, we expect to improve protocol adherence and reduce participant dropout due to problems experienced with the HR sensor.

Incorporation and automation of incentive management is a vital aspect of the BART app. We are using monetary incentives, an important component of research projects, to support adherence to study procedures and reward participants for carrying out certain tasks, such as completing a survey or resilience training. Since participants have individual start dates, their individual calendar of study activities will differ across the study population, making manual monitoring of study adherence both time consuming and error prone. By embedding adherence management, we can check protocol activities frequently, then reward participants immediately using automatic, incremental incentive payments. Weekly reports to study staff on incentive payments yielded useful feedback on protocol adherence and the potential for intervention by study staff to help keep individual participants on track.

Retention of users is a common issue with mobile apps in general. Bonnie [43] reported that 90% of users stopped using apps within 30 days and 95% by 90 days after installation. In contrast, the BART app methodology of supporting participants with incentives and usage feedback allowed the study to retain over 20% of enrolled participants after 60 days and roughly 10% at 90 days.

While a key component of ensuring optimal study participation, automation of incentive management was not without issues. Initially we had a somewhat complex set of requirements for participant incentives, including requiring their resilience trainings plus completion of the weekly survey in each of the first 4 weeks to receive the incentive payment. Furthermore, participants were given 4 days to compete the survey, and then it was removed from the activities menu. Despite instructions via the embedded incentive requirements table, several participants complained that they did not receive incentives. Upon review, we found that they did not fully meet the requirements but, as they met most, we decided to award the incentives anyway. We then relaxed the requirements, while still asking that these tasks be completed (or at least initiated), so that such persons would not drop out of the study. Nonetheless, having the automated incentive checking and database recording was helpful to review these cases and to consider the participant's actions and understanding how to better incentivize study activities.

Conclusion

Results presented is this paper merely showcase features and capabilities of the BART app, along with preliminary data on app usage and demonstration of analyses of real-time sensor-streaming data, such as the psychophysiological HRV response to cognitive stress and paced breathing training.

Currently distributed for the BART research study, the BART app is being used to collect self-reported survey and HR sensor data for comparative evaluation of paced breathing relaxation training with and without HRV biofeedback. Our preliminary ad hoc analyses indicate that the app acquires high-quality data for studying changes in psychophysiological stress according to mind-body activity states, including relaxation and cognitive stress conditions. However, no conclusion of effectiveness, or noneffectiveness, of the biofeedback-assisted relaxation training intervention should be drawn from these data.

Acknowledgments

We thank the Navy, Marine Corps, and Army Reserves, and the North Carolina Air National Guard for their participation in the BART Study.

This research was supported by the US Army Medical Research and Materiel Command (W81XWH-16-1-0346, W81XWH-16-1-0347, W81XWH-11-2-0129). The conclusions do not necessarily reflect the position or policy of the US Government, and no endorsement should be inferred.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Heart rate variability methodology and validation.

[PDF File (Adobe PDF File), 209KB - mhealth_v7i9e12590_app1.pdf]

References

1. Mustillo SA, Kysar-Moon A, Douglas SR, Hargraves R, Wadsworth SM, Fraine M, et al. Overview of depression, post-traumatic stress disorder, and alcohol misuse among active duty service members returning from Iraq and Afghanistan, self-report and diagnosis. Mil Med 2015 Apr;180(4):419-427. [doi: 10.7205/MILMED-D-14-00335] [Medline: 25826347]
2. Capaldi VF, Guerrero ML, Killgore WDS. Sleep disruptions among returning combat veterans from Iraq and Afghanistan. Mil Med 2011 Aug;176(8):879-888. [Medline: 21882777]
3. McLay RN, Klam WP, Volkert SL. Insomnia is the most commonly reported symptom and predicts other symptoms of post-traumatic stress disorder in U.S. service members returning from military deployments. Mil Med 2010 Oct;175(10):759-762. [Medline: 20968266]
4. Ramchand R, Miles J, Schell T, Jaycox L, Marshall GN, Tanielian T. Prevalence and correlates of drinking behaviors of previously deployed military personnel and matched civilian population. Mil Psychol 2011 Jan 01;23(1):6-21 [FREE Full text] [doi: 10.1080/08995605.2011.534407] [Medline: 25324594]
5. Clinton-Sherrod AM, Barrick K, Gibbs DA. Soldier characteristics, alcohol abuse risk, and mental health risk as treatment predictors. Mil Psychol 2011 Jan 13;23(1):22-35 [FREE Full text] [doi: 10.1080/08995605.2011.534408]
6. Brown JM, Williams J, Bray RM, Hourani L. Postdeployment alcohol use, aggression, and post-traumatic stress disorder. Mil Med 2012 Oct;177(10):1184-1190. [Medline: 23113445]
7. Porges SW. The Polyvagal Theory. Neurophysiological Foundations of Emotions, Attachment, Communication, and Self-regulation (Norton Series on Interpersonal Neurobiology). New York, NY: WW Norton & Company; 2011.
8. Porges SW. The polyvagal perspective. Biol Psychol 2007 Feb;74(2):116-143 [FREE Full text] [doi: 10.1016/j.biopsycho.2006.06.009] [Medline: 17049418]
9. Zucker TL, Samuelson KW, Muench F, Greenberg MA, Gevirtz RN. The effects of respiratory sinus arrhythmia biofeedback on heart rate variability and posttraumatic stress disorder symptoms: a pilot study. Appl Psychophysiol Biofeedback 2009 Jun;34(2):135-143. [doi: 10.1007/s10484-009-9085-2] [Medline: 19396540]
10. Hauschildt M, Peters MJV, Moritz S, Jelinek L. Heart rate variability in response to affective scenes in posttraumatic stress disorder. Biol Psychol 2011 Dec;88(2-3):215-222. [doi: 10.1016/j.biopsycho.2011.08.004] [Medline: 21856373]
11. Thayer JF, Ahs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012 Feb;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
12. Nagpal ML, Gleichauf K, Ginsberg JP. Meta-analysis of heart rate variability as a psychophysiological indicator of posttraumatic stress disorder. J Trauma Treat 2013;3(182). [doi: 10.4172/2167-1222.1000182]
13. Minassian A, Geyer MA, Baker DG, Nievergelt CM, O'Connor DT, Risbrough VB. Heart rate variability characteristics in a large group of active-duty marines and relationship to posttraumatic stress. Psychosom Med 2014;76(4):292-301. [doi: 10.1097/psy.0000000000000056]
14. Shah A, Vaccarino V. Heart rate variability in the prediction of risk for posttraumatic stress disorder. JAMA Psychiatry 2015 Oct;72(10):964-965 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1394] [Medline: 26352544]
15. Cohen H, Benjamin J, Geva AB, Matar MA, Kaplan Z, Kotler M. Autonomic dysregulation in panic disorder and in post-traumatic stress disorder: application of power spectrum analysis of heart rate variability at rest and in response to recollection of trauma or panic attacks. Psychiatry Res 2000 Sep 25;96(1):1-13. [Medline: 10980322]
16. Schelling G, Briegel J, Roozendaal B, Stoll C, Rothenhäusler HB, Kapfhammer HP. The effect of stress doses of hydrocortisone during septic shock on posttraumatic stress disorder in survivors. Biol Psychiatry 2001 Dec 15;50(12):978-985. [Medline: 11750894]
17. Creswell JD, Pacilio LE, Lindsay EK, Brown KW. Brief mindfulness meditation training alters psychological and neuroendocrine responses to social evaluative stress. Psychoneuroendocrinology 2014 Jun;44:1-12. [doi: 10.1016/j.psyneuen.2014.02.007] [Medline: 24767614]

18. Hourani L, Tueller S, Kizakevich P, Lewis G, Strange L, Weimer B, et al. Toward preventing post-traumatic stress disorder: development and testing of a pilot predeployment stress inoculation training program. Mil Med 2016 Dec;181(9):1151-1160. [doi: 10.7205/MILMED-D-15-00192] [Medline: 27612367]
19. Krygier JR, Heathers JAJ, Shahrestani S, Abbott M, Gross JJ, Kemp AH. Mindfulness meditation, well-being, and heart rate variability: a preliminary investigation into the impact of intensive Vipassana meditation. Int J Psychophysiol 2013 Sep;89(3):305-313. [doi: 10.1016/j.ijpsycho.2013.06.017] [Medline: 23797150]
20. Watson CG, Tuorila JR, Vickers KS, Gearhart LP, Mendez CM. The efficacies of three relaxation regimens in the treatment of PTSD in Vietnam War veterans. J Clin Psychol 1997 Dec;53(8):917-923. [Medline: 9403395]
21. Gevirtz R. The promise of heart rate variability biofeedback: evidence-based applications. Biofeedback 2013 Sep;41(3):110-120 [FREE Full text] [doi: 10.5298/1081-5937-41.3.01]
22. Reyes FJ. Implementing heart rate variability biofeedback groups for Veterans with posttraumatic stress disorder. Biofeedback 2014 Dec;42(4):137-142 [FREE Full text] [doi: 10.5298/1081-5937-42.4.02]
23. Tan G, Wang P, Ginsberg J. Heart rate variability and posttraumatic stress disorder (PTSD). Biofeedback 2013 Sep;41(3):131-135 [FREE Full text] [doi: 10.5298/1081-5937-41.3.05]
24. Schoenberg PLA, 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]
25. Wahbeh H, Oken BS. Peak high-frequency HRV and peak alpha frequency higher in PTSD. Appl Psychophysiol Biofeedback 2013 Mar;38(1):57-69 [FREE Full text] [doi: 10.1007/s10484-012-9208-z] [Medline: 23178990]
26. Hourani L, Kizakevich PN, Hubal R. Predeployment stress inoculation training for primary prevention of combat-related stress disorders. J Cyberther Rehabil 2011;4(1):101-117 [FREE Full text]
27. Lewis GF, Hourani L, Tueller S, Kizakevich P, Bryant S, Weimer B, et al. Relaxation training assisted by heart rate variability biofeedback: implication for a military predeployment stress inoculation protocol. Psychophysiology 2015 Sep;52(9):1167-1174. [doi: 10.1111/psyp.12455] [Medline: 26095854]
28. Kizakevich PN, Eckhoff R, Brown J, Tueller SJ, Weimer B, Bell S, et al. PHIT for Duty™, a mobile application for stress reduction, sleep improvement, and alcohol moderation. Mil Med 2018 Mar 01;183(suppl_1):353-363. [doi: 10.1093/milmed/usx157] [Medline: 29635566]
29. Eckhoff RP, Kizakevich PN, Bakalov V, Zhang Y, Bryant SP, Hobbs MA. A platform to build mobile health apps: the Personal Health Intervention Toolkit (PHIT). JMIR Mhealth Uhealth 2015 Jun 01;3(2):e46 [FREE Full text] [doi: 10.2196/mhealth.4202] [Medline: 26033047]
30. Kizakevich PN, Eckhoff R, Weger S, Weeks A, Brown J, Bryant S, et al. A personal health information toolkit for health intervention research. Stud Health Technol Inform 2014;199:35-39. [Medline: 24875686]
31. Porges SW, Bohrer RE. Analyses of periodic processes in psycho-physiological research. In: Cacioppo JT, Tassinary LG, editors. Principles of Psychophysiology: Physical, Social, and Inferential Elements. New York, NY: Cambridge University Press; 1990:708-753.
32. Denver JW, Reed SF, Porges SW. Methodological issues in the quantification of respiratory sinus arrhythmia. Biol Psychol 2007 Feb;74(2):286-294 [FREE Full text] [doi: 10.1016/j.biopsycho.2005.09.005] [Medline: 17067734]
33. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception Psychophys 1974 Jan;16(1):143-149. [doi: 10.3758/bf03203267]
34. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
35. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014 Nov;179(11):1218-1222. [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
36. Plaza I, Demarzo MMP, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
37. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
38. 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]
39. Li KHC, White FA, Tipoe T, Liu T, Wong MC, Jesuthasan A, et al. The current state of mobile phone apps for monitoring heart rate, heart rate variability, and atrial fibrillation: narrative review. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e11606 [FREE Full text] [doi: 10.2196/11606] [Medline: 30767904]
40. Owens OL, Beer JM, Reyes LI, Gallerani DG, Myhren-Bennett AR, McDonnell KK. Mindfulness-based symptom and stress management apps for adults with chronic lung disease: systematic search in app stores. JMIR Mhealth Uhealth 2018 May 15;6(5):e124 [FREE Full text] [doi: 10.2196/mhealth.9831] [Medline: 29764800]

41. Bernardi L, Wdowczyk-Szulc J, Valenti C, Castoldi S, Passino C, Spadacini G, et al. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. J Am Coll Cardiol 2000 May;35(6):1462-1469 [FREE Full text] [Medline: 10807448]
42. Shinba T, Kariya N, Matsui Y, Ozawa N, Matsuda Y, Yamamoto K. Decrease in heart rate variability response to task is related to anxiety and depressiveness in normal subjects. Psychiatry Clin Neurosci 2008 Oct;62(5):603-609 [FREE Full text] [doi: 10.1111/j.1440-1819.2008.01855.x] [Medline: 18950382]
43. Bonnie E. The mobile marketer's guide to mastering user retention. Mountain View, CA: CleverTap; 2017 Oct 17. URL: https://clevertap.com/blog/guide-to-user-retention/

Abbreviations

BART: Biofeedback-Assisted Resilience Training
HR: heart rate
HRV: heart rate variability
ID: identification
PHIT: Personal Health Informatics and Intervention Toolkit
PIN: personal identification number
PTSD: posttraumatic stress disorder
RSA: respiratory sinus arrhythmia

Edited by G Eysenbach; submitted 23.10.18; peer-reviewed by CC Chang, C Stepnowsky, F López Seguí; comments to author 21.03.19; revised version received 30.05.19; accepted 17.06.19; published 06.09.19.

Please cite as:
Kizakevich PN, Eckhoff RP, Lewis GF, Davila MI, Hourani LL, Watkins R, Weimer B, Wills T, Morgan JK, Morgan T, Meleth S, Lewis A, Krzyzanowski MC, Ramirez D, Boyce M, Litavecz SD, Lane ME, Strange LB
Biofeedback-Assisted Resilience Training for Traumatic and Operational Stress: Preliminary Analysis of a Self-Delivered Digital Health Methodology
JMIR Mhealth Uhealth 2019;7(9):e12590
URL: https://mhealth.jmir.org/2019/9/e12590/
doi:10.2196/12590
PMID:31493325

Original Paper

Response Time as an Implicit Self-Schema Indicator for Depression Among Undergraduate Students: Preliminary Findings From a Mobile App–Based Depression Assessment

Kyungmi Chung[1,2], PhD; Jin Young Park[1,2], MD, PhD; DaYoung Joung[1,2], MA; Kyungun Jhung[3], MD, PhD

[1]Department of Psychiatry, Yonsei University College of Medicine, Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea

[2]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Psychiatry, International St. Mary's Hospital, Catholic Kwandong University College of Medicine, Incheon, Republic of Korea

Corresponding Author:
Kyungun Jhung, MD, PhD
Department of Psychiatry
International St. Mary's Hospital
Catholic Kwandong University College of Medicine
Simgokro 100gil 25 Seo-gu
Incheon, 22711
Republic of Korea
Phone: 82 32 290 3878
Fax: 82 32 290 3879
Email: kyungun12@gmail.com

Abstract

Background: Response times to depressive symptom items in a mobile-based depression screening instrument has potential as an implicit self-schema indicator for depression but has yet to be determined; the instrument was designed to readily record depressive symptoms experienced on a daily basis. In this study, the well-validated Korean version of the Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R) was adopted.

Objective: The purpose of this study was to investigate the relationship between depression severity (ie, explicit measure: total K-CESD-R Mobile scores) and the latent trait of interest in schematic self-referent processing of depressive symptom items (ie, implicit measure: response times to items in the K-CESD-R Mobile scale). The purpose was to investigate this relationship among undergraduate students who had never been diagnosed with, but were at risk for, major depressive disorder (MDD) or comorbid MDD with other neurological or psychiatric disorders.

Methods: A total of 70 participants—36 males (51%) and 34 females (49%)—aged 19-29 years (mean 22.66, SD 2.11), were asked to complete both mobile and standard K-CESD-R assessments via their own mobile phones. The mobile K-CESD-R sessions (binary scale: yes or no) were administered on a daily basis for 2 weeks. The standard K-CESD-R assessment (5-point scale) was administered on the final day of the 2-week study period; the assessment was delivered via text message, including a link to the survey, directly to participants' mobile phones.

Results: A total of 5 participants were excluded from data analysis. The result of polynomial regression analysis showed that the relationship between total K-CESD-R Mobile scores and the reaction times to the depressive symptom items was better explained by a quadratic trend—F (2, 62)=21.16, P<.001, R^2=.41—than by a linear trend—F (1, 63)=25.43, P<.001, R^2=.29. It was further revealed that the K-CESD-R Mobile app had excellent internal consistency (Cronbach alpha=.94); at least moderate concurrent validity with other depression scales, such as the Korean version of the Quick Inventory for Depressive Symptomatology-Self Report (ρ=.38, P=.002) and the Patient Health Questionnaire-9 (ρ=.48, P<.001); a high adherence rate for all participants (65/70, 93%); and a high follow-up rate for 10 participants whose mobile or standard K-CESD-R score was 13 or greater (8/10, 80%).

Conclusions: As hypothesized, based on a self-schema model for depression that represented both item and person characteristics, the inverted U-shaped relationship between the explicit and implicit self-schema measures for depression showed the potential of an organizational breakdown; this also showed the potential for a subsequent return to efficient processing of schema-consistent information along a continuum, ranging from nondepression through mild depression to severe depression. Further, it is expected

that the updated K-CESD-R Mobile app can play an important role in encouraging people at risk for depression to seek professional follow-up for mental health care.

(JMIR Mhealth Uhealth 2019;7(9):e14657) doi:10.2196/14657

KEYWORDS

depressive symptoms; response time; self-concept; mobile phone; mobile apps; diagnostic screening programs; self-assessment; treatment adherence; compliance

Introduction

Background

Why do most psychometric instruments screen for or diagnose mental health problems (eg, depression, anxiety, and stress) only based on a summed total score, requiring that the same items be administered to all individuals? On the grounds of classical test theory [1-3], traditional psychometric measurements for depressive symptoms, such as the Beck Depression Inventory-II [4], the Patient Health Questionnaire-9 (PHQ-9) [5], the Geriatric Depression Scale [6], and the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R) [7], have assumed that all items are equally weighted and that the characteristics of items cannot be separated from those of the person. In an attempt to reduce the burden on respondents of repeated exposures to long, fixed questionnaires, a number of researchers have, until recently, developed and validated a short-form of the self-report depression screening scales [8-10]. Given that there is the trade-off between efficiency and accuracy, a brief, efficient instrument can apply the fewest items to all respondents but may fail to employ the most informative items required to adequately and accurately measure the full range of clinical severity for each respondent. On the contrary, item response theory (IRT) [11,12] has not assumed that all people are measured with the same level of certainty, in that individuals with the same total score may display a wide variation in the relative severity and frequency of depressive symptoms. To deal with these limitations of existing, classical test, theory-based instruments, computerized adaptive testing based on IRT has been widely adopted; this has been used to estimate the respondent's true score on the latent trait of interest in the individual item, thereby ensuring that a small, optimal number of items are administered to each individual until a previously determined level of measurement precision of the severity estimate is obtained [13].

Previous studies on the development of IRT-based computerized adaptive testing for depression [14-17] have had a greater emphasis on increased efficiency without loss of accuracy in assessing the presence and severity of depressive symptoms. However, what these studies have neglected is that examining the potential existence and function of a self-schema in nondepressed and depressed individuals should come first. In a cognitive model of the self [18], the self is viewed as a schema whose content is built up and organized from the individual's day-to-day experiences in his or her world. As an interpretive frame for the encoding of personal data, the self-schema is activated and becomes an important part of the information processing system when the individual encounters personally relevant information. In Beck's cognitive model of depression, depressive or negative content is defined as "an enduring characteristic of the cognitive organization, present in the depression-prone individual, even when the person is not feeling depressed [19]." Due to the enduring nature of negative schemata that contribute to the occurrence and reoccurrence of other depressive symptoms [20], the existence of negative self-referent information in individuals with different levels of depression (ie, nondepressed, high-risk individuals) needs to be tracked in their everyday life.

According to a self-schema model for depression, self-referent recall enhancement can be achieved only when schema-consistent information is processed in a highly efficient manner via one's view of self, particularly for nonclinical and clinical depressives [21,22]. With a well-organized and efficient cognitive schema, the two groups would exhibit shorter response times to recall negative or depressive content in *yes/no* ratings for the self-referent judgement on experienced depressive symptoms; however, they would represent the only substantial difference in the actual content of personal information schematically represented [23]. Nonclinical and clinical depressives with similar total scores in severity of depression would produce a *content-specific* depressive self-schema, showing changes in the self-reported frequency of experienced symptoms. The self-schema model postulates not only an organizational breakdown, but also a subsequent return to effective processing of schema-consistent information [23]. In terms of the content and efficiency parameters, it would be difficult for mild depressives with a disruption in their organized and consistent view of self to efficiently process either positive or negative personal information. The lack of efficiency may result from their uncertainty regarding applicable self-referent attributes, which, in turn, would exhibit longer response times for self-referent judgements on negative information. In other words, those who have already begun to experience depressive symptoms and view themselves with negative or depressive content in their self-schema may have difficulty in the positive and precise identification of symptom severity; this is the case because positive or nondepressive content has yet to be displaced [24]. In this respect, the self-schema of mild depressives whose depression level is not severe enough to use a negative self-schema would differ from that of nonclinical and clinical depressives.

In general, it has been assumed by most cognitive theories and related works [25-27] that implicit cognitive biases stemming from activated negative self-schemata would be evident across all study designs (ie, measurement paradigms) and facets of cognition (ie, attention, memory, self-belief and interpretation, and self-esteem). Not until triggered by environmental stress do individuals vulnerable to depression possess relatively stable,

negative, self-referential implicit cognitions that remain latent [25,26]. Because of their latent trait, these cognitions are posited to affect all aspects of information processing when activated [28,29]. Particularly, automatic (ie, implicit) dysfunctional attitudes about the self, known as the key vulnerability factor for the first onset and recurrences of depression [25,30], are more likely to remain undetected if self-report questionnaires are administered to explicitly measure their beliefs and feelings. From a dual-process perspective [26], the autonomic (ie, reflexive) nature of implicit cognitions can be assessed by means of reaction time or memory association measures, such as the Implicit Associations Test [31], compared to explicit cognitions measured by individuals' deliberate (ie, reflective) consideration. If it may be possible to identify depression-vulnerable individuals and their self-schema based on patterns of explicit and implicit cognitions, using a person-level approach will be recommended to investigate how each self-schema is uniquely associated with different levels of depression severity [27]. With advancements in hardware and software technology, a wide variety of computerized implicit measures could be run on mobile devices, such as mobile phones and tablets, as well as in a laboratory or other experimentally controlled setting. However, explicit and implicit measures based more on the standard approach may contribute to relatively low accessibility for self-administered depression assessment tools in one's daily life.

For measuring an accurate *latent trait of interest* in schematic self-referent processing of depressive symptom items, a mobile-based experience sampling method, also known as mobile-based ecological momentary assessment (mEMA), can be utilized. As a *time-stamped*, self-reported data collection method [32-34], mHealth apps can help users capture momentary psychological symptoms in their everyday lives in a timely and unconscious manner by recording entry and completion times with high contextual precision. More importantly, the apps can motivate those at high risk for depression to seek professional help, to discuss their screening test results with mental health care practitioners and professionals, and to take appropriate action against previously undiagnosed mental health problems [33,35,36]. While a majority of researchers have recently developed and validated a mobile-based *prospective* assessment tool [37-39], others have chosen and improved one of the well-established, standardized screening or diagnostic instruments to be optimized for mobile platforms [33,40-42]. While many measures ask an individual to recall depressive symptoms present in the previous weeks, mEMA is less influenced by recall bias, as individuals report symptoms that were present on that day. Taken together, there is a need to bridge the gap between standard and applied depression assessment tools. A number of depression apps available in app stores featured a therapeutic treatment (33.7%) or psychoeducation (32.1%) function, followed by medical assessment (16.9%), symptom management (8.2%), and supportive resources (1.6%) [43]. However, it should be determined whether the use of mobile phone and app technology in screening and management of depression is ecologically and clinically valid in order for these technologies to be employed in clinical practice as well as in large-scale epidemiological studies.

Objective

The objective of this study is to allow people to readily record depressive symptoms they have experienced on a daily basis via their own mobile phones. For this purpose, we chose the Korean version of the CESD-R (K-CESD-R) [44], which is available for public use as one of the most widely used and well-validated depression screening instruments in the field of psychiatric epidemiology. We developed the K-CESD-R Mobile app in our previous study [42]. The purpose of this study is to examine the relationship between depression severity (ie, explicit measure: total K-CESD-R Mobile scores) and the latent trait of interest in depressive symptom items (ie, implicit measure: response times) in undergraduate students who had never been diagnosed with, but were at risk for, major depressive disorder (MDD) or comorbid MDD with other neurological or psychiatric disorders. We could thereby trace and understand the possible differences in schematic self-referent processing along a continuum ranging from nondepression through increasing levels of severity to clinical depression. It can be hypothesized that participants would more quickly respond to schema-compatible information than schema-incompatible information, thus presenting an inverted-U pattern between the total scores and response times. Based on the findings of this study, the potential of response times to depressive symptom items as an implicit self-schema indicator for depression will be determined. Furthermore, methodological discussion will be helpful to enhance the quality of the depression assessment to be used in both community and clinical samples.

Methods

Recruitment

This study was part of a government-driven project for developing mobile app-based intervention technology to identify South Korean college and university students vulnerable to mental health problems and to help them seek professional help. Therefore, undergraduate students who were 19 years of age or older and had never been diagnosed with either MDD or comorbid MDD with other neurological or psychiatric disorders were eligible to participate in this study. The study application was posted via online advertisements on several university websites in Seoul, South Korea. The online advertisement included the following information: aim of the study, inclusion and exclusion criteria, reward for participation, number of target participants, study period and procedure, and contact information.

A total of 70 undergraduate students—36 males (51%) and 34 females (49%)—who returned the participation application via email were recruited as healthy controls; they ranged in age from 19 to 29 years (mean 22.66, SD 2.11). In addition to the other inclusion and exclusion criteria for recruitment, the volunteers were required to have their own mobile phones with a screen size of at least 4 inches diagonally to control for variables that might affect reaction times. Based on Fitts's Law [45], the size of a target (eg, either a *yes* or *no* button) to tap and its distance from the user's current position (ie, hand gestures and fingertip locations) within the user interface had to be carefully considered. After providing signed informed

consent, all the volunteers were enrolled. On the basis of the exclusion criteria for data analysis, those who did not assess depressive symptoms for at least 7 days and complete both standard and mobile K-CESD-R assessments were excluded from the statistical analysis. All were paid KRW 30,000 for their participation. This study was approved by the Institutional Review Board of Gangnam Severance Hospital.

In this study, we attempted to determine whether the updated K-CESD-R Mobile app could motivate its users to adhere to the self-administered assessment for 2 weeks. We also sought to determine whether the app could motivate those at risk for depression to seek further diagnostic interviews, as provided by the guidance on the interpretation of test results from the K-CESD-R Mobile app. As we intended to observe the adherence rate by the users of the app, information on further follow-up after finishing the 2-week course of the standard and mobile K-CESD-R assessments was not given to volunteers via the advertisement.

Standard K-CESD-R Scale Versus Applied K-CESD-R Mobile App

To overcome the limitation of the retrospective recall-based K-CESD-R assessment, we had previously developed *K-CESD-R Mobile*, a mobile-optimized, daily self-report, depression screening tool; for a review, see Chung et al's study [42]. Based on a *frequency* approach, an original version of the K-CESD-R scale instructed participants to indicate how often they have experienced each of the 20 symptom items, as defined by the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria. Participants were to indicate the frequency of symptoms over the past 2 weeks using the following 5-point response format ranging from 0 to 4: 0 (*Not at all* or *less than one day*), 1 (*1-2 days*), 2 (*3-4 days*), 3 (*5-7 days*), and 4 (*Nearly every day for 2 weeks*) [7]. In addition, the K-CESD-R Mobile app asked participants to indicate whether or not they have experienced depressive symptoms (20 items) in the past 24 hours using a *Yes* or *No* response format for the following 2 weeks. Responses of *Yes* and *No* were coded as 1 and 0, respectively. It was recommended to them that each session should be completed 24 hours after the previous session. Otherwise, participants could freely complete the assessment at any time within the specific time window of 6 hours before or after the recommended time (ie, every 24 hours), as displayed on the home screen of the app.

Furthermore, the K-CESD-R Mobile app applied a *ratio* approach to deal with the problem of missing data in case participants administered the assessment for at least 7 days or more, but not for all the days, during the 2-week study period. If participants completed less than 7 daily sessions during the study period, their final scores were not computed after completing the final session. To apply the same standard to compare the two K-CESD-R scores, we developed a new algorithm to convert a binary response to a 5-point response with different cutoff criteria: 0 ($0 \leq Y < 2/14$), 1 ($2/14 \leq Y < 5/14$), 2 ($5/14 \leq Y < 9/14$), 3 ($9/14 \leq Y < 13/14$), and 4 ($13/14 \leq Y = 14/14$); Y=Q/P, where Q is equal to the total number of times users responded *Yes* to each item, and P is equal to the total number of days that users completed sessions over 2 weeks.

We tested the feasibility and validity of the K-CESD-R Mobile scale and its converting algorithm in our previous research [42]. According to the validation study of the K-CESD-R [44], both of the total K-CESD-R scores could range from 0 to 80, with a cutoff score of 13 or more.

In addition to the K-CESD-R scores, response times were recorded as the latent trait of interest in schema-compatible information, particularly on depressive symptoms experienced. The *latent trait of interest* was defined as the interval between the initial presentation of each item via the K-CESD-R Mobile app installed on participants' mobile phones and their *Yes* or *No* responses to the item. To minimize the Hawthorne effect and improve the generalizability of results, no records of the response times were displayed on the app, nor were participants informed that their response time data would be collected. Only authorized investigators were allowed to access the raw response time data by signing in to a Web dashboard with a user ID and password and unlocking the data file with a different password.

Procedure

For the initial visit, participants gathered at the same time in the grand auditorium at Gangnam Severance Hospital. Participants who returned their signed informed consent forms were asked to fill out a paper-and-pencil prequestionnaire to collect their demographic information and self-report ratings on depression scales, such as the Korean version of Quick Inventory for Depressive Symptomatology-Self Report (KQIDS-SR) [46,47] and the Korean version of the PHQ-9 [48].

To collect mental health data only from enrolled participants via the K-CESD-R Mobile app, they had to be registered in advance as beta testers of the app for the study period. iOS users were guided to download the beta app, K-CESD-R Mobile, only through Apple's TestFlight platform for beta testing; Android users could directly search for and download the K-CESD-R Mobile app on the Google Play store. The K-CESD-R Mobile app required the participants' consent to collect and use their data for research purposes under Korea's Personal Information Protection Act, particularly at the final step of the sign-up process; this was to further ensure security and privacy for sensitive personal information collected and transmitted via mobile devices to a cloud services platform (ie, Amazon Web Services). After reviewing and agreeing to all terms of use and a privacy statement, participants could create an account and start a new session.

To lead them to perform a 2-week K-CESD-R test in a comfortable but controlled manner, a warm-up screen with a *Start* button was sequentially followed by a guidance screen and 20 K-CESD-R item screens. The guidance screen was presented to instruct participants that they could start a test when they were mentally ready to administer it. At the last item screen, participants were asked to tap a *Save and Send* button to transfer their responses to the Amazon Web Services platform. On the first visit, and even on the guidance screen, the app did not let participants know that response time data was being acquired while responding to the K-CESD-R items; the reason for this was to control the quality of an implicit measure as a self-schema indicator as well as to prevent participants from

misunderstanding the aim of the study. In an effort to ensure adequate measurement of response times during the test, the app did not allow users to sign in or start the test unless a stable Internet connection via Wi-Fi or cellular network could be guaranteed; this was to prevent results being affected by Internet quality.

Following the completion of the first mobile K-CESD-R session, all participants received guidance from experimenters on the given tasks: (1) the remaining mobile K-CESD-R sessions should be administered on a daily basis for 2 weeks and (2) a standard K-CESD-R assessment created with SurveyMonkey [49] should be administered on the final day of the 2-week study period; this survey was delivered via text message and included a link to the survey. Accordingly, both mobile and standard K-CESD-R assessments ended on the same day.

After scoring was completed, participants whose online K-CESD-R score or mobile K-CESD-R score was 13 or above were recommended to have clinician-administered diagnostic interviews. It was explained to all participants at the first visit that the CESD-R was designed as a quick and reliable self-administered screening tool for depression, regardless of the platforms on which they were provided. To make the diagnosis of clinical depression, an initial screening of participants with these instruments would need to be followed by clinical interviews based on their K-CESD-R scores. Follow-up visits took place at the outpatient clinic in the Department of Psychiatry, Gangnam Severance Hospital; each participant was individually scheduled to come at a convenient time in order to motivate him or her to discuss mental health problems with a medical doctor. During the 30-minute clinical interview, the following scales were administered: the original English version of the Clinical Global Impressions-Severity of Illness Scale (CGI-S) [50], the Korean version of the Montgomery-Asberg Depression Rating Scale (K-MADRS) [51,52], the Korean version of the Hamilton Anxiety Rating Scale (K-HAM-A) [53,54], the Korean version of the Hamilton Depression Rating Scale (K-HAM-D) [55,56], and the Korean version of the Mini-International Neuropsychiatric Interview (MINI) version 5.0.0. [57,58]. After the second interview, further follow-up was not required.

Statistical Analysis

Statistical analyses were performed using PASW Statistics 18 software (SPSS Inc). Cronbach alpha was calculated to evaluate the internal consistency of the standard and mobile K-CESD-R scales. As nonparametric alternatives to the paired-samples *t* test and Pearson's correlation test, a Wilcoxon signed-rank test was used to compare the difference between the standard and mobile K-CESD-R scores whose normality assumptions were not satisfied. Spearman's correlation coefficient was then calculated to measure concurrent validity of the K-CESD-R Mobile scale with other depression screening scales, such as the standard K-CESD-R, the KQIDS-SR, and the PHQ-9. In order to determine whether the relationship between the explicit and implicit self-schema measures for depression would be better explained by a quadratic trend than by a linear trend, the polynomial regression analysis was conducted after the normal distributions of the variables were confirmed by a Kolmogorov-Smirnov test.

Results

Participant Characteristics

After ensuring that enrolled undergraduate students met our inclusion and exclusion criteria for the statistical analysis, 5 out of the 70 participants were excluded (93% adherence rate). This is because the K-CESD-R Mobile app was designed to calculate the test results only if its users assessed depressive symptoms for at least 7 days in the 2-week study period. Furthermore, it was also required that students complete the standard K-CESD-R assessment on the last day of the mobile K-CESD-R assessment, following the experimental protocol of this study. The detailed demographic information on all participants included in the data analysis is presented in Table 1.

Table 1. Participant demographic characteristics (N=65).

Participant characteristic	Value
Age (years), mean (SD)	22.63 (2.13)
Age (years), n (%)	
19	2 (3)
20-29	63 (97)
Gender, n (%)	
Male	32 (49)
Female	33 (51)
Mobile operating system, n (%)	
Android	35 (54)
Apple iOS	30 (46)
Marital status (single), n (%)	65 (100)
Current smoking status, n (%)	
Smoker	5 (8)
Duration of smoking in years, n (%)	
3	2 (3)
4	1 (2)
6	2 (3)
Cigarettes smoked per day, n (%)	
4	1 (2)
8	1 (2)
10	1 (2)
11	1 (2)
13	1 (2)
Nonsmoker	60 (92)
Current alcohol drinking status, n (%)	
Drinker	45 (69)
Frequency of alcohol intake per week, n (%)	
Once	25 (39)
Twice	12 (19)
Three times	7 (11)
Four times	1 (2)
Nondrinker	20 (31)

Depression Screening by Self-Reported Scales and Clinician Interview

A total of 65 participants completed all the standard K-CESD-R assessments (median 3.00, interquartile range [IQR] 0-7.50; scored from 0 to 63) and the mobile K-CESD-R assessments (median 2.00, IQR 0-6.50; scored from 0 to 59) with high-variance distributions: coefficient of variation (CV) was 1.73 and 1.90, respectively. The distribution of the standard K-CESD-R scores had a positive skew (skewness 3.91, SE 0.30) and was leptokurtic (kurtosis 17.68, SE 0.59). Similarly, the mobile K-CESD-R scores had positively skewed (skewness 3.90, SE 0.30) and leptokurtic (kurtosis 17.39, SE 0.59) distributions. The internal consistencies of the standard K-CESD-R (Cronbach alpha=.94) and the mobile K-CESD-R (Cronbach alpha=.94) scales were equivalently high.

With use of a Wilcoxon signed-rank test, we found a significant difference between the standard K-CESD-R and mobile K-CESD-R scores (Z=–2.69, *P*=.007), with a Spearman's correlation coefficient of .82 (*P*<.001). The number of participants whose depression screening score was 13 or above were as follows: (1) 9 participants based on the standard K-CESD-R, (2) 6 participants based on the mobile K-CESD-R, and (3) 5 participants based on both scales. Out of 10

participants, only 1 (10%) was consistently assessed as being above the diagnostic thresholds for depression through a clinician-administered diagnostic interview, structured using the CGI-S (score ≥3), the K-MADRS (score ≥16), the K-HAM-A (score ≥25), the K-HAM-D (score ≥19), and the Korean version of the MINI, with modules based on the DSM-IV diagnostic criteria for a major depressive episode. According to their standard and mobile K-CESD-R scores, all 10 of the participants at high risk for depression were advised to visit the clinic for a 30-minute clinical interview at their desired date and time. However, 2 out of 10 (20%) did not seek further professional follow-up, as recommended by both the K-CESD-R Mobile app and the experimenters.

Concurrent Validity of the K-CESD-R Mobile Scale

The concurrent validity of the K-CESD-R Mobile scale was assessed through Spearman's correlation between the K-CESD-R Mobile and other depression screening scales: KQIDS-SR with a total score of 0-26 (median 6.00, IQR 4.00-9.50, CV=.71; ρ=.38, *P*=.002) and the PHQ-9 with a total score of 0-19 (median 2.00, IQR 1.00-4.50, CV=1.11; ρ=.48, *P*<.001).

Inverted U-Shaped Relationship Between Depression Score and Response Time

We tested the hypothesized curvilinear (ie, inverted U-shaped) relationship between the severity of participants' depression level as an independent variable and personal negative information about themselves (ie, latent trait of interest) as a dependent variable; this was done whether the association between the two variables was best characterized by a quadratic trend or by a linear trend. The hypothesis was tested using a polynomial regression analysis. To justify the use of the parametric test, a Kolmogorov-Smirnov test was conducted for testing the normality assumption (Z=1.11, *P*=.17). Before hypothesis testing, we subtracted the mean response time for each item from each participant's response time to control for the item effect only (ie, item length and vocabulary level). This is because the widely used double standardization method, which controlled for both person effects (ie, reading and motor speed) and item effects, was criticized for its artefactual negative correlation between items varying in mean response times [59]; it was also criticized because it was revealed that the severity of depressive symptoms accounted for impairments in information processing speed and psychomotor retardation [60].

The polynomial regression analysis revealed that the relationship between total K-CESD-R Mobile scores and the reaction times to the depressive symptom items was better captured by the quadratic trend—F (2, 62)=21.16, *P*<.001, R^2=.41—than by the linear trend—F (1, 63)=25.43, *P*<.001, R^2=.29. As shown in Figure 1, this finding reflects the inverted U-shaped reaction time effect as the self-schema evidence from faster reaction times for low and high K-CESD-R Mobile scores than for intermediate ones. To consider the quality of both models and the potential outliers, scatterplots of residuals by fit values for the linear model and quadratic model were produced. Figure 2 illustrates that the residuals of the quadratic model are more evenly dispersed than those of the linear model, showing their skewed distribution. It was also found that potential outliers are less identified in the quadratic model compared to that of the linear model.

Figure 1. Results of polynomial regression analysis predicting the curvilinear relationship between total scores on the Korean version of the Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R) Mobile app and mean standardized reaction times for the depression items.

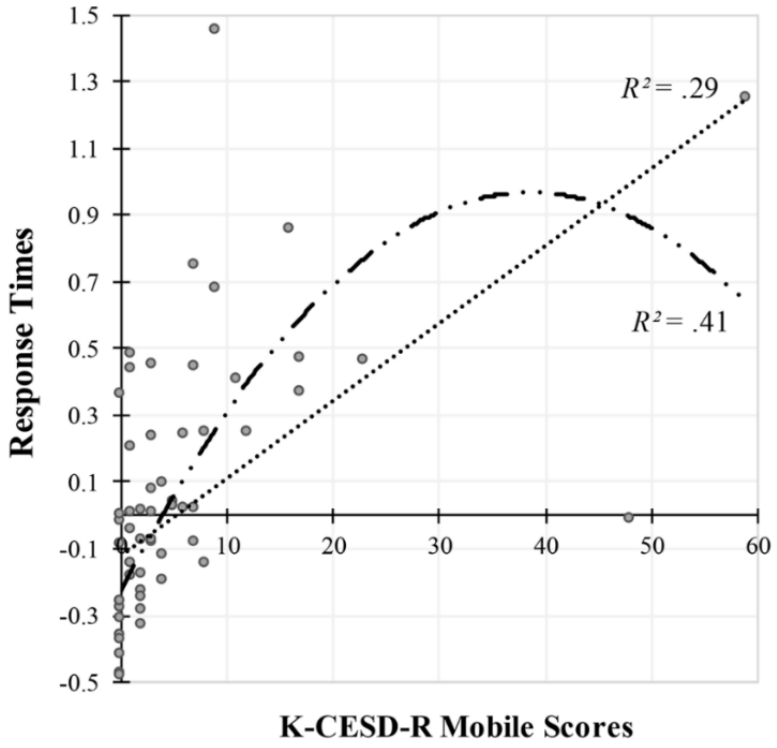
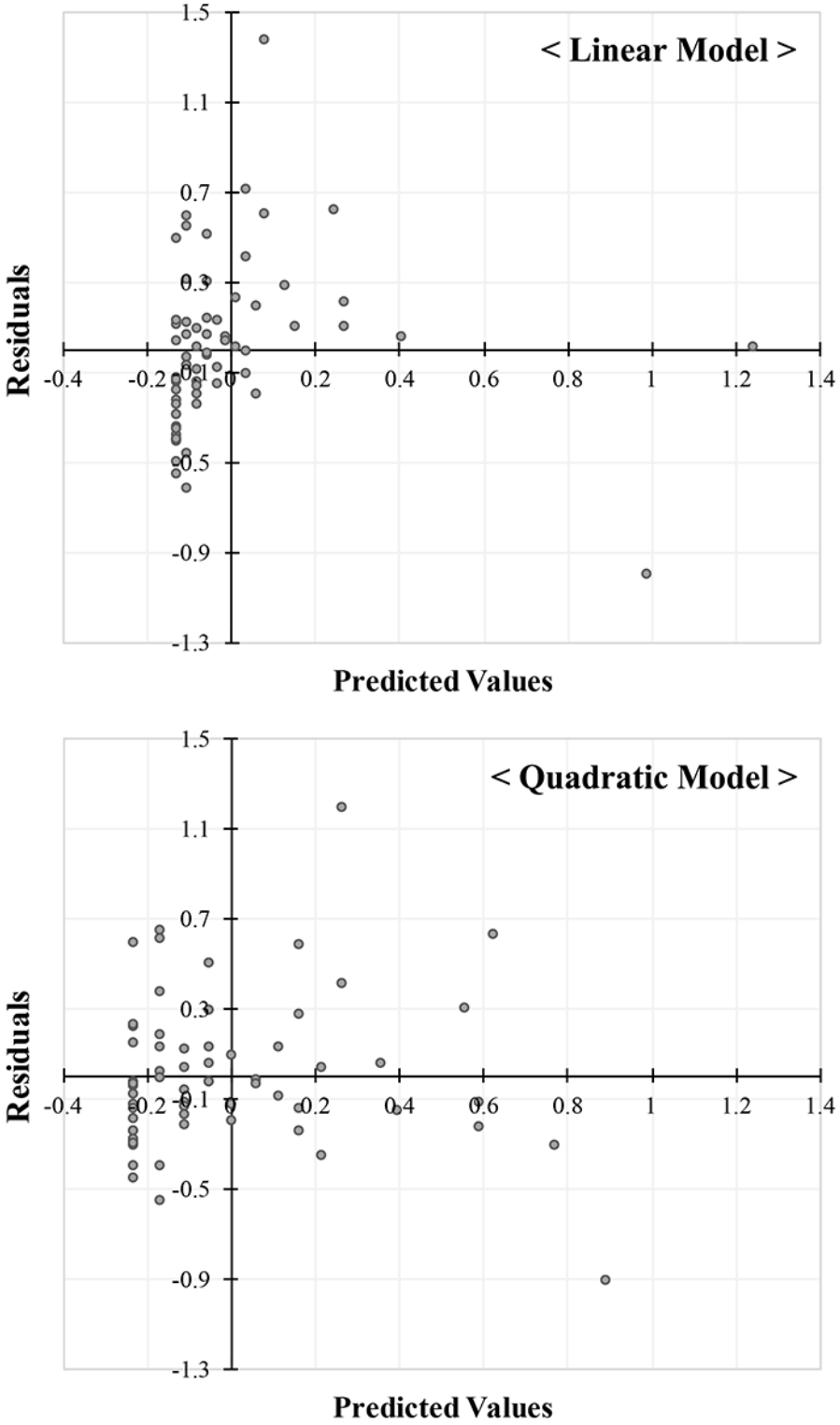


Figure 2. Scatterplots of residuals by fit values for linear and quadratic models.



Discussion

Principal Findings

The aim of this study was to investigate the relationship between the severity of depression and the latency of response to depressive symptom items in a sample of undergraduate students at risk for mental health problems. In this study, we postulated the potential existence and function of a depressive self-schema as the individual's idiosyncratic cognitive structure, with which content (ie, item) and efficiency (ie, speed) would be responsive to variations in depression level when self-relevant information is processed. Given the impairment and breakdown process in the self-schema and subjective organization of personal information in depression, we hypothesized that nondepressed and severely depressed individuals would be faster than mildly depressed individuals in making self-referent judgments on

experienced depressive symptoms with a *yes/no* response format, showing an inverted U-shaped pattern in their self-schemata.

As hypothesized by the self-schema model for depression [23], the inverted U-shaped relationship between the total K-CESD-R Mobile scores and response times to the items was found in a sample of young adult, university students who have not previously received a diagnosis of depression. It appeared to be empirically supported that individuals at risk of depression might exhibit a disruption in their organized and consistent views of self with both positive and negative information components; this is the case because positive information has yet to be displaced by negative or depressive information in their self-schema [23,24,61]. According to the statistical results and figures, the relationship between depression severity and reaction times was better explained by the quadratic model than the linear model. Figure 2 revealed that the residuals of the quadratic model were independent of the predicted fit values, and the residual distribution of the quadratic model was less skewed and more evenly dispersed than that of the linear model. It was observed in the quadratic model that potential outliers tended to be more acceptably scattered than in the linear model. However, our preliminary study had the limitations of sample size; as well, the study sample lacked participants who had been diagnosed with MDD as positive controls to show that the K-CESD-R Mobile app has the power and sensitivity to identify those students and demonstrate the inverted U-shaped relationship. In fact, we failed to determine whether the self-schema model for depression fit well with the inverted U-shaped curve in undergraduate students at all levels of depression. Beyond the limitations of the sample used and its size, it is also true that the response time, as an implicit self-schema indicator for depression, showed a *potential* for enhancing the quality of mobile app-based depression assessment and screening between nondepressed, mildly depressed, and severely depressed individuals.

Most of all, the K-CESD-R Mobile app was designed to make up for the weak points of traditional depression screening tools, which have rarely been used after the diagnosis of depression and have discriminated against individuals at risk for depression with optimal cutoff values. To develop a mobile-based depression intervention for identifying undergraduate students with mental health problems, we adopted the K-CESD-R Mobile app [42], with which individuals could assess their depressive symptoms experienced in the past 24 hours with *yes* or *no* ratings during a 2-week period. Considering that the standard K-CESD-R scale [44] asked respondents to choose response options from 1 to 5 based on how many days they have experienced the given symptoms during the past 1-2 weeks, the mobile K-CESD-R scale would contribute to reducing the possibility of recall bias from the retrospective depression assessment with a longer recall period. As far as the treatment for depression and the assessment of remission status are concerned, defining remission status from depression based only on the total scores of explicit symptom-based measures is not recommended; this is because there exists both discordance and concordance between the self-ratings of depression symptom severity and psychosocial functioning impairments [62]. As the next step, to estimate the respondents' *true* scores, the app was updated to explicitly and implicitly measure depression severity on a daily basis by adding a new feature for acquiring the response latencies for all items with *total* K-CESD-R Mobile scores. The key feature was based on a developmental approach to the acquisition of *latent* negative self-schemata; prior depressive experience or repeated associations between current depressed mood and thoughts or memories is more likely to increase accessibility to negative cognitions once the self-schema has been activated [63]. Compared to the depression Implicit Associations Test that respondents should complete via computer- or mobile-based apps (eg, E-prime or Inquisit) in a controlled, uncomfortable environment to support the validity of the data collection, the K-CESD-R Mobile app installed on their own mobile phones automatically measures the response times to all items without requiring further test procedures. The mobile app employs a streamlined approach to a well-established depression screening tool for epidemiologic studies; therefore, the app would make it possible for both its users and practitioners to rapidly but accurately detect depressive symptoms and their severity and monitor any shift in the content and efficiency of the self-schema throughout the lifetime.

Moreover, the K-CESD-R Mobile app was expected to encourage positively screened individuals to seek professional help; this was to be done before an organizational breakdown and possible shifts for a subsequent return to efficient processing of schema-consistent information are completed. To do so, the K-CESD-R Mobile app provides clinicians and mental health professionals with access to the online dashboard to implement a personal-level intervention for individuals, simultaneously monitoring their adherence to the app and total scores from separate remote locations. However, we limited the scope of this study to the enhancement of the K-CESD-R Mobile app, not to that of its dashboard. Once participants produced their total scores on the last day of the standard and mobile K-CESD-R assessments, those whose online or mobile K-CESD-R scores (or both) were 13 or over (n=10) were directly contacted via text message and phone call, in order to make an appointment for the structured clinical interview. With an adherence rate of approximately 93% (65/70) on the depression assessments, 8 of the 10 participants (80%) visited the outpatient clinic for the diagnostic interview. In fact, it would be difficult to rule out the possibility that the high adherence rate to the app might result from the financial remuneration for their participation or the Hawthorne effect. Another possibility is that the participants perceived the app to be so credible that they decided to seek further medical attention for their depressive symptoms. Despite the benefits of successful screening and brief intervention for groups at high risk of depression, this study points out the need to further develop and implement more detailed, tailored, and evidence-based interventions for those whose depressive symptoms do not interfere with or cause difficulties in their lives. When they do not consider their psychosocial functioning as severely impaired, they are less likely to seek professional help and more likely to believe themselves to be in remission [62]. To shed light on this issue, we suggest major updates to enable users to decide what other explicit measures (eg, psychological impairment and quality of life) to include with the mobile K-CESD-R scale in the app. We also suggest that

practitioners be allowed to send user-centered push notification messages via the online dashboard, thereby motivating users to adhere to the Internet- and mobile-based intervention platforms.

Limitations

This study was part of a government-driven project for developing intervention technology to identify and help college and university students at high risk for mental health problems; therefore, the sample was only composed of undergraduate students from different university campuses who have not previously been diagnosed with clinical depression. Given the main purpose of this study to achieve the aim of the project, the sampling method limits the generalizability of the findings by recruiting students who are easily accessible and willing to participate in the research, in comparison to subclinical or clinical samples and even other nonclinical South Korean samples. Furthermore, the small sample size may lead this study to have insufficient power to identify clinically relevant differences. In addition to the sample size and characteristics, another concern about the rating algorithm used in the K-CESD-R Mobile app can be raised. To convert binary response data to 5-point response data, and to automatically calculate a total score if response data for at least 7 days was collected, we employed the ratio approach-based algorithm, which was developed in our previous study [42]. Despite this attempt to deal with the absence of response data from the possible missing days, the algorithm could be biased because the number of days that the mobile K-CESD-R scale had been completed could be influenced by the *yes* or *no* type of response. For example, people could be less likely to respond on the day when they were more depressed, which might underestimate their depression severity. Taken together, these limitations can be dealt with by replicating this experimental protocol and assessing test-retest reliability and validity of the app in a large sample of clinical depressives, as well as among nonclinical and subclinical depressives. This would allow us to test the feasibility of the intervention platforms and extend the findings of this study.

Conclusions

In conclusion, this study showed preliminary evidence that the inverted U-shaped pattern of response times to all items would reflect the self-schema for depression, which was organized for the efficient processing of schema-consistent personal information on depressive symptoms experienced. High-risk adult students with unstable and incomplete depressive self-schemata, as well as mental health professionals, could benefit from measuring and analyzing response latency as an implicit self-schema indicator for depression; this could be done particularly via the K-CESD-R Mobile app and its compatible online dashboard for early intervention in depression management.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HM16C1991).

Authors' Contributions

KC and JYP contributed to the conception and design of the study. KC and DJ organized the database and performed the statistical analysis. KC wrote the first draft and all sections of the manuscript. KJ and JYP supervised the research. KJ acquired the funding for this study.

Conflicts of Interest

None declared.

References

1. Allen MJ, Yen WM. Introduction to Measurement Theory. Long Grove, IL: Waveland Press; 2002.
2. DeVellis RF. Classical test theory. Med Care 2006 Nov;44(11 Suppl 3):S50-S59. [doi: 10.1097/01.mlr.0000245426.10853.30] [Medline: 17060836]
3. Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Charlotte, NC: Information Age Publishing (IAP); 2008.
4. Beck AT, Steer RA, Brown GK. Beck Depression Inventory®—II (BDI®–II): Manual. Boston, MA: Psychological Corporation; 1996.
5. 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]
6. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
7. Eaton WW, Smith C, Ybarra M, Muntaner C, Tien A. Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults. Volume 3. 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:363-377.

8. Mitchell A, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: A meta-analysis of GDS30 and GDS15. J Affect Disord 2010 Sep;125(1-3):10-17. [doi: 10.1016/j.jad.2009.08.019] [Medline: 19800132]
9. Björgvinsson T, Kertz SJ, Bigda-Peyton JS, McCoy KL, Aderka IM. Psychometric properties of the CES-D-10 in a psychiatric sample. Assessment 2013 Aug 18;20(4):429-436. [doi: 10.1177/1073191113481998] [Medline: 23513010]
10. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. Gen Hosp Psychiatry 2015 Nov;37(6):567-576. [doi: 10.1016/j.genhosppsych.2015.06.012] [Medline: 26195347]
11. Hulin CL, Drasgow F, Parsons CK. Item Response Theory: Application to Psychological Measurement. Belmont, CA: Dorsey Press; 1983.
12. Hambleton RK, Swaminathan H. Item Response Theory: Principles and Applications. Dordrecht, the Netherlands: Springer Netherlands; 1985.
13. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016 Mar 28;12(1):83-104. [doi: 10.1146/annurev-clinpsy-021815-093634] [Medline: 26651865]
14. Smits N, Cuijpers P, van Straten A. Applying computerized adaptive testing to the CES-D scale: A simulation study. Psychiatry Res 2011 Jun 30;188(1):147-155 [FREE Full text] [doi: 10.1016/j.psychres.2010.12.001] [Medline: 21208660]
15. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012 Nov;69(11):1104-1112 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.14] [Medline: 23117634]
16. Iwata N, Kikuchi K, Fujihara Y. The usability of CAT system for assessing the depressive level of Japanese: A study on psychometric properties and response behavior. Int J Behav Med 2016 Aug 14;23(4):427-437. [doi: 10.1007/s12529-015-9503-1] [Medline: 26272358]
17. Loe BS, Stillwell D, Gibbons C. Computerized adaptive testing provides reliable and efficient depression measurement using the CES-D scale. J Med Internet Res 2017 Sep 20;19(9):e302 [FREE Full text] [doi: 10.2196/jmir.7453] [Medline: 28931496]
18. Kuiper NA, Derry PA. The self as a cognitive prototype: An application to person perception and depression. In: Cantor N, Kihlstrom JF, editors. Personality, Cognition, and Social Interaction. Abingdon, UK: Routledge; 2017:215-232.
19. Weintraub M, Segal RM, Beck AT. An investigation of cognition and affect in the depressive experiences of normal man. J Consult Clin Psychol 1974 Dec;42(6):911. [doi: 10.1037/h0037522] [Medline: 4612061]
20. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, NY: The Guilford Press; 1979.
21. Derry PA, Kuiper NA. Schematic processing and self-reference in clinical depression. J Abnorm Psychol 1981;90(4):286-297. [doi: 10.1037/0021-843X.90.4.286]
22. Kuiper NA, Olinger LJ, MacDonald MR, Shaw BF. Self-schema processing of depressed and nondepressed content: The effects of vulnerability to depression. Soc Cogn 1985 Mar;3(1):77-93. [doi: 10.1521/soco.1985.3.1.77]
23. Kuiper NA, Derry PA, MacDonald MR. Self-reference and person perception in depression: A social cognition perspective. In: Weary G, Mirels HL, editors. Integrations of Clinical and Social Psychology. New York, NY: Oxford University Press; 1982:79-103.
24. Kuiper NA, MacDonald MR. Self and other perception in mild depressives. Soc Cogn 1982 Sep;1(3):223-239. [doi: 10.1521/soco.1982.1.3.223]
25. Beck AT. Depression: Causes And Treatment. Philadelphia, PA: University Of Pennsylvania Press; 1970.
26. Beevers CG. Cognitive vulnerability to depression: A dual process model. Clin Psychol Rev 2005 Nov;25(7):975-1002. [doi: 10.1016/j.cpr.2005.03.003] [Medline: 15905008]
27. Phillips WJ, Hine DW, Thorsteinsson EB. Implicit cognition and depression: A meta-analysis. Clin Psychol Rev 2010 Aug;30(6):691-709. [doi: 10.1016/j.cpr.2010.05.002] [Medline: 20538393]
28. Beck AT. The evolution of the cognitive model of depression and its neurobiological correlates. Am J Psychiatry 2008 Aug;165(8):969-977. [doi: 10.1176/appi.ajp.2008.08050721] [Medline: 18628348]
29. Ingram RE, Miranda J, Segal Z. Cognitive vulnerability to depression. In: Alloy LB, Riskind JH, editors. Cognitive Vulnerability to Emotional Disorders. Mahwah, NJ: Lawrence Erlbaum Associates; 2006:63-91.
30. Clark DA, Beck AT. Scientific Foundations of Cognitive Theory and Therapy of Depression. New York, NY: John Wiley & Sons; 1999.
31. Greenwald AG, McGhee DE, Schwartz JL. Measuring individual differences in implicit cognition: The implicit association test. J Pers Soc Psychol 1998 Jun;74(6):1464-1480. [doi: 10.1037//0022-3514.74.6.1464] [Medline: 9654756]
32. 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]
33. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]

34. van Ballegooijen W, Ruwaard J, Karyotaki E, Ebert DD, Smit JH, Riper H. Reactivity to smartphone-based ecological momentary assessment of depressive symptoms (MoodMonitor): Protocol of a randomised controlled trial. BMC Psychiatry 2016 Oct 21;16(1):359 [FREE Full text] [doi: 10.1186/s12888-016-1065-5] [Medline: 27769201]
35. Wichers M, Simons CJ, Kramer IM, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011 Oct;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
36. BinDhim NF, Alanazi EM, Aljadhey H, Basyouni MH, Kowalski SR, Pont LG, et al. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? J Med Internet Res 2016 Jun 27;18(6):e156 [FREE Full text] [doi: 10.2196/jmir.5726] [Medline: 27349441]
37. van Os J, Delespaul P, Wigman J, Myin-Germeys I, Wichers M. Beyond DSM and ICD: Introducing "precision diagnosis" for psychiatry using momentary assessment technology. World Psychiatry 2013 Jun;12(2):113-117 [FREE Full text] [doi: 10.1002/wps.20046] [Medline: 23737412]
38. Nahum M, Van Vleet TM, Sohal VS, Mirzabekov JJ, Rao VR, Wallace DL, et al. Immediate mood scaler: Tracking symptoms of depression and anxiety using a novel mobile mood scale. JMIR Mhealth Uhealth 2017 Apr 12;5(4):e44 [FREE Full text] [doi: 10.2196/mhealth.6544] [Medline: 28404542]
39. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional well-being predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
40. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: Cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan 17;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
41. Kim J, Lim S, Min YH, Shin Y, Lee B, Sohn G, et al. Depression screening using daily mental-health ratings from a smartphone application for breast cancer patients. J Med Internet Res 2016 Aug 04;18(8):e216 [FREE Full text] [doi: 10.2196/jmir.5598] [Medline: 27492880]
42. Chung K, Jeon M, Park J, Lee S, Kim CO, Park JY. Development and evaluation of a mobile-optimized daily self-rating depression screening app: A preliminary study. PLoS One 2018;13(6):e0199118 [FREE Full text] [doi: 10.1371/journal.pone.0199118] [Medline: 29944663]
43. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: A review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
44. Lee S, Oh S, Ryu S, Jun J, Lee K, Lee E. Validation of the Korean version of Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R). Korean J Psychosom Med 2016;24(1):83-93. [doi: 10.1093/ijnp/pyw043.223]
45. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 1954;47(6):381-391. [doi: 10.1037/h0055392]
46. Rush A, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep;54(5):573-583. [doi: 10.1016/S0006-3223(02)01866-8]
47. Yoon J, Jon D, Hong H, Hong N, Seok J. Reliability and validity of the Korean version of Inventory for Depressive Symptomatology. Mood Emot 2012 Oct;10(3):131-151 [FREE Full text]
48. Park S, Choi H, Choi J, Kim K, Hong J. Reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9). Anxiety Mood 2010;6(2):119-124 [FREE Full text]
49. SurveyMonkey. URL: https://www.surveymonkey.com/ [accessed 2019-08-02]
50. Guy W. Clinical Global Impressions (CGI) Scale, Modified. In: Rush AJ, editor. Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association; 2000.
51. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr 01;134(4):382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
52. Ahn Y, Lee K, Yi J, Kang M, Kim D, Kim J. A validation study of the Korean version of the Montgomery-Asberg Depression Rating Scale. J Korean Neuropsychiatry Assoc 2005;44(4):466-476 [FREE Full text]
53. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [Medline: 13638508]
54. Kim C. Psychiatric Assessment Instruments. Seoul, South Korea: Hana Medical Books; 2000.
55. Hamilton M. The Hamilton Rating Scale for Depression. In: Sartorius N, Ban TA, editors. Assessment of Depression. Berlin, Germany: Springer, Berlin, Heidelberg; 1986:143-152.
56. Yi J, Bae S, Ahn Y, Park D, Noh K, Shin H. Validity and reliability of the Korean version of the Hamilton Depression Rating Scale (K-HDRS). J Korean Neuropsychiatry Assoc 2005;44(4):456-465 [FREE Full text]
57. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. Eur Psychiatry 1997;12(5):224-231 [FREE Full text] [doi: 10.1016/S0924-9338(97)83296-8]
58. Yoo S, Kim Y, Noh J, Oh K, Kim C, Nam K. Validity of Korean version of the Mini-International Neuropsychiatric Interview. Anxiety Mood 2006 Jun;2(1):50-55 [FREE Full text]

59. Neubauer AC, Malle BF. Questionnaire response latencies: Implications for personality assessment and self-schema theory. Eur J Psychol Assess 1997 May;13(2):109-117. [doi: 10.1027/1015-5759.13.2.109]
60. McDermott LM, Ebmeier KP. A meta-analysis of depression severity and cognitive function. J Affect Disord 2009 Dec;119(1-3):1-8. [doi: 10.1016/j.jad.2009.04.022] [Medline: 19428120]
61. Rogers TB, Kuiper NA, Kirker WS. Self-reference and the encoding of personal information. J Pers Soc Psychol 1977;35(9):677-688. [doi: 10.1037//0022-3514.35.9.677]
62. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Boerescu D, Attiullah N. Discordance between self-reported symptom severity and psychosocial functioning ratings in depressed outpatients: Implications for how remission from depression should be defined. Psychiatry Res 2006 Feb 28;141(2):185-191. [doi: 10.1016/j.psychres.2005.05.016] [Medline: 16499976]
63. Teasdale JD. Negative thinking in depression: Cause, effect, or reciprocal relationship? Adv Behav Res Ther 1983 Jan;5(1):3-25. [doi: 10.1016/0146-6402(83)90013-9]

Abbreviations

CESD-R: Center for Epidemiologic Studies Depression Scale-Revised
CGI-S: Clinical Global Impressions-Severity of Illness Scale
CV: coefficient of variation
DSM-IV: fourth edition of the Diagnostic and Statistical Manual of Mental Disorders
IQR: interquartile range
IRT: item response theory
K-CESD-R: Korean version of the CESD-R
K-HAM-A: Korean version of the Hamilton Anxiety Rating Scale
K-HAM-D: Korean version of the Hamilton Depression Rating Scale
K-MADRS: Korean version of the Montgomery-Asberg Depression Rating Scale
KQIDS-SR: Korean version of the Quick Inventory for Depressive Symptomatology-Self Report
MDD: major depressive disorder
mEMA: mobile-based ecological momentary assessment
MINI: Mini-International Neuropsychiatric Interview
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 08.05.19; peer-reviewed by WJ Choi, JD Ni; comments to author 31.05.19; revised version received 21.06.19; accepted 04.07.19; published 13.09.19.

Please cite as:
Chung K, Park JY, Joung D, Jhung K
Response Time as an Implicit Self-Schema Indicator for Depression Among Undergraduate Students: Preliminary Findings From a Mobile App–Based Depression Assessment
JMIR Mhealth Uhealth 2019;7(9):e14657
URL: https://mhealth.jmir.org/2019/9/e14657/
doi:10.2196/14657
PMID:31586362

Original Paper

Comparison of On-Site Versus Remote Mobile Device Support in the Framingham Heart Study Using the Health eHeart Study for Digital Follow-up: Randomized Pilot Study Set Within an Observational Study Design

Nicole L Spartano[1,2], PhD; Honghuang Lin[3], PhD; Fangui Sun[4], PhD; Kathryn L Lunetta[4], PhD; Ludovic Trinquart[4], PhD, MPH; Maureen Valentino[2], BSc; Emily S Manders[2], BSc; Mark J Pletcher[5], MD, MPH; Gregory M Marcus[6], MD, MAS; David D McManus[7], MD; Emelia J Benjamin[2,8,9], MD, ScM; Caroline S Fox[2,10], MD, MPH; Jeffrey E Olgin[6*], MD; Joanne M Murabito[2,8,11*], MD, ScM

[1]Section of Endocrinology, Diabetes, Nutrition, and Weight Management, Boston University School of Medicine, Boston, MA, United States
[2]Framingham Heart Study, Framingham, MA, United States
[3]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States
[4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States
[5]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States
[6]Division of Hospital Medicine, University of California, San Francisco, San Francisco, CA, United States
[7]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
[8]Boston University School of Medicine, Boston, MA, United States
[9]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States
[10]Merck Research Laboratories, Boston, MA, United States
[11]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Nicole L Spartano, PhD
Section of Endocrinology, Diabetes, Nutrition, and Weight Management
Boston University School of Medicine
720 Harrison Ave, Suite 8100
Boston, MA
United States
Phone: 1 3154152040
Email: spartano@bu.edu

Abstract

Background: New electronic cohort (e-Cohort) study designs provide resource-effective methods for collecting participant data. It is unclear if implementing an e-Cohort study without direct, in-person participant contact can achieve successful participation rates.

Objective: The objective of this study was to compare 2 distinct enrollment methods for setting up mobile health (mHealth) devices and to assess the ongoing adherence to device use in an e-Cohort pilot study.

Methods: We coenrolled participants from the Framingham Heart Study (FHS) into the FHS–Health eHeart (HeH) pilot study, a digital cohort with infrastructure for collecting mHealth data. FHS participants who had an email address and smartphone were randomized to our FHS-HeH pilot study into 1 of 2 study arms: remote versus on-site support. We oversampled older adults (age ≥65 years), with a target of enrolling 20% of our sample as older adults. In the remote arm, participants received an email containing a link to enrollment website and, upon enrollment, were sent 4 smartphone-connectable sensor devices. Participants in the on-site arm were invited to visit an in-person FHS facility and were provided in-person support for enrollment and connecting the devices. Device data were tracked for at least 5 months.

Results: Compared with the individuals who declined, individuals who consented to our pilot study (on-site, n=101; remote, n=93) were more likely to be women, highly educated, and younger. In the on-site arm, the connection and initial use of devices

was ≥20% higher than the remote arm (mean percent difference was 25% [95% CI 17-35] for activity monitor, 22% [95% CI 12-32] for blood pressure cuff, 20% [95% CI 10-30] for scale, and 43% [95% CI 30-55] for electrocardiogram), with device connection rates in the on-site arm of 99%, 95%, 95%, and 84%. Once connected, continued device use over the 5-month study period was similar between the study arms.

Conclusions: Our pilot study demonstrated that the deployment of mobile devices among middle-aged and older adults in the context of an on-site clinic visit was associated with higher initial rates of device use as compared with offering only remote support. Once connected, the device use was similar in both groups.

(JMIR Mhealth Uhealth 2019;7(9):e13238) doi:10.2196/13238

KEYWORDS

wearable electronic devices; cell phone; fitness trackers; electrocardiography; epidemiology

Introduction

Background

Recent advances in mobile health (mHealth) technology have improved the feasibility of collecting digital data and have the potential to revolutionize both research and health care delivery [1-4]. The term mHealth technology refers to the use of smartphones and other mobile devices for personal health monitoring, health care delivery, or research [5]. Expert recommendations from the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute (NHLBI) advocated for using innovative approaches, such as study designs that utilize mHealth technology, to provide new opportunities for population science [6]. Innovative electronic cohort (*e-Cohort*) study designs that incorporate mHealth technology into traditional cohort studies have been proposed, minimizing the requirement of physical resources by collecting data *remotely* (reducing or completely eliminating in-person clinical examinations) [7-10]. In 2015, the NIH funded a national resource to *mobilize research* by creating an infrastructure for conducting research using mHealth technology and has recently initiated the All of Us Research Program (formerly called the Precision Medicine Initiative) [11]. The All of Us program is a large, national study, with the goal of recruiting 1 million participants, which differs from other national cohorts such as the United Kingdom Biobank Study [12], by allowing for electronic (remote) enrollment. Successful recruitment in previous e-Cohort studies such as Health eHeart (HeH) Study and MyHeart Counts, which do not require on-site visits [13,14], have paved the way for new, large e-Cohorts such as All of Us.

The e-Cohort approach may provide a cost-effective methodology to remotely collect population-level data outside of standard research clinic settings, using mHealth devices and internet-based questionnaires [7-10], but may introduce substantial selection bias beyond that of typical research studies [13,15]. Investigators from HeH reported that HeH participants are more likely to be female, white/non-Hispanic, college-educated, nonsmokers, in excellent general health, but are also more likely to have cardiovascular disease and risk factors, compared with a national research study with more traditional recruitment practices [13]. Moreover, the level of technical support that may be required by participants for mHealth device data collection is unclear, especially with regard to middle-aged and older adults who may have less familiarity and require more support with mHealth technology [16]. Finally, despite several theoretical advantages of merging these newer remote studies (lacking on-site visits) with established conventional cohorts, this practice has not yet been carefully studied [6].

Objectives

We conducted a 5-month pilot study in the well-characterized Framingham Heart Study (FHS) cohort to test the feasibility of incorporating mHealth technology in a long-standing epidemiologic cohort study using remote versus in-person device set up. Our approach to pilot test and scale up the use of mHealth technology and electronic surveys (e-surveys) within FHS [17] leveraged the committed study participants and infrastructure of FHS. For the pilot study, we partnered with the HeH Study, which had an established protocol and infrastructure for collecting mHealth data.

The main purpose of our FHS-HeH pilot study was to assess whether remote mHealth data collection supported by email was equivalent to a strategy that involved in-person support *on-site* at the FHS Research Center by measuring the rates of mHealth device set up and continued use over the 5-month study. In addition to testing the feasibility and optimal data collection strategy, we also assessed the clinical characteristics of enrolled versus declined participants, completion rates of internet-based self-report data, and study design acceptability among participants.

Methods

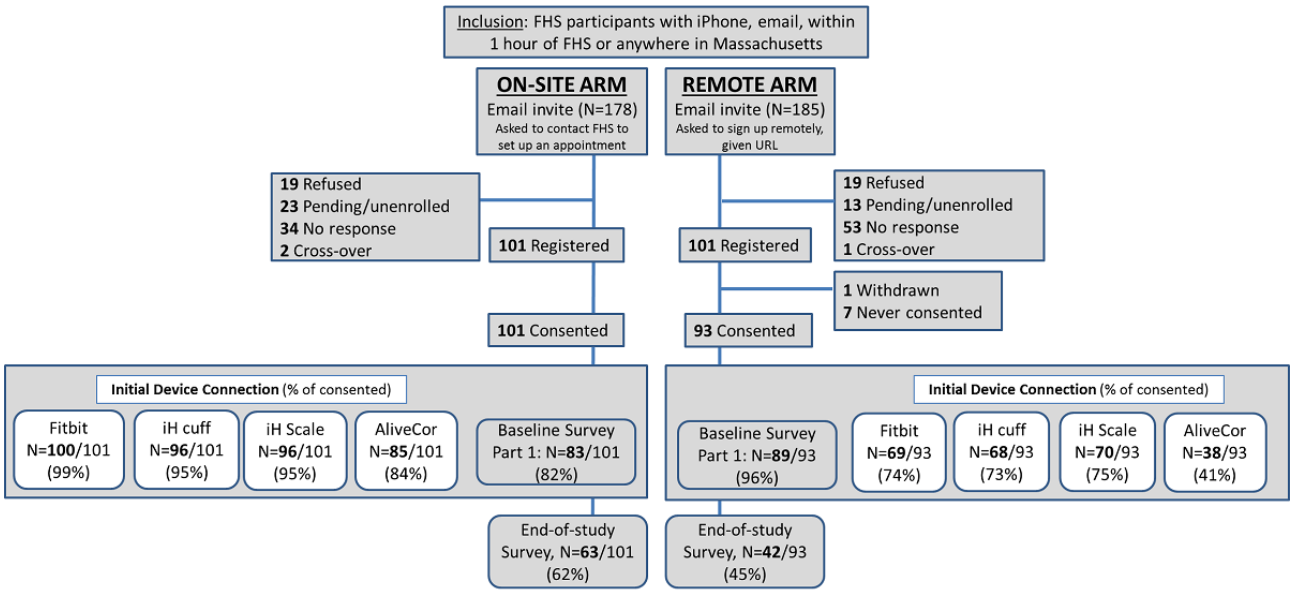
Study Design

The FHS began enrolling participants for the Original cohort in 1948 [18]. In 1971, the offspring of the Original cohort and the spouses of these offspring were enrolled in the Offspring study [19]. In 1994 and 2002, ethnic/racial minority Omni cohorts were recruited to increase the diversity represented in FHS to better reflect the contemporary diversity of the town of Framingham, Massachusetts. In addition, in 2002, Third Generation participants were recruited from a sample of individuals that had at least 1 parent in the Offspring cohort [20]. These participants have been followed at 2- to 8-year intervals in the subsequent years and the study is ongoing. The most recent Offspring examination (including Omni cohort 1) occurred between 2011 and 2014 and the last Third Generation (including Omni cohort 2) examination was conducted during

2008 to 2011. Previous FHS examinations primarily used phone calls to recruit participants to return to the FHS Research Center.

FHS Offspring, Third Generation, and Omni participants [19,20] who had an email address, lived within a 1-hour drive of the FHS Research Center and owned an iPhone [21] were eligible for participation in this investigation. The iPhone requirement was included as, at the time, not all devices were supported by Android. A previous report from FHS, Framingham Digital Connectedness Survey, permitted us to identify participants reporting iPhone ownership and internet use for recruitment purposes [21]. During the recruitment (May-October 2015), 363 participants were sent an email invitation Figure 1. Our goal was to recruit 100 participants in each of the study arms (*remote* vs *on-site* support) and to sample at least 20% older participants (age ≥65 years). Our study protocol followed Zelen design [22], in which participants were randomized to one of the following 2 groups before invitations were sent and consent was obtained:

Figure 1. Flow chart of recruitment and initial device connection for the Framingham Heart Study–Health eHeart pilot study. Pending/unenrolled participants responded to the initial email invitation, but they did not respond to further communications. FHS: Framingham Heart Study; iH: iHealth; AliveCor: electrocardiogram device.



- Remote support: Participants randomized to the remote support group received an email invitation with an explanation about the FHS-HeH pilot study and a URL they could follow to learn more and register for the study (first figure, Multimedia Appendix 1). For those who did not register within 1 week of the initial email, a second email was sent. After a second week of no response, a phone call was placed to their home. No more than 3 phone calls were placed to any individual for recruitment purposes.
- On-site support: Participants randomized to the on-site support group were contacted by the same email/phone call protocol to register for the study and set up a study visit (second figure, Multimedia Appendix 1). Trained FHS staff members assisted the participants in-person to register with the FHS-HeH pilot study, sign the Web-based consent, and connect the devices to their iPhones and the study website. If requested, participants were able to return to the FHS Center if they required additional in-person support.

After the study termination (March 2016), all participants were emailed an *end-of-study survey*, through an internet link, to assess the participant burden and the overall FHS-HeH experience. The survey went out after 98% of the participants had completed the 5-month study (4 participants had not yet completed 5 months). The FHS-HeH study was approved by the Institutional Review Board (IRB) at the University of California, San Francisco, and the participants provided written informed consent. The Boston University Medical Center had an approved IRB authorization agreement.

Covariates

The following demographic information was collected from the most recent FHS examination attended: age, sex, body mass index (BMI), physical activity index [23], history of smoking (defined as former or current smokers, having at least 1 cigarette per day in the past year), hyperlipidemia (total cholesterol ≥200 mg/dL or being on lipid treatment), education, diabetes mellitus (defined as fasting glucose ≥126 mg/dL or treatment with hypoglycemic agent or insulin), hypertension (defined as systolic BP ≥140 mm Hg or diastolic BP ≥90 mm Hg or being on treatment), atrial fibrillation, and cardiovascular disease (includes myocardial infarction, coronary insufficiency, atherothrombotic brain infarct, transient ischemic attack, intermittent claudication, and heart failure). Participants with missing demographic data (detailed in the Results section) either did not attend their last FHS examination cycle or did not complete that part of the examination. Participants with missing covariate data were included in all tables.

Statistical Analysis

Demographic information was reported as mean (SD) for each study arm and for FHS participants who declined to participate in this investigation. Study adherence was defined conservatively as simply taking 1 measurement each month to

get a broad assessment of continued device use. Study adherence and survey responses were compared between the 2 study arms in the total study sample by calculating the mean percent differences and 95% CIs. All statistical analyses were performed by using SAS, version 8 (SAS Institute Inc). Significant differences were reported at the $P<.05$ level.

Results

Study Enrollment

Of the 363 participants invited, 87 participants did not respond to the initial recruitment efforts, 38 declined to participate, and 36 communicated an intent to participate but did not follow through with enrollment (Figure 1). There were 101 participants who completed enrollment in each of the randomized study arms (n=202 total). Owing to the 2 early withdrawals (1 withdrawal in each study arm), additional participants were allowed to enroll to replace these withdrawals. In the *on-site* arm, there was a study technician available to answer questions and we observed 100% completion of the consent process. In contrast, individuals in the *remote* arm were emailed a link to initiate the consent process; only 93/101 (92%) completed the consent. In total, 82 participants responded to the invitation but did not complete the consent (38 participants declined, 36 were pending/not enrolled, and 8 enrolled but did not complete consent). Consenting participants were more likely to be women, tended to be younger, were less likely to smoke or have diabetes mellitus, and were more likely to have attended at least some college (Table 1). The rates of missing demographic data from Table 1 were low (BMI, missing [m]=11; physical activity index, m=13; history of smoking, m=3; hyperlipidemia, m=11; education, m=5; diabetes mellitus, m=15; and hypertension, m=11). Missing data were because of either missing the most recent FHS examination or missing the questionnaire/biomarker data at the most recent examination. None of the participants missing diabetes mellitus data had a diagnosis of diabetes mellitus on FHS examinations that occurred before the most recent FHS examination.

Table 1. Demographic information from study participants collected at their last Framingham Heart Study examination.

Demographics	Consented to study (n=194)		Responded to invitation, but not consented[a] (n=82)	*P* value for difference between consented and not consented[b]
	Randomized to on-site arm (n=101)	Randomized to remote arm (n=93)		
Age (years), mean (SD)	55 (11)	53 (10)	58 (12)	.009
Women, n (%)	60 (59)	57 (61)	38 (46)	.04
Cohort, n (%)				
Offspring	19 (19)	12 (13)	30 (37)	—[c]
Third Generation	76 (75)	75 (81)	49 (60)	—
Omni 1	2 (2)	—	—	—
Omni 2	4 (4)	6 (6)	3 (4)	—
Education, n (%)				
Less than high school	—	—	—	—
High school	6 (6)	3 (3)	14 (17)	—
Some college	10 (10)	19 (20)	17 (21)	—
College and higher	85 (84)	71 (76)	51 (62)	—
Body mass index (kg/m^2), mean (SD)	27 (5)	29 (6)	28 (6)	.48
Physical Activity Index, mean (SD)	35 (5)	35 (7)	36 (5)	.26
History of smoking, n (%)	15 (15)	21 (23)	29 (35)	.002
Hyperlipidemia, n (%)	47 (46)	47 (53)	40 (50)	.99
Diabetes mellitus, n (%)	5 (5)	2 (2)	8 (10)	.09
Hypertension, n (%)	20 (20)	22 (25)	19 (24)	.87
Cardiovascular disease, n (%)	2 (2)	7 (8)	4 (5)	.99
Atrial fibrillation, n (%)	2 (2)	1 (1)	1 (1)	.99

[a]The *not consented* column includes 38 participants who declined, 36 pending/not enrolled, and 8 enrolled but did not complete consent.

[b]*P* values were not calculated for differences in cohort and education because of low numbers in some groups.

[c]Not applicable.

Importantly, recruitment of the older adults (age ≥65 years) for this e-Cohort study was less efficient (50% of individuals consented, 27 out of the 54 individuals who responded to the email invitation to participate) compared with the recruitment of adults aged <65 years (75% consented, 167 out of the 222 individuals who responded to the email invitation), as calculated

from Table 1 and the first table of Multimedia Appendix 1. Older adults choosing to participate in our study had completed more education (100% completing at least some college) than those choosing not to participate, of which 26% (n=7/27) had not continued on to college after high school.

Device Use

In the *on-site* arm, 99% of the consenting participants (100/101) initially connected to the Fitbit device, 95% (96/101) to the iHealth BP cuff and scale, and 84% (85/101) to the AliveCor ECG. As for the *remote* arm, 74% of those that consented initially (69/93) connected to the Fitbit device, 73% (n=68/93) to the iHealth BP cuff, 75% (70/93) to the iHealth scale, and 41% (38/93) to the AliveCor ECG (Figure 1 and Table 2). The *on-site* arm had 20% to 43% more participants initially connected to the devices at baseline (mean percent difference was 25% [95% CI 17-35] for activity monitor, 22% [95% CI 12-32] for BP cuff, 20% [95% CI 10-30] for scale, and 43% [95% CI 30-55] for ECG).

After the initial connection, the proportion of participants that continued to use the devices declined consistently in both arms of the study (Table 3 and Figure 2). Although 4 study participants in the *on-site* arm did not have the opportunity to participate in the full 5-month study, removal of these participants in sensitivity analyses did not change the results considerably (second and third table of Multimedia Appendix 1).

Table 2. Primary analysis: Rate of device connection at baseline and continued use at 5 months.

Device	On-site (n=101), n (% consent)		Remote (n=93), n (% consent)		Difference in proportion of device connection rate between study arms	
	Baseline connection	Fifth month device use[a]	Baseline connection	Fifth month device use	Mean percent difference between study arms in baseline connection rate (95% CI)	Mean percent difference between study arms in fifth month device use rate (95% CI)
Fitbit device	100 (99)	79 (78)	69 (74)	54 (58)	25 (17 to 35)	20 (7 to 33)
iHealth blood pressure cuff	96 (95)	54 (53)	68 (73)	40 (43)	22 (12 to 32)	10 (−4 to 24)
iHealth scale	96 (95)	57 (56)	70 (75)	40 (43)	20 (10 to 30)	13 (−1 to 27)
AliveCor	85 (84)	54 (53)	38 (41)	33 (35)	43 (30 to 55)	18 (4 to 31)

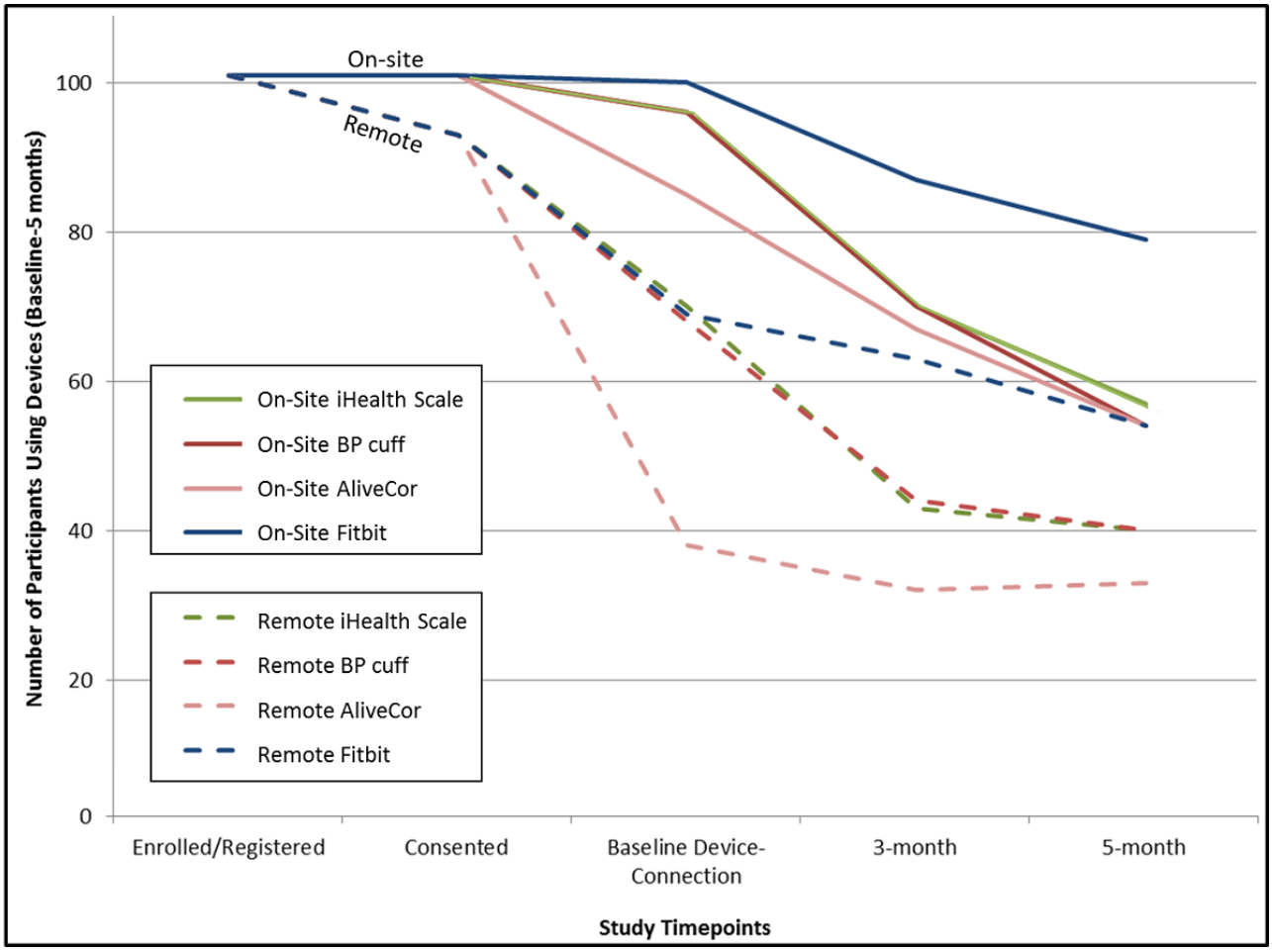
[a]A total of 4 participants in the *on-site* arm did not have the opportunity to participate for the full 5 months owing to study termination.

Table 3. Secondary analysis: Continued use of devices for participants who were initially able to connect to the devices during the first month. The n (%) values are given with regard to baseline device connection.

Device	On-site (N=101)			Remote (N=93)			Difference in proportion of continued device use between study arms	
	Baseline connection, n	Third month device use, n (% baseline)	Fifth month device use, n (% baseline)[a]	Baseline connection, n	Third month device use, n (% baseline)	Fifth month device use, n (% baseline)	Mean percentage difference between study arms in baseline connection rate (95% CI)	Mean percentage difference between study arms in fifth month device use rate (95% CI)
Fitbit device	100	87 (86)	79 (79)	69	63 (91)	54 (78)	−4 (−14 to 6)	1 (−12 to 14)
iHealth blood pressure cuff	96	70 (69)	54 (56)	68	44 (65)	40 (59)	8 (−6 to 23)	−3 (−18 to 13)
iHealth scale	96	70 (69)	57 (59)	70	43 (61)	40 (57)	11 (−3 to 26)	2 (−13 to 17)
AliveCor	85	67 (66)	54 (64)	38	32 (84)	33 (87)	−5 (−9 to 11)	−23 (−37 to −6)

[a]A total of 4 participants in the *on-site* arm did not have the opportunity to participate for the full 5 months owing to study termination.

Figure 2. Number of participants using devices throughout the study from study enrollment through the 5-month follow-up period. BP: blood pressure.



Survey Data

All consenting participants were sent links to participate in the 2 internet-based surveys: a baseline core survey and an *end-of-study survey* after the study termination. The baseline core survey comprised 34 separate parts assessing self-reported health outcomes that could be completed in any order and was well attended by participants in both arms. The first survey was completed by 83 (82%) participants from the *on-site* arm and 89 (96%) participants from the *remote* arm (Figure 1.) After the study completion, all participants were sent an end-of-study survey, of which only 63% of the *on-site* arm and 45% of the *remote* arm participated (fourth table, Multimedia Appendix 1). Overall, the participants endorsed positive statements about their study participation. At least 95% of the participants in both study arms agreed to the statement, "I would participate in this type of study again in the future." Over 85% of the participants in the *on-site* arm agreed with almost all the survey questions (as demonstrated by the shaded boxes in the fourth table of Multimedia Appendix 1), whereas there was slightly lower agreement for the *remote* arm.

Discussion

Principal Findings

Our FHS-HeH pilot study was conducted in collaboration with the HeH Study to test feasibility of mHealth and digital data collection in FHS participants using remote versus in-person support for device set up and use over a 5-month period. Participants in our *on-site* study arm had the opportunity to visit the FHS Research Center for consent and mobile device set up. We observed that the *on-site* participants were more likely to consent and had better success with initial device connection and use compared with the individuals who received only *remote* support by phone or email. However, once connected to the devices, the rates of continued device use were similar in both groups. Our findings suggest that it is possible to maximize participation by leveraging in-person support for e-Cohort studies. Furthermore, we observed reasonable adherence with mHealth technology by older adults.

In both study arms combined, almost 79% of the participants who successfully initialized the Fitbit device at the beginning of the study continued to use the device for the full 5-month study, representing 69% of the total sample of consenting study participants. We defined *continued use* very conservatively, as 1 measurement per month, to get a broad assessment of continued device use. Preliminary data from a new FHS initiative separate from FHS-HeH, called *eFHS*, reported that 76% (306 of 402 participants given an Apple Watch device) wore the device at least weekly over 3 months and received reminder messages if no data were sent for 14 days [17].

In 2 other studies that recruited participants using *snowball* (social network/internet-based) sampling strategies specifically to enroll participants into e-Cohort studies, surprisingly, the frequency of device use did not appear to be more successful,

and may have even been lower, than in FHS-HeH or eFHS which enrolled from within the ongoing FHS cohort [14,24]. In the MyHeart Counts study, investigators reported that 47% of their >48,000 consented study participants completed just 2 consecutive days of fitness monitor data as measured by a smartphone app in the first week and adherence only declined from there [14]. In the mPower substudy of HeH, a 6-month smartphone-based study, 87% of 9520 study participants completed at least one task on the smartphone app after consenting to the study, but only 9% contributed data on ≥5 separate days, confirming that consistency in device or app use is one of the major challenges of this type of research [24]. Physical activity intervention studies provide additional comparative data, with considerable drop-off in device use over the short term (3-6 months) and over longer periods (6 months to 1 year), especially after the participant incentives are terminated [25-27]. Unfortunately, owing to our study termination after at least 5-months follow-up, we are unable to test whether there would be an effect of device setup strategy (on-site versus remote) on longer follow-up of continued device use. It is also unclear what type of communication, support, or incentives might maximize adherence with mHealth devices. In our study, participants were only sent reminders to sync their devices, briefly, midstudy. Our study was not designed to assess whether these reminders affected device use. However, there is a burgeoning field of study testing communication methods/strategies to increase and sustain health behavior [28-31]. Messaging may need to be tailored to participants based on the current adherence, and investigators should be cautious that the language does not infer that data are not received, unless that is the message meant to be communicated.

Overall use of the BP cuff, scale, and AliveCor ECG were somewhat lower than the continued use of the Fitbit device in our FHS-HeH *on-site* arm, but generally, once connected, the use was similar for both the study arms. Across both arms, 56% to 59% of the participants who successfully connected the BP cuff or scale at baseline, continued to use it through the 5-month study duration. For comparison, in a meta-analysis, rates of adherence to self-monitoring BP in hypertensive patients participating in an intervention to lower BP varied widely by study, but true comparison with our study is difficult as most of our FHS-HeH participants were not hypertensive [32]. In addition, most studies from the meta-analysis used traditional nonconnected BP devices, instead of mHealth devices with smartphone apps.

Device connection to the AliveCor ECG device was lower than other devices. Our technical staff reported that the AliveCor was typically the last device they connected during the in-person visit. Another contributing factor could be the more complex instructions for setting up the AliveCor device, including multiple steps in which the participants were required to log in to their email. Other than these reasons, it is unclear why the connection rates were much lower in the remote arm (41%) compared with the on-site arm (84%).

In contrast to the diminishing rates of the BP cuff or scale use over time, the rate of continued adherence for those that were initially able to connect to the AliveCor ECG remained relatively high at 5 months (especially in the remote arm, 87%). However, enthusiasm about apparently high AliveCor adherence should be tempered by the fact that only a small number of participants connected to this device at baseline. Thus, participants who successfully connected to the AliveCor at baseline may differ from those who connected to other devices. We hypothesize that AliveCor users may be more interested in their health, more motivated study participants, and/or more technologically savvy. However, one limitation to our study is that we did not measure the reasons for differences in device use, so we are not able to determine the facilitators and barriers to the use of specific devices [13,14,16,33,34].

Internet-Based Survey Data Can Be Successfully Administered Via Different Strategies

In addition to answering important questions about device connection and use, our study was able to assess the rate of internet-based survey initiation using our 2 study arms. Until recently, the FHS has conducted most questionnaires in-person and only administered short health history updates in the interim between examinations by phone or using traditional mail via the postal service. Although consent and device connection appeared more successful in the *on-site* arm of the study, the participation in the baseline core survey was higher in the remote arm (82% vs 96% in the *on-site* and *remote* arms, respectively). Therefore, in-person contact may not be an important part of a study designed only to perform surveys with participants. Instead, higher survey participation rates in the *remote* arm may be reflective of the lower burden imposed initially in the *remote* arm before devices were shipped. These results provide some evidence that internet-based surveys may be effective means to conduct a health history questionnaire in FHS participants.

Other e-Cohort studies have had variable success with participant engagement in e-surveys, which may depend on the timing and strategies used to present surveys to participants. In MyHeart Counts, 41% of the study participants completed a cardiovascular health survey, whereas 73% completed a physical activity survey, and only 17% provided race/ethnicity [14]. The HeH Study (with >210,000 participants) reported that 86% of participants completed at least one survey, but 37% provided complete survey data [13]. Another traditional cardiovascular epidemiology cohort, Coronary Artery Risk Development in Young Adults has also explored the electronic administration of surveys through the internet (eCARDIA), reporting 52% survey completion [35]. On the basis of the results from these studies, it may be important to prioritize survey administration in e-Cohort studies to ensure that the most important surveys have strong adherence.

In contrast to the high participation rates in the baseline core survey, the end-of-study survey was not completed as frequently (62% vs 45% of *on-site* and *remote* participants, respectively). Study design and communication with participants are not only important for the baseline connection and use of the device, but also for good adherence to device use at follow-up. These considerations are especially important for longitudinal studies that continue to engage participants over a long follow-up period as poor communication and frustration from participants may impact future participation. On the basis of data from approximately half of the study participants who provided

feedback, approximately 96% of the participants said that they would participate in this type of study again, regardless of the study arm. Although participation bias influences our ability to interpret results from the end-of-study survey, it does appear that the *on-site* participants responded more favorably overall.

Strengths, Bias, and Limitations of Our Study Design

The strengths of our study lie in our study design, which leveraged infrastructure and the strengths of FHS and HeH, including a recruitment sample of committed study participants across middle and older age. Our design not only enabled the examination of different methodologies for incorporating consumer-facing mHealth technology into an epidemiological study, but may also provide insight for other study designs, including clinical trials.

Important limitations to consider include the limited exploration of participation bias by demographic factors other than age. The study was small, and we had limited power to examine subgroup findings. The FHS primarily comprises white individuals residing in New England; therefore, we were unable to analyze how the study design influenced participation by racial/ethnic group or region. Certain demographic groups may be more unlikely to be eligible for participation in mHealth studies, such as those that do not have a smartphone [21]. In our FHS *Digital Connectedness Survey*, administered during 2014 to 2015, we reported that smartphone users in FHS were younger, more highly educated, with less cardiovascular risk factors than individuals without a smartphone [21]. However, even among the participants who were eligible for our study (ie, had an iPhone and email address), those who agreed to participate were more likely to have attended at least some college (95% vs 82% among participants who were eligible but declined to participate) and were less likely to be smokers. Both trends are similar to what was seen in other e-Cohort studies [35,36], including the preliminary HeH recruitment analysis in which participants were less likely to smoke and were more likely to be women, had higher educational attainment, reported excellent general health, and were likely to be white (rather than black, Hispanic, or Asian) when compared with the traditional National Health and Nutrition Examination Survey study design [13]. Although issues of generalizability plague all epidemiological studies, it may be a particular concern in e-Cohort studies.

Previous studies in minority communities in the United States cited concerns and misconceptions by the participants in mHealth studies, such as the type of information that would be tracked by mobile technology, legal risks that might be introduced through participation, a lack of familiarity with certain devices, and unwanted attention from others when wearing or using devices in public [37,38]. These concerns can impact both study participation and adherence and may require cultural sensitivity (or age/generational sensitivity), creativity, and patience from the study team. The study team must weigh cost-effectiveness of potential adaptations, with limiting selection bias and maximizing the equity in research across diverse populations [37,38]. We did not analyze the cost between the study arms, so we are not able to compare the differences in our study. It is possible that personal communications with knowledgeable study coordinators and the research team may help to overcome some of the barriers mentioned above. The introduction of mHealth technology raised some concerns even in FHS participants who are familiar with research studies. We observed a barrier to consent that was somewhat overcome through the *on-site* study design, in which participants spoke with study coordinators who could explain the study, answer questions, and provide in-person support for setting up the mHealth devices. Future studies should assess whether other forms of participant engagement, such as text messaging, will influence mobile device use and study adherence.

We also acknowledge the conservative measure of study adherence (device use once per month) as we were most interested in assessing the overall adherence as a primary study aim. In future studies, it will also be important to understand the barriers preventing study adherence and to investigate the factors contributing to the frequency of use and how to improve these metrics. We acknowledge that providing 4 devices might have been burdensome for some participants, especially as participants needed to visit 3 different consumer-facing websites/apps to create accounts for each device (iHealth, Fitbit, and AliveCor) to connect the devices to the HeH platform, adding complexity to the initial user experience. Using 1 single app to connect multiple devices may improve connection for participants, especially for participants connecting remotely. Another key future step will be testing different methods of supporting and engaging participants, including assessing how participants engage with the website/apps using Web analytics tools. Providing in-person support, as we showed in our FHS-HeH pilot study, has the potential to increase study efficiency and may also minimize participation bias.

Conclusions

Our feasibility study demonstrated that offering *on-site* support for studies involving mHealth technology maximizes participation and initial rates of device use, compared with offering only *remote* support. However, once connected, drop-off rates were similar in both groups. Future studies may find it to be cost-effective to provide in-person support for studies involving mHealth technology for middle-aged and older populations.

Acknowledgments

The FHS is supported by the following grants from the NIH and the US Department of Health and Human Services (HHS): NHLBI-N01-HC25195, HHSN268201500001I, and NHLBI Intramural Research Program. EJB, JMM, and DMM are supported by Robert Wood Johnson award 74624. DDM is supported by R01HL126911, R01HL136660, R15HL121761, and UH2TR000921. EJB is supported by R01HL128914, 2R01 HL092577, 2U54HL120163, American Heart Association, and 18SFRN34110082. JEO, MJP, and GMM are supported by U2CEB021881. LT is supported by 18SFRN34110082. HL is also supported by the Boston University Digital Health Initiative; Boston University Alzheimer's Disease Center Pilot Grant; and the National Center

for Advancing Translational Sciences, NIH, through BU-CTSI grant number 1UL1TR001430. The views expressed in this paper are those of the authors and do not necessarily represent the views of the NIH or the US Department of HHS.

Conflicts of Interest

GMM receives research support from Jawbone Heath. CSF is currently employed by Merck Research Laboratories and owns stock in the company. DDM discloses equity stakes or consulting relationships with Flexcon, Bristol-Myers Squibb, Mobile Sense, ATRIA, Pfizer, Boston Biomedical Associates, Rose consulting group, and research funding from Sanofi Aventis, Flexcon, Otsuka Pharmaceuticals, Philips Healthcare, Biotronik, Bristol-Myers Squibb, and Pfizer. No other authors have relevant disclosures.

Editorial notice: This randomized study was not prospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Supplemental material containing the participant recruitment emails and tables including demographics of participants at least 65 years of age, sensitivity analyses, and "end of study" survey responses.
[PDF File (Adobe PDF File)372 KB - mhealth_v7i9e13238_app1.pdf]

Multimedia Appendix 2
CONSORT‐EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File)2331 KB - mhealth_v7i9e13238_app2.pdf]

References

1. Lauer MS. Time for a creative transformation of epidemiology in the United States. J Am Med Assoc 2012 Nov 7;308(17):1804-1805. [doi: 10.1001/jama.2012.14838] [Medline: 23117782]
2. Sorlie PD, Bild DE, Lauer MS. Cardiovascular epidemiology in a changing world--challenges to investigators and the National Heart, Lung, and Blood Institute. Am J Epidemiol 2012 Apr 1;175(7):597-601 [FREE Full text] [doi: 10.1093/aje/kws138] [Medline: 22415032]
3. Manolio TA, Weis BK, Cowie CC, Hoover RN, Hudson K, Kramer BS, et al. New models for large prospective studies: is there a better way? Am J Epidemiol 2012 May 1;175(9):859-866 [FREE Full text] [doi: 10.1093/aje/kwr453] [Medline: 22411865]
4. Khoury MJ, Lam TK, Ioannidis JP, Hartge P, Spitz MR, Buring JE, et al. Transforming epidemiology for 21st century medicine and public health. Cancer Epidemiol Biomarkers Prev 2013 Apr;22(4):508-516 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0146] [Medline: 23462917]
5. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
6. Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, et al. Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. Am J Epidemiol 2015 Mar 15;181(6):363-368 [FREE Full text] [doi: 10.1093/aje/kwv011] [Medline: 25743324]
7. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
8. Topol EJ, Steinhubl SR, Torkamani A. Digital medical tools and sensors. J Am Med Assoc 2015 Jan 27;313(4):353-354 [FREE Full text] [doi: 10.1001/jama.2014.17125] [Medline: 25626031]
9. Powell AC, Landman AB, Bates DW. In search of a few good apps. J Am Med Assoc 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
10. Kaiser J. Epidemiology. Budget woes threaten long-term heart studies. Science 2013 Aug 16;341(6147):701. [doi: 10.1126/science.341.6147.701] [Medline: 23950503]
11. Sankar PL, Parker LS. The precision medicine initiative's All of Us research program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750. [doi: 10.1038/gim.2016.183] [Medline: 27929525]

12. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
13. Guo X, Vittinghoff E, Olgin JE, Marcus GM, Pletcher MJ. Volunteer participation in the health eheart study: a comparison with the US population. Sci Rep 2017 May 16;7(1):1956 [FREE Full text] [doi: 10.1038/s41598-017-02232-y] [Medline: 28512303]
14. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart Counts cardiovascular health study. JAMA Cardiol 2017 Jan 1;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
15. Wirth KE, Tchetgen EJ. Accounting for selection bias in association studies with complex survey data. Epidemiology 2014 May;25(3):444-453 [FREE Full text] [doi: 10.1097/EDE.0000000000000037] [Medline: 24598413]
16. Anderson M, Perrin A. Pew Research Center. 2017. Tech Adoption Climbs Among Older Adults URL: https://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/
17. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and preliminary findings from a new electronic cohort embedded in the Framingham heart study. J Med Internet Res 2019 Mar 1;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]
18. Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. Circulation 1966 Oct;34(4):553-555. [doi: 10.1161/01.cir.34.4.553] [Medline: 5921755]
19. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979 Sep;110(3):281-290. [doi: 10.1093/oxfordjournals.aje.a112813] [Medline: 474565]
20. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham heart study: design, recruitment, and initial examination. Am J Epidemiol 2007 Jun 1;165(11):1328-1335. [doi: 10.1093/aje/kwm021] [Medline: 17372189]
21. Fox CS, Hwang SJ, Nieto K, Valentino M, Mutalik K, Massaro JM, et al. Digital connectedness in the Framingham heart study. J Am Heart Assoc 2016 Apr 13;5(4):e003193 [FREE Full text] [doi: 10.1161/JAHA.116.003193] [Medline: 27076568]
22. Torgerson DJ, Roland M. What is Zelen's design? Br Med J 1998 Feb 21;316(7131):606 [FREE Full text] [doi: 10.1136/bmj.316.7131.606] [Medline: 9518917]
23. Kannel WB, Belanger A, D'Agostino R, Israel I. Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham study. Am Heart J 1986 Oct;112(4):820-825. [doi: 10.1016/0002-8703(86)90480-1] [Medline: 3766383]
24. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The m-power study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 3;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
25. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
26. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
27. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017 Nov 1;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
28. Adams MA, Hurley JC, Todd M, Bhuiyan N, Jarrett CL, Tucker WJ, et al. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health 2017 Mar 29;17(1):286 [FREE Full text] [doi: 10.1186/s12889-017-4197-8] [Medline: 28356097]
29. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
30. Thakkar J, Kurup R, Laba TL, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
31. 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 9;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]
32. Fletcher BR, Hartmann-Boyce J, Hinton L, McManus RJ. The effect of self-monitoring of blood pressure on medication adherence and lifestyle factors: a systematic review and meta-analysis. Am J Hypertens 2015 Oct;28(10):1209-1221. [doi: 10.1093/ajh/hpv008] [Medline: 25725092]

33. Deloitte US. 2017. Global Mobile Consumer Survey: US Edition. A New Era in Mobile Continues URL: https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html [accessed 2019-07-30]
34. Levine DM, Lipsitz SR, Linder JA. Changes in everyday and digital health technology use among seniors in declining health. J Gerontol A Biol Sci Med Sci 2018 Mar 14;73(4):552-559. [doi: 10.1093/gerona/glx116] [Medline: 28605446]
35. Kershaw KN, Liu K, Goff Jr DC, Lloyd-Jones DM, Rasmussen-Torvik LJ, Reis JP, et al. Description and initial evaluation of incorporating electronic follow-up of study participants in a longstanding multisite cohort study. BMC Med Res Methodol 2016 Sep 23;16(1):125 [FREE Full text] [doi: 10.1186/s12874-016-0226-z] [Medline: 27664124]
36. Andreeva VA, Deschamps V, Salanave B, Castetbon K, Verdot C, Kesse-Guyot E, et al. Comparison of dietary intakes between a large online cohort study (etude nutrinet-santé) and a nationally representative cross-sectional study (etude nationale nutrition santé) in France: addressing the issue of generalizability in e-epidemiology. Am J Epidemiol 2016 Nov 1;184(9):660-669 [FREE Full text] [doi: 10.1093/aje/kww016] [Medline: 27744386]
37. Murray KE, Ermias A, Lung A, Mohamed AS, Ellis BH, Linke S, et al. Culturally adapting a physical activity intervention for Somali women: the need for theory and innovation to promote equity. Transl Behav Med 2017 Mar;7(1):6-15 [FREE Full text] [doi: 10.1007/s13142-016-0436-2] [Medline: 27558245]
38. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]

Abbreviations

BMI: body mass index
BP: blood pressure
ECG: electrocardiogram
e-Cohort: electronic cohort
e-survey: electronic survey
FHS: Framingham Heart Study
HeH: Health eHeart
HHS: Health and Human Services
IRB: Institutional Review Board
mHealth: mobile health
NHLBI: National Heart, Lung, and Blood Institute
NIH: National Institutes of Health

Edited by G Eysenbach; submitted 22.03.19; peer-reviewed by R Sloan, D Kerr; comments to author 27.04.19; revised version received 21.06.19; accepted 19.07.19; published 30.09.19.

Please cite as:
Spartano NL, Lin H, Sun F, Lunetta KL, Trinquart L, Valentino M, Manders ES, Pletcher MJ, Marcus GM, McManus DD, Benjamin EJ, Fox CS, Olgin JE, Murabito JM
Comparison of On-Site Versus Remote Mobile Device Support in the Framingham Heart Study Using the Health eHeart Study for Digital Follow-up: Randomized Pilot Study Set Within an Observational Study Design
JMIR Mhealth Uhealth 2019;7(9):e13238
URL: https://mhealth.jmir.org/2019/9/e13238
doi:10.2196/13238
PMID:31573928

Original Paper

A Smartphone Attention Bias App for Individuals With Addictive Disorders: Feasibility and Acceptability Study

Melvyn Zhang[1], MBBS, MRCPsych; Jiangbo Ying[1], MBBS, MRCPsych; Syidda B Amron[1], BSc (Hons); Zaakira Mahreen[1], BSc (Hons); Guo Song[1], PhD; Daniel S S Fung[2], MBBS, MMED; Helen Elizabeth Smith[3], DM, FFPHM

[1]National Addiction Management Service, Institute of Mental Health, Singapore, Singapore

[2]Department of Developmental Psychiatry, Institute of Mental Health, Singapore, Singapore

[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Melvyn Zhang, MBBS, MRCPsych
National Addiction Management Service
Institute of Mental Health
10 Buangkok Green Medical Park
Singapore, 529747
Singapore
Phone: 65 98556631
Email: melvynzhangweibin@gmail.com

Abstract

Background: Conventional psychology therapies are unable to address automatic biases that result in individuals relapsing into their substance use disorder. Advances in experimental psychology have led to a better understanding of attention and approach biases and methods to modify these biases. Several studies have demonstrated the effectiveness of bias modification among clinical cohorts. The advances in mobile health technologies have allowed remote delivery of these interventions. To date, there is a lack of studies examining bias modification in a substance-using non-Western sample.

Objective: This study was designed to determine the feasibility of an attention bias modification intervention and an attention bias modification smartphone app for the reduction of attention biases among treatment-seeking individuals. The secondary aim is to determine the acceptability of the intervention.

Methods: A feasibility study was conducted among inpatients who were in their rehabilitation phase at the National Addictions Management Service. Participants were to complete a set of baseline questionnaires, and on each day that they are in the study, undertake an attention bias assessment and modification task while completing a visual analogue scale to assess their craving. Feasibility was determined by the acceptance rate of participation and participants' adherence to the interventions. Acceptability was assessed by a perception questionnaire. Descriptive statistical analyses were performed using SPSS version 22. A thematic analysis approach was used in the qualitative synthesis of users' perceptions.

Results: Of the 40 participants invited to participate in the feasibility study, 10 declined, yielding an acceptance rate of 75%. Of the recruited participants, 6 participants were diagnosed with alcohol dependence; 17, with opioid dependence; 2, with cannabis dependence; and 5, with stimulant dependence. In addition, of the final 30 participants, 11 (37%) failed to complete all the planned interventions and 22 (73%) completed the perspective questionnaires; of these 22 participants, 100% rated the app as extremely and very easy, 77% rated it as extremely or very interactive, 54% rated it as extremely or very motivating, and 33% reported a change in their confidence levels.

Conclusions: Our results highlight the feasibility of recruiting participants to undertake attention bias modification interventions. Participants generally accept use of a mobile version of such an intervention. Nevertheless, our acceptability data indicate that there could be improvements in the existing app, and a participatory design approach might be helpful in its future conceptualization.

International Registered Report Identifier (IRRID): RR2-10.2196/11822

(JMIR Mhealth Uhealth 2019;7(9):e15465) doi:10.2196/15465

KEYWORDS

attention bias; cognitive bias; psychiatry; substance abuse; alcohol abuse; opioid abuse; cannabis abuse; addiction; digital health; mhealth

Introduction

According to the United Nations Office on Drugs and Crime, substances like cannabis, opioids, and stimulants are commonly abused [1]. Substance abuse and substance dependence are associated with significant morbidity and mortality. An estimated 190,000 deaths were attributed to substance use in 2015 [1]. The recent statistics from the World Health Organization also highlighted the high prevalence of alcohol use [2], particularly in higher-income countries. In Singapore, the most recent study found that the prevalence of alcohol use disorders was 0.19% and 1.40% among women and men, respectively, and the prevalence of drug use disorders was 0.07% and 0.28% among women and men [3]. Given the prevalence of these disorders, there is clearly a need for effective interventions. Conventionally, treatment and interventions of addictive disorders involve a combination of medications and psychological therapies. For psychological therapies to help individuals achieve abstinence, frequently used therapies include cognitive behavioral therapy, cue-exposure therapy, contingency management, and mindfulness-based relapse prevention. Cognitive behavioral therapy has had an immediate effect size of 0.45 [4]; 40%-50% of the individuals relapsed within a year and 70% relapsed within 3 years [4]. This may be because such therapies mainly address the cognitive control processes, but fail to address the underlying unconscious automatic processes that contribute to an individual's lapse and relapse.

Among individuals with addictive disorders, there are two common biases: attentional bias and approach bias. Attentional biases are automatic, unconscious processes that result in the preferential allocation of attention toward substance-related stimuli [5,6], and approach biases are the automated tendencies for individuals to reach out and approach substance stimuli [7]. These biases have been well-studied with a theoretical underpinning based on the dual-process model, which posits that the chronic administration of the substance leads to enhanced automatic processing of the substance-related cue, with a corresponding inhibition in the normal cognitive control processes [8]. Interventions to modify bias have been extensively evaluated. Ziaee et al [9] reported that the introduction of bias modification reduced attentional biases among individuals maintained on methadone as well as their cravings to use, the dose of methadone, and the number of relapses. Several other studies that have involved participants undergoing detoxification on an addiction treatment unit [10-12] also found that bias modification was effective in reducing attentional biases. More recently, Cristea et al [13] highlighted that bias modification interventions were effective in modifying both attentional and approach biases in individuals with alcohol and tobacco use issues [13].

Although conventional bias modification interventions are confined to a laboratory setting, advances in technologies now enable remote delivery of such interventions. A recent review [14] showed that seven of eight studies reported the effectiveness of a mobile-based cognitive bias modification intervention. These studies were targeting conditions such as insomnia, social anxiety, tobacco use, and alcohol use disorders [14]. Another recent review [15] explored the utilization of technology for retraining of attentional biases in individuals with tobacco use disorder and reported that mobile delivery of bias retraining was effective. Subsequently, the effectiveness of a mobile health approach for methamphetamine use disorder [16] was evaluated, and improvements in cognitive impairments and impulsive control, but not attentional biases, were reported [16].

To date, no study has evaluated bias modification in a substance-using, treatment-seeking, non-Western cohort. Although technologies have been used for the delivery of remote bias modification intervention, the evaluation was limited to alcohol and tobacco disorders and stimulant disorders. Although a previous study reported negative findings in this regard [16], future research should seek to evaluate the effectiveness of mobile apps for modification of attentional biases, given that the training task used in that study was a newly developed task and differs significantly from conventional bias assessment and modification paradigms. In addition, according to the recommendations of the National Institute of Health Research [17], a feasibility study is crucial, as such a study seeks to determine primarily whether a study could be conducted. This is pertinent in our case, as there are no prior studies that examined bias assessment and retraining in a treatment-seeking, non-Western cohort. In addition, feasibility studies are typically limited to the evaluation of important parameters that will be crucial in the design of the main study and do not routinely evaluate the main outcome of interest [17]. Typically, a feasibility study is conducted first, before a pilot, as a pilot study. It is essentially a version of the main study that is run on a small scale to test whether the components of the main study can all work together [17].

Our study aimed to examine the feasibility of a mobile-based attention bias modification intervention among treatment-seeking individuals with alcohol or substance use disorders. If deemed feasible, this will guide further pilot and definitive randomized trials investigating the effectiveness of the mobile intervention. The objectives of the study were to determine (1) the feasibility of participants undertaking a mobile attention bias modification intervention, (2) the feasibility of the mobile intervention for reducing attention biases, and (3) the acceptability of the intervention. The specific research questions were as follows: (1) Will the mobile attention bias modification intervention be feasible and acceptable among individuals with addictive disorders? (2) Is the developed mobile intervention capable of assessing for and reducing attention biases?

Methods

Study Setting and Design

The target population comprised individuals admitted for inpatient medication-assisted detoxification and rehabilitation (total duration of 14 days, with 7 days in detoxification and 7 days in rehabilitation) at the National Addictions Management Service (NAMS), Institute of Mental Health, Singapore. At NAMS, the treatment is entirely voluntarily, which implies that patients and participants could self-discharge at any time. The NAMS inpatient unit has approximately 22 beds, and most of these beds are occupied by patients who are undergoing detoxification. Patients who had completed their detoxification treatment and were in the rehabilitation phase of the program were recruited. The study design is that of a feasibility study, where participants are recruited by means of convenience sampling.

Ethics Approval

This study was approved by the National Healthcare Group's Domain Specific Research Board (reference number: 2018/00316) on May 2, 2018.

Recruitment and Sample Size

Patients were recruited on completion of their inpatient detoxification treatment (7 days) and at the start of day 1 of their rehabilitation treatment. Potential participants were identified by their primary psychiatrist, provided with further information by the study team, and given time to consider participation. Participants who agreed to participate completed an informed consent form, which was signed in the presence of an impartial witness, in accordance with the Human and Biomedical Act regulations. As the study was designed to assess feasibility and acceptability, power computation was not performed. Given the diversity of the disorders, the minimum recruitment target was 30 participants and the maximum was 34 participants.

Inclusion and Exclusion Criteria

Patients were included in the study if they were aged between 21 and 65 years; diagnosed with a primary psychiatric disorder of alcohol, opioid, cannabis, stimulants, or polysubstance dependence; diagnosed with polysubstance dependence, with alcohol, opioid, cannabis, or stimulants as the main substance of use; able to read and write in English; and capable of using a smartphone or tablet device.

Patients were excluded from the study if they had a known history of cognitive impairment or dementia, a history of seizures or a prior history of withdrawal seizures, a history of migraines triggered by flashing lights, and moderate to severe comorbid psychiatric disorders based on clinical assessment.

Measures

Baseline demographic and clinical information was collected from the participants. This included information about nationality, gender, marital status, race, religion, highest level of education, housing conditions, current substance use, method of consumption of substance, quantity of substance consumed each time, frequency of use, previous treatment history, chronic diseases (psychiatric or physical disorders), and current psychiatric medications. Participants also completed a modified Addiction Severity Index (ASI)-Lite, Severity of Drug Dependence Scale (SDS), and the Short Form (SF-12) questionnaires.

The ASI-Lite collated information for the following domains: drug and alcohol use, medical, employment/school, legal, family, and social and psychiatric aspects [18]. In our modified version, we retained only the drug and alcohol use questions. Participants were asked about their alcohol and substance use in the last 30 days, last month, and lifetime. Participants were asked whether they had used alcohol, nicotine, heroin, amphetamine-type stimulants, cannabis, other opioids, benzodiazepines and other sedatives, barbiturates, ketamine, cocaine, inhalants, hallucinogens, and new psychoactive substances. The SDS comprised five items, all of which are explicitly concerned with the psychological components of dependence [19]. A previous study [20] reported that the total severity score is highly positively correlated with the severity of dependence, as measured by the Diagnostic and Statistical Manual of Mental Disorders-IV. The SF-12 has been widely used in the assessment of the self-reported quality of life. It only covers the eight health domains from the original SF-36 [21] and has demonstrated good content and criterion validity without any evidence of any systematic biases [22].

Intervention

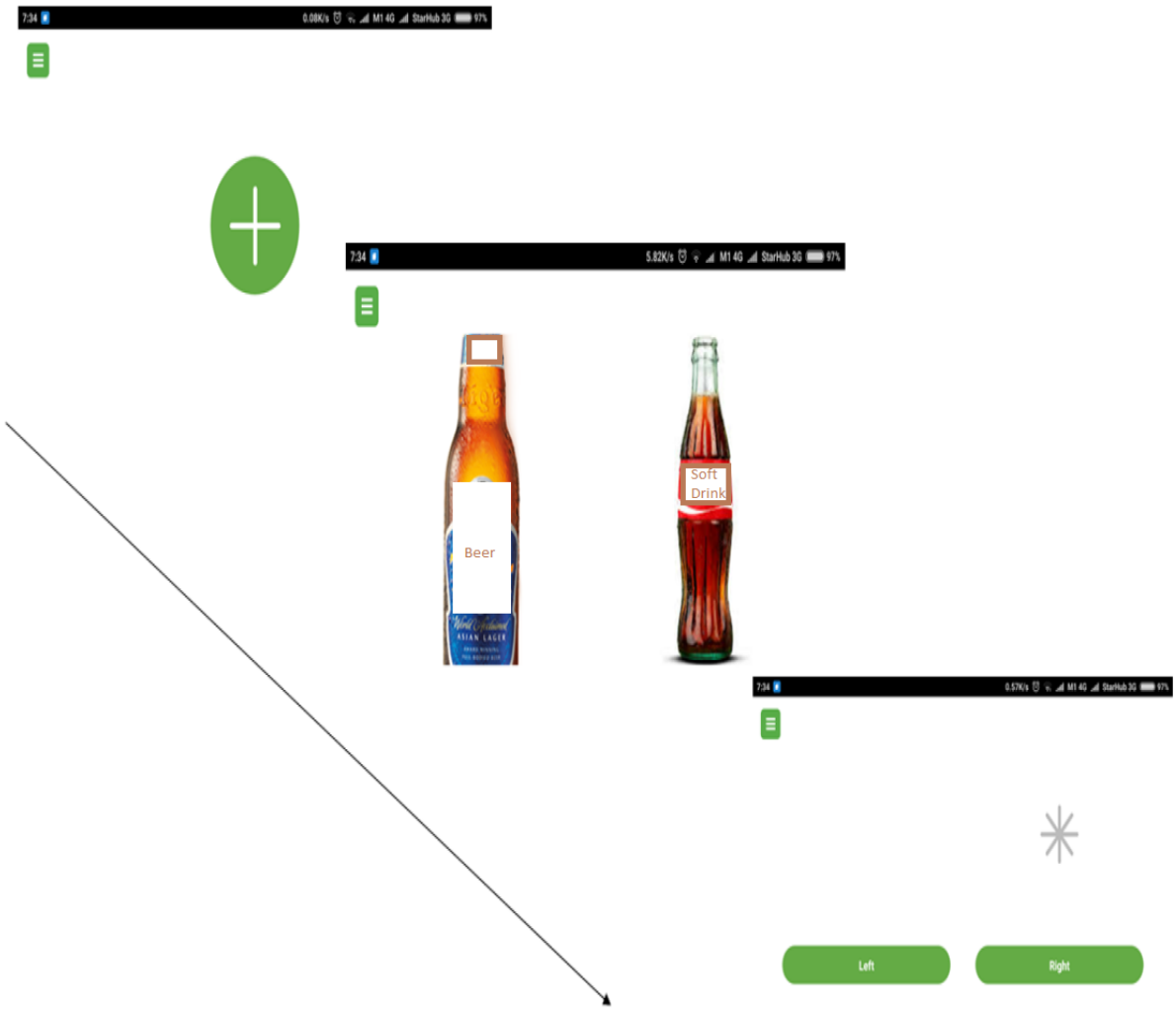
Following completion of enrollment, participants were required to complete a visual-analogue scale for craving before and after the completion of each session. Members of the study team familiar with the app gave participants a 15-minute briefing on the use of the mobile app before the commencement of the assessment and intervention. The study team provided participants tablets to use the mobile attention bias modification intervention.

On day 1 of the intervention, participants were required to complete both a baseline attention bias assessment task and an attention bias modification task. They could rest for 15 minutes before they completed a reassessment of their attention bias. On the subsequent days (days 2-7) of their rehabilitation, they completed the attention bias modification task and were allowed 10 minutes of rest before retaking an attention bias assessment task. Participants were required to complete the visual-analogue scale for craving before and after completion of each bias modification task. Participants who completed three sessions were asked to complete the app perception questionnaire. Participants were expected to undertake the intervention on each day of their rehabilitation stay, except for weekends and public holidays. Participants were allowed to undertake the intervention a maximum of five times in total.

The mobile version of the visual probe task was the same as the original visual probe task. In the attention bias assessment task, participants were required to complete a total of 200 trials (with 10 sets of images repeated 20 times). In each trial, participants were presented with a fixation cross in the center of the screen for 500 milliseconds. Subsequently, they were presented with a set of two images for another 500 milliseconds. In each set of

images, one of the images was neutral but closely related to the image of the alcohol or drug (for example, an image of a man drinking from a can of beer, which was paired with an image of a man drinking from a soft drink can). Following the disappearance of the images, an asterisk replaced the position of one of the images (either on the right or left). The participants were required to indicate where the position of the asterisk was by selecting the physical on-screen buttons as fast as they could. The next set of images was presented once the participant has indicated a response (by pressing the left or right button, depending on where the asterisk was) or if the time of 2000 milliseconds had lapsed (Figure 1). In the assessment phase, 50% of the time, the asterisk replaced the neutral image and 50% of the time, the asterisk replaced the alcohol or substance image. For the intervention or bias modification task, the participant was required to take the same task as that described, but the asterisk replaced the position of the neutral image 100% of the time, enabling retraining of attentional bias. The substance images presented to participants are either a picture of the drugs, pictures of individuals using substances, or paraphernalia used for the administration of the drugs.

Figure 1. An overview of the task that participants undertake on the smartphone/tablet device.



Outcomes

Feasibility was the primary outcome and defined by the number of participants recruited and participants' adherence to the intervention. The study was considered feasible if 25% of the recruitment target (of 30 participants) was met and 60% of the patients managed to adhere to the planned interventions (ie, completed all the planned interventions up until day 5 of their program).

The secondary outcome of acceptability was assessed through a perception questionnaire, which included the following questions:

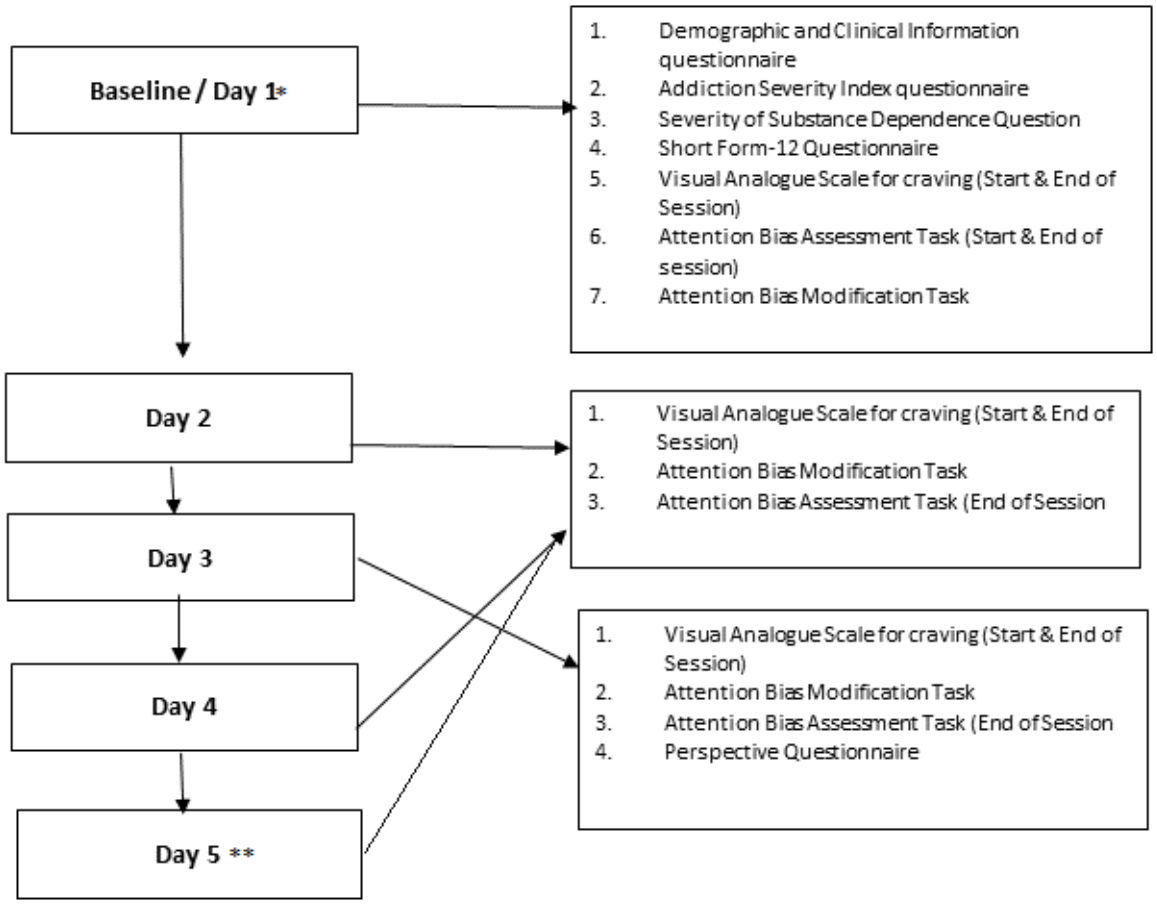
1. Prior to using the app, how confident are you in managing your addiction problems? (5-point Likert scale, ranging from not at all to extremely)
2. How easy was it to use the app? (5-point Likert scale, ranging from not at all to extremely)
3. How interactive was the app? (5-point Likert scale, ranging from not at all to extremely)
4. Do you feel motivated to continue using the app? (5-point Likert scale, ranging from not at all to extremely)
5. Do the images in the app remind you of your substance use? (5-point Likert scale ranging from not at all to extremely)
6. After using the app, how confident are you in managing your addiction problem? (5-point Likert scale, ranging from not at all to extremely)

For questions 2, 3, and 4, respondents were also asked to provide free-text comments.

Acceptability was predefined as a willingness to use the app daily, and if at least 30% of the participants rated ease of use, interactivity, motivation, and reality (questions 2, 3, 4, and 5) positively (either very or extremely on the 5-point Likert scale), and if at least 30% of the participants perceived there to be a change in their confidence level after receiving three sessions of the intervention task (questions 1 and 6). The app was also deemed acceptable by the absence of any severe adverse events (such as intense cravings leading to premature discharge from the inpatient program).

Figure 2 provides an overview of the task that participants undertake each day they are in the study.

Figure 2. Overview of the outcomes measures that participants need to complete for each session. *Attention bias modification assessment task will be completed twice on the first day. The first assessment will provide information pertaining to the baseline attentional biases. The second assessment will assess for the change in attentional biases following the first intervention. **Participants will undertake a maximum of 5 sessions, taking into consideration that the study will not be conducted on weekends.



Data Management and Monitoring

All participants were allocated a subject number upon recruitment, and no participant-related identifiers were captured on the hard-copy forms. These forms, together with the questionnaires, were stored in secured, locked cabinets in a restricted area. The electronic data from the smartphone app was automatically synchronized onto a secured, password-protected cloud database. The main investigator backed-up a copy of the electronic data records onto a local secured computer daily. The principal investigator and the research assistants took the responsibility of coding the data from the hard-copy forms. An independent coinvestigator routinely checked the data entry for accuracy, ensuring that the translation of scores from the hard-copy forms to the electronic

form was free of errors. All records will be kept securely for at least 6 years after completion of the study.

Statistical Analyses

Data collated was analyzed using SPSS (version 22. IBM Corp, Armonk, NY). Baseline demographic information of the subjects was summarized using descriptive statistics, including means and SD. The presence of attentional biases was determined based on the mean reaction times taken to respond to the position of the probes that replace drug or neutral stimuli. The formula used for the computation of attentional biases was $(\sum T_1/n_1) - (\sum T_0/n_0)$, where T_1 refers to the time for probes that replaced the neutral stimulus, n_1 refers to the number of trials for probes that replaced the neutral stimulus, T_0 refers to the time for probes that replaced the substance stimulus, and N_0 refers to the number of trials for probes that replaced the substance stimulus.

Qualitative Analysis of Acceptability Data

Patients' perspective and feedback were collated by means of the perspective questions. Two separate independent researchers coded their verbatim, handwritten comments using NVivo, version 12.0 (QSR International, London, United Kingdom). Similar codes were grouped together and further analyzed, giving rise to higher-order themes.

Results

Feasibility of Recruitment and Adherence

Of the 40 participants invited to participate in the feasibility study, 10 declined, yielding an acceptance rate of 75%. Of the recruited participants, 6 participants were diagnosed with alcohol dependence; 17, with opioid dependence; 2, with cannabis dependence; and 5, with stimulant dependence. In addition, 11 participants of the 30 participants failed to complete all the planned interventions. The adherence rate was thus 63%. For 10 participants, discontinuation was linked with them electing for premature discharge from the ward, and another participant withdrew from the study after the initial intervention. Table 1 provides an overview of the baseline demographic characteristics of the 30 participants recruited.

The mean age of the participants with alcohol and opioid dependence was 43.7 (SD 11.6) years and 47.9 (SD 11.8) years, respectively, and that for participants with stimulant dependence and cannabis dependence was 37.6 (SD 7.0) and 58.0 (SD 1.4), respectively. Most of the participants were Singaporean (90%), and most were of male gender (86%). In addition, 53% had a secondary school education, 76% were unemployed, and 20% of the participants reported being homeless. Furthermore, 50% of the participants with alcohol dependence and 50% of the participants with cannabis dependence had comorbid medical conditions. Moreover, 60% of the participants with stimulant dependence reported having an underlying psychiatric disorder. Participants with alcohol dependence had a mean score of 11.2 (SD 1.9) on the severity of substance dependence questionnaire; those with opioid dependence, stimulant dependence, and cannabis dependence had mean scores of 11.7 (SD 2.2), 9.0 (SD 5.7), and 8.8 (SD 4.5), respectively. These scores demonstrated that participants sampled had a psychological dependence on the substances they were using. The physical health and mental health composite scores were lower for individuals with alcohol use disorders as compared to those with the other disorders.

Table 2 provides the mean attention bias scores for each participant across the trials. Based on the protocol, participants were expected to complete a total of five training sessions. However, not all participants have completed a total of five sessions, as some participants had a public holiday during their stay. Of the 30 participants, 14 participants had positive attentional biases at baseline, whereas the other 16 participants did not have any underlying baseline attentional biases. For those with baseline attentional biases, there was a general decrease in the attention bias scores from baseline to the end of the planned intervention trials. The changes in the scores ranged from 12.0 to 409.5 milliseconds, comparing the final attention bias scores (upon the completion of the intervention) with the baseline scores (at the start of the intervention).

Table 1. Baseline demographic characteristics of participants (n=30).

Demographic characteristics	Alcohol dependence (n=6)	Opioid dependence (n=17)	Cannabis dependence (n=2)	Stimulant dependence (n=5)
Age (years), mean (SD)	43.7 (11.64)	47.9 (11.8)	58.0 (1.4)	37.6 (7.0)
Nationality, n (%)				
Singaporean	4 (66.7)	16 (94.1)	2 (100)	5 (100)
Others	2 (33.3)	1 (5.9)	0 (0)	0 (0)
Gender, n (%)				
Male	4 (66.7)	15 (88.2)	2 (100)	5 (100)
Female	2 (33.3)	2 (11.8)	0 (0)	0 (0)
Race, n (%)				
Chinese	2 (33.3)	2 (13.3)	0 (0)	3 (60.0)
Malay	1 (16.7)	11 (64.7)	0 (0)	0 (0)
Indian	3 (50.0)	3 (17.6)	1 (50.0)	2 (40.0)
Others	0 (0)	1 (5.9)	1 (50.0)	0 (0)
Religion, n (%)				
Christianity	2 (33.3)	4 (23.5)	0 (0)	2 (40)
Hinduism	1 (16.7)	0 (0)	0 (0)	2 (40)
Muslim	2 (33.3)	12 (70.6)	2 (100)	0 (0)
Others	1 (16.7)	1 (5.9)	0 (0)	1 (20)
Education, n (%)				
Primary education	2 (33.3)	3 (17.6)	1 (50.0)	1 (20.0)
Secondary education	2 (33.3)	11 (64.7)	1 (50.0)	2 (40.0)
Junior college/polytechnic/technical studies	1 (16.7)	3 (17.6)	0 (0)	1 (20.0)
Undergraduate studies	1 (16.7)	0 (0)	0 (0)	1 (20.0)
Employment, n (%)				
Unemployed	6 (100)	12 (70.6)	1 (50.00)	4 (80.0)
Part-time employment	0 (0)	1 (5.9)	1 (50.0)	1 (20.0)
Full-time employment	0 (0)	4 (23.5)	0 (0)	0 (0)
Housing, n (%)				
Homeless	1 (16.7)	3 (17.6)	0 (0)	2 (40.0)
1 room	3 (50.0)	5 (29.4)	0 (0)	0 (0)
2 rooms	0 (0)	2 (11.8)	0 (0)	0 (0)
3 rooms	1 (16.7)	1 (5.9)	0 (0)	1 (20)
4 rooms	1 (16.7)	3 (17.6)	1 (50.0)	2 (40.0)
5 rooms	0 (0)	2 (11.8)	0 (0)	0 (0)
Others	0 (0)	1 (5.9)	1 (50.0)	0 (0)
Presence of other chronic diseases, n (%)	3 (50.0)	7 (41.2)	1 (50.0)	1 (20.0)
Presence of psychiatric disorder, n (%)	1 (16.7)	0 (0)	0 (0)	2 (40)
Severity of Substance Dependence scores, mean (SD)	11.2 (1.9)	11.7 (2.2)	9.0 (5.7)	8.8 (4.5)
Short Form-12 questionnaire scores, mean (SD)				
Physical health composite scores	40.5 (24.5)	54.7 (21.6)	82.0 (15.6)	70.2 (21.2)
Mental health composite scores	29.3 (11.1)	41.0 (17.3)	52.0 (0)	39.2 (24.3)
Total scores	33.7 (9.2)	46.5 (16.3)	64.5 (6.4)	51.4 (15.2)

Table 2. Change in attentional bias scores.

Participant	Drug	Baseline		Session 1		Session 2		Session 3		Session 4		Session 5		Overall change in attentional bias
		Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	
1	Stimulants	30.3	96:100	70.6	99:101	36.3	98:102	9.3	98:102	–23.6	96:98	13.3	99:99	17
2[a]	Stimulants	–22.4	103:97	–23.4	97:103	–11.7	95:105	N/A[b]	N/A	N/A	N/A	N/A	N/A	10.7 (increased)
3	Stimulants	6.7	100:100	–3.6	98:101	–28.9	99:99	–11.3	100:99	4.1	100:99	–7.3	100:99	14
4[c]	Opioid	32.1	96:93	28.7	99:100	12.2	99:100	31.2	100:99	20.1	100:99	N/A	N/A	12
5[c]	Alcohol	91.2	97:98	–23.3	100:99	–37.4	100:99	–5.3	99:100	–33.2	99:100	N/A	N/A	124.4
6[c]	Opioid	98.9	28:116	33.4	60:46:00	14.5	71:66	–9.7	88:86	–31.5	86:84	N/A	N/A	130.4
7	Stimulants	–30.5	100:100	–13.5	100:99	–23.4	98:99	–27.7	100:99	–28.2	98:99	–7.6	96:99	22.9 (increased)
8[a]	Opioids	58.7	36:56:00	85.1	40:57:00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	26.4 (increased)
9[a]	Opioids	25.8	97:97	13.5	96:98	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	12.3
10	Cannabis	–9.9	97:99	–20.7	60:40:00	–54.6	93:90	–14.1	96:95	–22.9	99:100	–48.4	100:99	38.5
11	Opioids	–30.9	33:66	2.4	97:97	–15.4	99:99	–7.4	97:99	14.2	99:99	7.3	98:99	38.2 (increased)
12[a]	Opioids	0.7	99:97	34.8	100:99	–15.3	98:99	–52.4	100:98	N/A	N/A	N/A	N/A	53.1
13	Alcohol[d]	–20.9	99:98	–12.3	100:98	N/A	N/A	N/A	N/A	–75.8	90:92	–48.6	99:100	27.7
14[a]	Alcohol	397.7	26:94	44.2	97:98	45.3	99:99	–32.0	98:100	–11.7	99:100			409.4
15	Cannabis	–7.7	100:96	–40.5	100:99	8.6	99:100	–33.6	99:99	–31.8	99:100	–50.3	100:99	42.6
16[a]	Opioids	–27.4	93:96	N/A[d]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
17	Opioids	–42.5	49:49:00	–64.8	75:71	63.4	63:61	8.9	73:78	–15.5	78:82	26.7	96:92	69.2 (increased)
18	Opioids	27.9	92:92	–17.8	99:96	3.2	100:99	–22.6	97:97	–104.6	99:99	–9.2	97:95	37.1
19[a]	Opioids	3.8	99:96	35.1	99:98	13.7	99:98	32.4	100:98	N/A	N/A	N/A	N/A	28.6 (increased)
20	Opioids	10.1	99:98	9.4	100:98	105.2	99:98	54.1	100:99	–1.7	100:99	20.3	99:100	10.2 (increased)
21	Alcohol	224.5	79:54:00	61.5	98:97	73.3	99:100	176.7	94:97	130.3	98:100	107	99:97	117.4
22	Opioids	–52.9	100:99	10.9	99:100	5.6	98:98	39.1	99:99	36.5	98:100	76.6	99:100	129.6 (increased)
23	Opioids	–36.4	96:97	18	95:96	45	98:97	74.8	97:97	35.9	94:96	41.3	100:97	77.7 (increased)
24	Opioids	–16.9	97:97	45.2	99:98	1.49	100:99	3.8	99:100	33.9	99:100	N/A	N/A	50.8 (increased)
26[a]	Opioids	–77.1	84:79	–82.5	92:91	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	5.4
27	Stimulants	–15.2	100:100	–11.6	98:100	10.1	98:100	–27.0	99:97	–25.7	100:98	–1.6	98:100	13.6 (increased)
28	Alcohol	–33.3	99:100	–29.4	99:98	–48.3	99:100	–10.9	99:100	–15.6	99:100	–11.6	100:99	21.8 (increased)

Participant	Drug	Baseline		Session 1		Session 2		Session 3		Session 4		Session 5		Overall change in attentional bias
		Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	Attentional bias	Task ratio (neutral: drug)	
29[a]	Alcohol	–41.4	98:100	–18.0	99:99	20.6	99:99	28.2	99:100	N/A	N/A	N/A	N/A	69.6 (increased)
30	Opioids	1.3	96:94	38.6	98:93	7.9	98:95	11.5	98:91	13	98:98	–8.9	99:98	10.2
31	Opioids	–38.4	88:62	–282.7	96:33:00	–166.4	96:03:00	–190.1	97:03:00	N/A	N/A	N/A	N/A	151.6

[a]Participants did not complete the study, as they left the voluntary program.

[b]N/A: Not available.

[c]There was a holiday during the participant's stay, and hence, the maximum number of sessions completed was four.

[d]Due to a technical issue, participant 13 was not administered an assessment task following the second intervention, and the participant took another intervention task instead. Attentional bias assessment was performed only after the fourth session.

Acceptability of the Intervention

Of the 30 participants, 22 (73%) completed the perspective questionnaires. All the participants sampled rated the app as at least very easy (10/22 participants) or extremely easy (11/22 participants) to use, with one participant describing it as "like a primary school application" [Participant 1]. Participants also commented on the adequacy of the provided instructions:

> *Easy to follow instructions and exercise.* [Participant 3]
>
> *First of all you give me instructions how to press.* [Participant 22]

Other participants felt that the app was easy to use due to the simplicity of the task:

> *No need to take time to think, just look at the photos and press either one of them.* [Participant 6]
>
> *Just to follow the asterisk * star.* [Participant 15]
>
> *Just need to follow the stars.* [Participant 24]

Participants commented on the ease of responding to the task:

> *Only 2 options – left or right that's why it is easy.* [Participant 4]
>
> *Just press only.* [Participant 10]

Participants also felt that the task could be undertaken by a diverse group of participants

> *But both young and old adults would be able to.* [Participant 7]
>
> *Very simple and easy to be administered to subjects at almost any level of intelligence and age.* [Participant 3]

With regard to interactivity, eight participants rated the app as extremely interactive, nine rated it as very interactive, and five rated it as moderately interactive. Participants commented that the application was "like playing the game" [Participant 4] and that the app "becomes more engaging" [Participant 3] over time.

With respect to motivation, there was a range of views: Eight participants reported being extremely motivated, four participants reported being very motivated, five participants reported being moderately motivated, four participants reported being slightly motivated, and one participant reported not being at all motivated. Participants who were motivated shared that the app helped them "pass time" [Participant 1] and that it was "just like playing game" [Participant 4]. Other participants were motivated, as they felt that the app "makes me feel better" [Participant 21] and could "help me with my treatment" [Participant 29]. Some participants who indicated that they were motivated in using the app highlighted possible reasons:

> *Continually using on a daily basis will become repetitive and boring.* [Participant 3]
>
> *It is a repetitive task and some may find it boring to continue using it, unlike a game which really interacts with the user.* [Participant 5]
>
> *If I concentrate, if I do this all the way, boring.* [Participant 10]

The participant who was not at all motivated commented that he/she finds that the app "doesn't help with my addiction problem" (Participant 22).

There was a range of responses to the question concerning whether the images reminded the participants of their drug use: 5 participants responded very, 4 responded moderately, 3 responded slightly, and 10 responded not at all. With regard to participants' confidence in managing their underlying addiction problem, 54% (12 participants) reported no change in their confidence level before and after using the app, and 10 participants reported a change in their confidence level, with 8 participants reporting a positive change.

All participants were invited to provide any additional feedback they had, and these were mainly related to the need to concentrate on the task:

> *Put your mind on it, follow the star, it would not go wrong.* [Participant 14]
>
> *Must be alert.* [Participant 23]
>
> *It requires your full attention because the switch between the image and the asterisk is very fast so I have to be really focused.* [Participant 5]

Discussion

Principal Findings

The results from our study answered our intended research questions. In the published protocol [23], the study was proposed to be feasible if 25% invited agreed to participate and 60% of the recruited patients adhered to the planned intervention. Our results demonstrated the feasibility of the study in terms of participation and adherence. A 25% recruitment rate was necessary because of the strict inclusion and exclusion criteria. Participation rates are expected to be lower, as individuals with a prior history of withdrawal seizures or any prior history of diagnosed seizures and individuals with moderate to severe psychiatric conditions (as assessed clinically) are excluded. The acceptance rates in our study are higher potentially due to the diversity of substance disorders considered and because most of the participants who sought help were patients with opioid dependence who did not have a prior history of withdrawal seizures, as withdrawal seizures are not common in opioid withdrawal. It is also important to recognize that our inpatient detoxification and rehabilitation program are entirely voluntary, and therefore, individuals are free to request discharge should they not be motivated to stay on. Despite the nature of our program, the adherence rate for this study was not affected. One of the other objectives of the feasibility study was to determine if the mobile attention bias modification intervention could assess and modify attentional biases. We found that the mobile intervention was capable of reducing attentional biases in most of the participants, although there were individuals who did not present with baseline biases.

In the protocol, acceptability is defined as the willingness to use the app daily, if at least 30% of the participants rate the ease of use, interactivity, and motivation positively and if at least 30% of the participants perceived a change in their confidence in managing their addictive disorders after receiving three intervention tasks. Except for one participant who withdrew from the study, the remaining participants were amenable to using the app daily (except for those who decided to leave the ward prematurely and hence did not complete the planned interventions). All these individuals decided to leave prematurely for reasons not related to the study, and there were no adverse outcomes reported during the course of the study. In addition, 100% of the participants rated the app to be either very easy or extremely easy to use, which exceeded our projection of 30%. Moreover, 77% of the participants rated the app as very or extremely interactive and 54% reported being very or extremely motivated to use the app, which exceeded our projection of 30%. Finally, 36% of the participants reported a change in their confidence in managing their addictive disorders, which is congruent with our projection.

In the qualitative feedback, 10 participants reported that the images included did not remind them of their substance use. The images used in the existing app might be different from the images of the substances that they have previously used and thus did not manage to capture their attention. This is in line with a previous commentary [24], which reported that one of the key factors leading to the poor reliability of the visual probe task is that of the nature of the stimulus used. That study [24] highlighted the importance of personalization of the stimulus presented to the participants, as it is postulated that stimulus that is relevant and identifiable to the participant would increase the baseline attentional bias score and provide evidence of greater change in the magnitude of attentional biases. Most of the images included in the existing mobile app were extracted from the internet through the United States Drug Enforcement Agency media library. Some of the images were extracted from Singapore's Central Narcotics Bureau's website. It might be possible that the images included do not approximate and are not realistic enough for participants. In our image set for opioids, we showed images of oxycodone and morphine pills, but in Singapore, these are not commonly abused. In addition, in our image set for cannabis, we showed images of spice, which is also not commonly abused in Singapore.

There are clearly several research implications arising from this study. The qualitative feedback from the acceptability part of the study suggests that it is important for us to consider involving patients in improving the app, to allow for personalization of the images and other functionalities of the app. Participatory research methods could be considered, in particular, for focus groups and codesign workshops. In their review, Zhang et al [25] reported that participatory research design methods have been widely applied in both medicine and psychiatry. For psychiatry, these methods have been applied mainly for perinatal depression, dementia, self-harm, and general and youth mental health issues. Their previous review [14] of attention bias and cognitive bias apps in the published literature and the commercial stores revealed that there is a disconnect between academics and developers. Through participatory design, there is potential to enhance the existing app by involving patients and health care professionals in a joint codesign in order to create an app that is more feasible; acceptable; and capable of detecting and modifying biases in opioid, cannabis, stimulants, and alcohol disorders. Our results might also be affected by the images that we chose to include in the app. In the next iteration of this intervention, we will recommend that participants rate the relevance of the images first, before embarking on the actual intervention. Our study also showed that some participants who are in the rehabilitation phase might not present with baseline attentional biases. Thus, this suggests that future research on such an intervention among individuals who are undergoing rehabilitation ought to consider assessing baseline attentional biases first; otherwise, the intervention would be futile in modifying biases.

Conclusions

To our knowledge, our study is the first study to recruit an Asian cohort of participants with substance use disorders and examine the feasibility and acceptability of the mobile bias intervention. Our results highlight the feasibility to recruit participants to undertake attention bias modification interventions and that participants generally accept a mobile version of such an intervention. Nevertheless, our acceptability data highlight that there could be improvements in the existing app. It is important for future research to take into consideration our findings and adopt a participatory design approach when refining the

conventional visual probe task to cater to the needs of the participants.

Acknowledgments

MZ, JY, SA, ZM, GS, DF, and HE jointly conceptualized the study. MZ, SA, and ZM were involved in data collation. MZ and JY coded and analyzed the initial data. MZ, JY, SA, ZM, GS, DF, and HE reviewed the final dataset. MZ wrote up the first draft of the manuscript, with guidance from DF and HE. All authors read and approved of the manuscript prior to submission.

MWZ is supported by a grant from the Singapore Ministry of Health's National Medical Research Council (grant number: NMRC/Fellowship/0048/2017) for PhD training. The funding source was not involved in any part of this project. We like to thank the National Addictions Management Service for providing finance support for this project.

Conflicts of Interest

None declared.

References

1. European Monitoring Centre for Drugs. 2018. Perspectives on Drugs: The role of psychosocial interventions in drug treatment URL: http://www.emcdda.europa.eu/publications/topics/pods/psychosocial-interventions [accessed 2019-09-02]
2. World Health Organization. Prevalence of alcohol use disorders URL: http://www.who.int/gho/substance_abuse/burden/alcohol_prevalence/en/ [accessed 2019-09-02]
3. World Health Organization. 2010. ATLAS of substance use disorders URL: http://www.who.int/substance_abuse/publications/atlas_report/profiles/singapore.pdf [accessed 2019-09-02]
4. Cutler RB, Fishbain DA. Are alcoholism treatments effective? The Project MATCH data. BMC Public Health 2005 Jul 14;5:75 [FREE Full text] [doi: 10.1186/1471-2458-5-75] [Medline: 16018798]
5. Field M, Marhe R, Franken IHA. The clinical relevance of attentional bias in substance use disorders. CNS Spectr 2014 Jun;19(3):225-230. [doi: 10.1017/S1092852913000321] [Medline: 23663386]
6. Cox WM, Fadardi JS, Intriligator JM, Klinger E. Attentional bias modification for addictive behaviors: clinical implications. CNS Spectr 2014 Jun;19(3):215-224. [doi: 10.1017/S1092852914000091] [Medline: 24642267]
7. 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]
8. 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]
9. Ziaee SS, Fadardi JS, Cox WM, Yazdi SAA. Effects of attention control training on drug abusers' attentional bias and treatment outcome. J Consult Clin Psychol 2016 Oct;84(10):861-873. [doi: 10.1037/a0040290] [Medline: 27281374]
10. Manning V, Staiger PK, Hall K, Garfield JB, Flaks G, Leung D, et al. Cognitive Bias Modification Training During Inpatient Alcohol Detoxification Reduces Early Relapse: A Randomized Controlled Trial. Alcohol Clin Exp Res 2016 Aug 04;40(9):2011-2019. [doi: 10.1111/acer.13163]
11. 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]
12. 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]
13. Cristea IA, Kok RN, Cuijpers P. The Effectiveness of Cognitive Bias Modification Interventions for Substance Addictions: A Meta-Analysis. PLoS ONE 2016 Sep 9;11(9):e0162226. [doi: 10.1371/journal.pone.0162226] [Medline: 27611692]
14. Zhang M, Ying J, Song G, Fung DS, Smith H. Attention and Cognitive Bias Modification Apps: Review of the Literature and of Commercially Available Apps. JMIR Mhealth Uhealth 2018 May 24;6(5):e10034 [FREE Full text] [doi: 10.2196/10034] [Medline: 29793899]
15. Tan YZ, Zhang MWB, Choo ACC. Perspectives on Modifying Attentional Biases Amongst Individuals with Tobacco Use Disorder Using Technology: A Review. Int J Environ Res Public Health 2019 Jul 24;16(15) [FREE Full text] [doi: 10.3390/ijerph16152644] [Medline: 31344921]
16. Zhu Y, Jiang H, Su H, Zhong N, Li R, Li X, et al. A Newly Designed Mobile-Based Computerized Cognitive Addiction Therapy App for the Improvement of Cognition Impairments and Risk Decision Making in Methamphetamine Use Disorder: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Jun 20;6(6):e10292. [doi: 10.2196/10292]
17. NIHR. National Institute for Health Research. Guidance on applying for feasibility studies URL: https://www.nihr.ac.uk/documents/nihr-research-for-patient-benefit-rfpb-programme-guidance-on-applying-for-feasibility-studies/20474 [accessed 2019-09-02]

18. McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. J Nerv Ment Dis 1980 Jan;168(1):26-33. [doi: 10.1097/00005053-198001000-00006] [Medline: 7351540]
19. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [Medline: 7795497]
20. Ferri CP, Marsden J, De Araujo M, Laranjeira RR, Gossop M. Validity and reliability of the Severity of Dependence Scale (SDS) in a Brazilian sample of drug users. Drug Alcohol Rev 2000 Dec;19(4):451-455. [doi: 10.1080/713659418] [Medline: 28474444]
21. Huo T, Guo Y, Shenkman E, Muller K. Assessing the reliability of the short form 12 (SF-12) health survey in adults with mental health conditions: a report from the wellness incentive and navigation (WIN) study. Health Qual Life Outcomes 2018 Feb 13;16(1). [doi: 10.1186/s12955-018-0858-2]
22. Tan ML, Wee HL, Salim A, Lee J, Ma S, Heng D, et al. Validity of a Revised Short Form-12 Health Survey Version 2 in Different Ethnic Populations. Ann Acad Med Singapore 2016 Jun;45(6):228-236 [FREE Full text] [Medline: 27412055]
23. Zhang M, Ying J, Amron SB, Mahreen Z, Song G, Fung DS, et al. A Smartphone Attention Bias Intervention for Individuals With Addictive Disorders: Protocol for a Feasibility Study. JMIR Res Protoc 2018 Nov 19;7(11):e11822 [FREE Full text] [doi: 10.2196/11822] [Medline: 30455170]
24. Field M, Christiansen P. Commentary on Ataya et al. (2012), 'Internal reliability of measures of substance-related cognitive bias'. Drug and Alcohol Dependence 2012 Aug;124(3):189-190. [doi: 10.1016/j.drugalcdep.2012.02.009]
25. Zhang M, Ying J. Incorporating Participatory Action Research in Attention Bias Modification Interventions for Addictive Disorders: Perspectives. IJERPH 2019 Mar 06;16(5):822. [doi: 10.3390/ijerph16050822]

Abbreviations

ASI: Addiction Severity Index
NAMS: National Addictions Management Service
SDS: Severity of Drug Dependence Scale
SF: Short Form

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by C Choo, F Fellendorf, N Khalili-Mahani; comments to author 21.08.19; revised version received 22.08.19; accepted 29.08.19; published 12.09.19.

Please cite as:
Zhang M, Ying J, Amron SB, Mahreen Z, Song G, Fung DSS, Smith HE
A Smartphone Attention Bias App for Individuals With Addictive Disorders: Feasibility and Acceptability Study
JMIR Mhealth Uhealth 2019;7(9):e15465
URL: http://mhealth.jmir.org/2019/9/e15465/
doi:10.2196/15465
PMID:31586359

Original Paper

A Mobile Phone App Designed to Support Weight Loss Maintenance and Well-Being (MotiMate): Randomized Controlled Trial

Emily Brindal[1], PhD; Gilly A Hendrie[1], PhD; Jill Freyne[2], PhD; Manny Noakes[1], PhD

[1]Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia

[2]Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Epping, Australia

Corresponding Author:
Emily Brindal, PhD
Health and Biosecurity
Commonwealth Scientific and Industrial Research Organisation
PO Box 10041
Adelaide, 5000
Australia
Phone: 61 088305 ext 0633
Email: emily.brindal@csiro.au

Abstract

Background: Few people successfully maintain lost weight over the longer term. Mobile phones have the potential to deliver weight loss management programs that can encourage self-monitoring while also providing some behavioral therapy to assist users in developing personal skills that may be necessary for improved longer term weight loss maintenance.

Objective: The aim of this study was to evaluate a program supporting weight maintenance, which uses a behaviorally based mobile phone app to manage weight, food, exercise, mood, and stress.

Methods: In a randomized controlled trial over 24 weeks, the full version of the app (MotiMate) was compared with a control app (monitoring only; excluding mood and stress) for its effect on weight, diet, and psychological well-being. Both apps had the same visual appearance and were designed to deliver all intervention content without face-to-face contact. The control version included features to track weight, food intake, and exercise with limited feedback and no encouraging/persuasive features. The intervention app included more persuasive and interactive features to help users track their weight, food intake, and physical activity and prompted users to enter data each day through notifications and included a mood and stress workshopping tool. Participants were recruited through advertising and existing databases. Clinic visits occurred at baseline, 4 weeks, 8 weeks, 12 weeks, and 24 weeks. At all visits, the clinical trial manager recorded body weight, and participants then completed a computer-delivered survey, which measured psychological and lifestyle outcomes. Objective app usage data were recorded throughout the trial.

Results: A total of 88 adults who had lost and maintained at least 5% of their body weight within the last 2 years were randomized (45 MotiMate and 43 control). Overall, 75% (66/88) were female, and 69% (61/88) completed week 24 with no differences in dropout by condition ($\chi^2_{1,87}$=0.7, *P*=.49). Mixed models suggested no significant changes in weight or psychological outcomes over 24 weeks regardless of condition. Of 61 completers, 53% (32/61) remained within 2% of their starting weight. Significant increases occurred over 24 weeks for satisfaction with life and weight loss self-efficacy regardless of app condition. Diet and physical activity behaviors did not vary by app or week. Negative binomial models indicated that those receiving the full app remained active users of the app for 46 days longer than controls (*P*=.02). Users of the full version of the app also reported that they felt more supported than those with the control app (*P*=.01).

Conclusions: Although some aspects of the intervention app such as usage and user feedback showed promise, there were few observable effects on behavioral and psychological outcomes. Future evaluation of the app should implement alternative research methods or target more specific populations to better understand the utility of the coping interface.

Trial Registration: Australia New Zealand Clinical Trials Registry ACTRN12614000474651; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=366120

(JMIR Mhealth Uhealth 2019;7(9):e12882) doi:10.2196/12882

KEYWORDS

mobile phone; body weight; lifestyle; mood; happiness

Introduction

Weight Management

According to the World Health Organization, 1.9 billion adults were overweight or obese in 2016 [1]. In response to the challenge of weight management, many weight loss programs have been developed. Although many people have initial success in changing their dietary and/or physical activity behaviors to lose weight, few successfully maintain their lost weight over the longer term [2]. For example, only 20% of people from the National Weight Control Registry in the United States managed to maintain initial weight losses after 2 years [3]. Successfully maintaining weight loss for 2 to 5 years greatly increases the likelihood of longer term success [4], as does increasing the duration of exposure to the weight loss program [5]. However, it currently appears as though weight loss is regained in a linear fashion with few mitigating factors [6].

Given the significant challenge of weight loss maintenance, it is unsurprising that few previous interventions have sought to tackle this issue. Targeting self-regulation skills is 1 strategy that is commonly suggested to assist in weight loss maintenance [2,7]. Wing et al [8] report on a study targeting these very skills, which compared 3 groups: a control group, which received only a quarterly newsletter; a group that received face-to-face intervention; and a group that received a Web-based intervention. One of the core features for both intervention groups was a bathroom scale, which gave color-coded feedback, indicating whether participants had a weight gain of 1.4 kg or less (green), between 1.4 and 2.2 kg (yellow), or more than 2.2 kg (red). Those in the green zone were sent minor reinforcements (mainly through positive messages). Those in the yellow zone were instructed to use problem solving to get back on track, and those in the red zone were instructed to reinitiate weight loss attempts. The face-to-face group attended monthly meetings, whereas the Web-based group received social support and advice through a Web interface. Over 18 months, there was no difference in weight regain for the Web-based intervention (mean 4.7 kg [SD 8.6]) compared with the control (mean 4.9 kg [SD 6.5]). However, the proportion of participants who stayed within 2.3 kg of their starting weight (ie, within the green or yellow zones) was significantly higher in the Web-based intervention compared with the control (45.6% vs 27.6%).

Weight Management Interventions Using Digital Technology

Recent technological progress has resulted in a shift from Web-based to mobile phone-based weight management interventions, with or without face-to-face support with some promising results [9,10]. Mobile phones could be used to extend the active duration of engagement with a weight management program, even through simple features such as a text message [11]. Therefore, apps may be a useful delivery mechanism for prolonging weight management attempts and, consequently, weight loss maintenance. Digital interventions are often described as more cost-effective and able to be wider reaching than more intensive face-to-face programs. As technology becomes more sophisticated, the ability to provide just-in-time intervention means that portable devices may also be able to provide intervention at critical times. Indeed, a review of just-in-time interventions suggested that portable devices may be useful to enhance cognitive behavioral therapy for weight loss programs [12]. Mobile phones also provide an avenue for regular self-monitoring, which have been strongly linked with successful behavior change, particularly in weight management [13,14].

Combining Behavioral Strategies and Digital Technology for Weight Loss Maintenance

In addition to behaviors such as self-monitoring in weight loss maintenance, Elfhag and Rossner [15] recognize the importance of stress and coping. They define coping as, “cognitive and behavioral efforts used to manage external and internal demands...that exceed available resources.” They suggest that people who regain lost weight have poorer coping strategies, use more avoidant coping methods, and use eating to regulate their mood. More recently, this has also been observed in an Australian sample who maintained weight losses 4 months after a weight loss program. This group showed stronger problem-solving skills and described more planning events than those who gained weight over the same period [16]. Conservation of resources, self-regulation theory, and, more recently, ego depletion help to explain these observations [17,18]. These theories suggest that an individual has limited capacity to navigate stresses successfully through each day. When demand exceeds supply, individuals are likely to get off-track, particularly in relation to behaviors that are not yet habitual. The more coping and psychological resources a person possesses, the less likely it is that the demand-supply equation be disrupted, which puts less strain on a person. Considering multiple strategies are needed to overcome a single issue or hassle [19], it is not surprising that the more coping strategies or resources a person possesses, the more likely it is that they will find one that successfully helps them to sustain their desired behavior change (eg, eating better or exercising more).

Very few weight loss maintenance interventions exist, and none has incorporated simple weight loss maintenance strategies into a supportive program that also targets well-being. Positive well-being and optimism can improve resilience and the ability to problem solve [20,21]. It can also help restore resources after depletion [22] and is likely to be a critical factor in the maintenance of behavior. Therefore, the aim of this study was to develop and test a theoretically and evidence-based mobile phone intervention for weight loss maintenance. Previous authors have emphasized the importance of theory-based interventions that use scientific evidence and use the functionality of modern phones [23-25]. Specifically, we aimed

to design and evaluate an app to improve psychological well-being, engagement with the intervention, and, ultimately, weight maintenance outcomes.

Methods

Overview

The study was approved by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Human Research Ethics Committee (14/02) in April 2014 and registered with the Australian New Zealand Clinical Trials Registry (ACTRN12614000474651). After being screened over the phone by the clinical trial manager, potential participants attended a study information session delivered by the principal investigator, received an information sheet, and then provided written consent to participate. A grocery voucher was given to participants at weeks 12 and 24 to thank them for their time (total 2×Aus $20 per participant). At the end of the study, participants could request access to the alternate app.

Participants

Power calculations were based on changes in mood observed in our previous study [26]. In a sample with 44 females divided into 2 conditions, we were able to detect a moderate effect (0.45) for changes in mood. The initial aim was to recruit 150 volunteers to allow for 30% dropout [26,27] and the inclusion of males, which may increase the variability in observations. The primary method of recruitment was through an existing clinical research unit database owned by CSIRO, which included the contact details of people who had consented to be contacted about future research. This method was supplemented by local print advertising, promotional news stories, and unaddressed promotional pamphlets delivered by Australia Post. In final recruitment efforts, an external recruitment company was engaged.

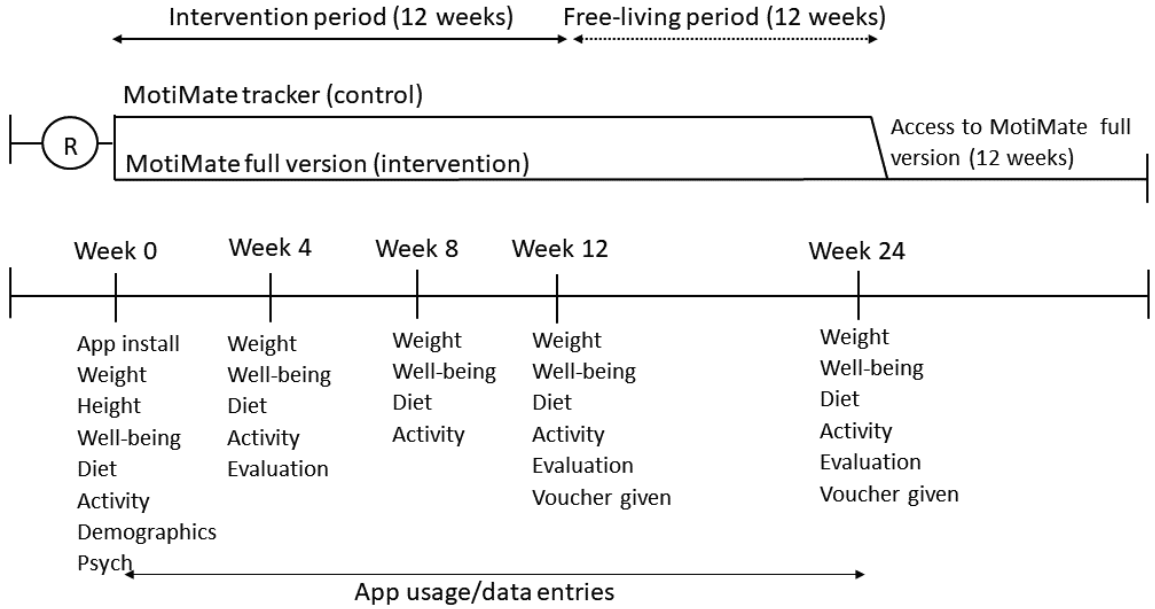
Participants had to meet the following eligibility criteria: adults (aged 18 years or older), lost at least 5% of their body weight within the last 2 years, access to bathroom scales, want to continue or maintain their weight loss, own a mobile phone with an operating system appropriate for the app (iPhone or Android), and willing to attend a clinic in the central business district 5 times over 6 months. Like a previous study, we asked participants to verify previous weight loss with a signed statement by friend and health professional [8]. Exclusion criteria were pregnancy (or planning pregnancy), active cancer, and type I diabetes.

Study Design

The trial was a 12-week, parallel, randomized, single-blind, controlled trial with 12-week follow-up. Participants were randomized to 1 of 2 groups (intervention or control) in a 1:1 ratio. The clinical trial manager allocated participants based on their ID using a random number generator. During randomization, subjects were balanced for sex, age, ownership of an iPhone (vs an Android), and obese versus not (based on self-reported information in the screening questionnaire). All participants received an app called MotiMate and were blinded regarding their allocation. None of the investigators were involved with participant allocation. Due to the collection of objective usage data (described further below), investigators could not be blinded surrounding participant allocation; some participants had app feature interactions only available on the MotiMate intervention app.

Between late 2014 and mid-2015, participants made 5 visits to the clinical research unit in Adelaide, South Australia. Visits occurred at baseline (week 0), 4 weeks, 8 weeks, 12 weeks, and 24 weeks (Figure 1). At all visits, the clinical trial manager recorded body weight in kilograms, and participants then completed a computer-delivered survey that was programmed in SurveyMonkey (SVMK, Inc). These visits generally took less than 15 min each.

Figure 1. Study protocol. Activity: physical activity; Diet: diet quality; Evaluation: evaluation questions; Psych: self-esteem, restraint, satisfaction with life; Well-being: battery of measures.



At the baseline visit, the intervention or control app was manually loaded onto the participants' phones. Clinic staff confirmed correct allocation and recorded allocation in the participant clinic record. They then showed participants the app icon and ensured that participants could log in to the app using the account credentials entered at setup. To replicate a real-world setting where the app would be downloaded from an app store, clinic staff did not provide an overview of the app to participants.

Interventions

Both groups received a mobile phone app designed by the research team to be used without any additional face-to-face support. The app was developed by an external company (Enabled) with close oversight from the study team and programmed for both iPhone and Android users. A helpline was established where technical enquiries or faults could be logged. The choice to make the intervention self-directed was partly to optimize cost-effectiveness and scalability but also because self-directed interventions have been shown to promote weight loss [28]. The content of the app did not change throughout the trial.

The development and features of the full version of the MotiMate app are described in detail elsewhere [29]. Briefly, both study apps had the same visual appearance, labeled MotiMate, and designed for daily use. The control version (also referred to as the tracker) included only features to track weight, food intake, and exercise. It primarily involved data entry with limited feedback and no encouraging/persuasive features (Table 1). The only feedback feature provided was a weight change graph. The full version of the app (or intervention version) was designed to include more persuasive and interactive features to help users track their weight, food intake, and physical activity and prompted users to enter data each day through notifications (Figure 2). In the intervention app, the food tracker gave users immediate feedback on whether they were meeting nutritional guidelines based on the number of serves of each food group that is recommended. Serving sizes were defined under the information tab. On the basis of the design used by Wing et al [8], the weight tracker in the intervention app also gave immediate feedback as to whether people were maintaining their weight (within 1.4 kg of their starting weight), gaining/in the danger zone (1.4 kg higher than starting weight), or had gained weight (2 kg or more over starting weight). These categories were indicated through colors, which went from green (maintaining) to faded green (gaining) to red (gained). The text displayed below the weight value also changed to be more directive with these categories. An automated email was sent to the study email address for participants who had entered a weight classified as gain. These people were contacted as soon as possible by a registered dietitian and asked if they had any questions or needed any advice. These phone calls were short and designed to provide just-in-time intervention to minimize further weight gain. Weekly summaries and graphs, which contextualized data entries in terms of success and areas for improvement, were also included in the intervention app.

Another major feature included in the full intervention app was a mood monitoring interface, which included a workshopping feature design to allow participants to develop their coping skills and emotional regulation. The workshopping interface was only triggered if a change in mood from positive to negative was detected or if a very positive mood became much less intense (large decrease in score). Once activated, users could workshop a cause of the change in their mood and then generate coping strategies to manage it. Once this process had been completed, if a positive change in mood was detected (from negative to positive), users then entered what had worked for them to help them change their mood. If users entered a highly negative mood (based on standard deviations) or prolonged negative mood states (of at least consecutive 7 days), an automated message was sent to the study email to contact this person regarding their mood. This was included to offer just-in-time intervention. In this instance, phone contact was made by a provisionally registered psychologist who followed a predefined protocol to assess if the user needed further support or referral to other services.

Despite creating slightly more cost and complication for a real-world translation of the MotiMate app, just-in-time contact with a registered dietitian and psychologist was included in the app to maximize its potential benefit. The ultimate vision for the app was it should be included as a tool as part of a wider health service with ready access to such professionals rather than employing these people specifically to support the app.

Table 1. Summary of core features of trial apps.

Feature	Intervention version	Control version
Daily notifications/prompts	The app sends a prompt to the user to remind them to enter data	Nil
System notifications	The system detects increases in weight or problematic mood patterns and emails the administrator who can then make person contact with the user to troubleshoot or direct to help	Nil
Weekly report	Summarizes all data entered each week and releases the report to the user	Nil
Motivational messages	At the top of the home screen, a motivational quote appears. The tone of these progresses with the duration of interaction	Nil
Weight entry	Participants slide the weight indicator to change their weight. Feedback is provided immediately with changes	Entry with no feedback.
Food entry	Users tap to indicate how much of food groups allowance they have consumed. A tick appears to indicate a satisfied group. An exclamation appears to indicate overconsumption	Entry with allowances but no tick or exclamation mark feedback
Diet action plans	System detects under- or overconsumption in certain target food groups and suggests that the user focuses on this area. In this interface, the user chooses from a prepopulated list of goals	Nil
Mood entry	Users can select from 6 different moods and then enter stress, location, and time	Nil
Mood change	The system detects negative changes or improvements in mood and triggers the coping workshop	Nil
Coping workshop	This guides people through planning how to overcome hassles (for negative change) or allows people to select coping strategies they did use to improve their mood (for improvement)	Nil
Exercise entry	Users enter duration, intensity, and type of exercise. They receive encouraging feedback for each entry	Entry with no feedback
Information text throughout	“i” buttons throughout the app give background information, instructions how to use features, and details about serving size for food groups	Same information with slightly less encouraging tone
Reviewing or entering data	The arrows at the top allow the user to navigate through previous data to enter data or review. Weight cannot be changed for previous days. Text at the top of the screen gives prompts encouraging more effective retrospective recall	Same as the intervention
Weight graph	Simple line graph showing changes in weight	Same as the intervention
Food graph	Summarizes food group intake according to whether each group is on target, over or under daily allowances	Nil
Exercise graph	Summarizes daily exercise entries according to moderate and intense minutes of exercise	Nil
Mood graph	Presents each mood recorded throughout the day. Moods can be tapped to see further details	Nil
Strategy graph	Summarizes all types of strategies used from the coping workshop into their parent groups: social, emotional, action, distraction, and others	Nil

Figure 2. (A) Home screen for intervention app. (B) Mood entry interface for intervention app. (C) Food intake interface for intervention app. (D) Home screen for control app. (E) Information regarding weight tracking tool for control app. (F) Information regarding weight tracking tool for intervention group.

Technical Errors Throughout the Trial

Due to the technical issues with the app, an update was released after the first week of the trial. This affected iPhone users only. The update was released within 2 days of a reported fault. A total of 4 participants reported technical issues with their app. Two of these (did not see a weekly report and last data entered not saved) resolved themselves and may have been related to a temporary outage of the external database. The other 2 reports related to the app opening slowly and were for Android systems. The developers could not replicate this issue, and participants persisted with the app despite this inconvenience.

Primary Study Outcomes

Unless stated otherwise, all outcomes were assessed at each visit (0, 4, 8, 12, and 24 weeks).

Subjective Well-Being

Subjective well-being was captured through 4 different variables including life satisfaction, depression, anxiety, stress, mood, and global happiness. Life satisfaction is considered an excellent indicator of a person's total well-being and was assessed using the 5-item Satisfaction with Life Scale [30]. The short form of the Depression Anxiety Stress Scales was used to assess depression, anxiety, and stress [31]. Mood was captured using the validated and widely implemented 20-item Positive and Negative Affect Schedule [32]. Fordyce's [33] simple 2-question scale was used to assess happiness. In addition to giving a total level of happiness between 0 and 10, respondents are asked to indicate the percentage of time that they feel happy, unhappy, and neutral.

Weight

The trial manager measured weight to 2 decimal places in the clinic using calibrated electronic scales (Mercury, AMZ 14) and standard operating procedures for collecting weight values.

Other Outcomes

App Engagement/Evaluation

Interactions with the app including logging in and accessing each of the core features were captured by the app and sent to an external database. In a questionnaire, participants were also asked to complete a formal evaluation of each of the components of the app at weeks 4, 12, and 24. The evaluation assessed perceptions of features and opinions toward the app. Perceived usefulness, ease of use, and overall attitude to the app were assessed based on the Technology Acceptance Model [34], which is widely used to evaluate new technologies in the discipline of information systems. These data were largely descriptive in nature and are not reported in this paper.

Self-Efficacy

A total of 3 forms of self-efficacy (nutrition, exercise, and weight loss) associated with weight maintenance were measured. Nutrition and exercise self-efficacy were measured using the 10-item Nutrition and Physical Activity Self-Efficacy Scale [35]. These assess a person's confidence in their ability to eat healthy foods and perform exercise in the presence of likely barriers. Weight loss self-efficacy refers to a person's feeling that they can resist from eating in several different scenarios, such as when feeling emotional and distracted and in social settings. It was measured using the 20-item Weight Loss Self-Efficacy Scale [36].

Resilience

Resilience refers to a person's belief in their ability to persist in the presence of difficulties. It is related to self-efficacy but encompasses a broader concept without being domain specific. To assess resilience, the 6-item scale, The Brief Resilience Scale, was used [37].

Coping

The 28-item Brief COPE questionnaire was used to assess coping style [38]. The tool assesses 14 different coping styles with higher scores representing greater use of each strategy. To minimize multiple analyses on each coping style, the subscales were factor analyzed (Multimedia Appendix 1). This indicated the presence of 2 factors used in the current analyses: the first included *Active* strategies (planning, active, reframing, emotional support, acceptance, and instrumental support), and the second included *Avoidant* strategies (denial, behavioral disengagement, substance use, and self-blame). Humor, religion, self-distraction, and venting did not load clearly on a single factor and were excluded. Factor scores were used to calculate an overall score for each of the 2 factors.

Lifestyle Behaviors

The 38-item short food survey gives a global score for diet quality out of 100 based on how well a person is meeting Australian Dietary Guidelines for the quantity, quality, and variety of different food groups [39]. The short form of the International Physical Activity Questionnaire was used to capture moderate, vigorous activity, and walking and sitting time throughout the previous 7 days [40]. It provides estimates for metabolic equivalent minutes, which represent a summary of total activity performed.

Demographics (Baseline Only)

Participants' characteristics were captured using a standard medical questionnaire administered by the clinical research unit. Participants also completed several items describing their previous weight loss history.

Confounding Variables (Measured at Baseline Only)

Self-Esteem

Self-esteem can influence many aspects of well-being [41]. It was assessed using the 10-item Rosenberg Self-Esteem Scale [42].

Dietary Restraint

To control for unwanted effects of dietary restraint, the 16-item Rigid Restraint Scale was used [43].

Neuroticism

Neuroticism describes the dispositional tendency to experience negative emotional states and is critical to outcomes such as mood. The 6-item Eysenck Personality Questionnaire Revised—Abbreviated was used to assess participants' levels of neuroticism [44].

Dispositional Optimism

Dispositional optimism refers to a person's tendency to have generally a more optimistic or positive outlook in the future. The 10-item Life Orientation Test was used to capture this [45]. Greater scores on this measure indicate higher levels of optimism relative to pessimism.

Analyses

All analyses were performed in SPSS version 20 (IBM). The primary analyses involved intention-to-treat methods using mixed modeling to assess differences in well-being, weight, dietary intake, and physical activity levels over the study period between the intervention groups.

Given the smaller-than-desired final sample, preliminary bivariate correlations were used to assess the relevance of

including all confounding variables. Dietary restraint was only weakly associated with a small number of the outcomes and, consequently, was not controlled for in any of the models. Neuroticism, self-esteem, and dispositional optimism (life orientation) related moderately to most of the psychological outcomes. For consistency, these variables were included in models assessing well-being, coping, resilience, and self-efficacy. All models also controlled for participants' sex and age. The primary dependent variables were app condition, changes over time (by week), and the interaction between these 2 variables. In the presence of significant interaction effects between app condition and week, pairwise comparisons were made using Bonferroni adjustments.

Due to the skew in the app interaction data, comparisons of usage of features were made using negative binomial linear models. These models were overdispersed; therefore, the parameter model was estimated by SPSS rather than set to 1. App condition was compared controlling for sex and age in these models.

Results

Final Participants

Despite various recruitment attempts to reach 150 starters, 88 people started the trial (58.7% of 150 target), and 61/88 completed the trial (69% of starters). There were no differences in dropout by condition ($\chi^2_{1,87}$=0.7, P=.49). Most people withdrew (n=11) because of being too busy with other commitments. Others were lost to contact (n=9; see Figure 3). On the basis of our previous observations [26], this number of completers should have provided 81.1% power to detect a moderate-sized difference in change in mood between groups and 97% power to detect a 2.5% difference in weight between groups.

Figure 3. Consolidated Standards of Reporting Trial (CONSORT) participant flow diagram for trial.

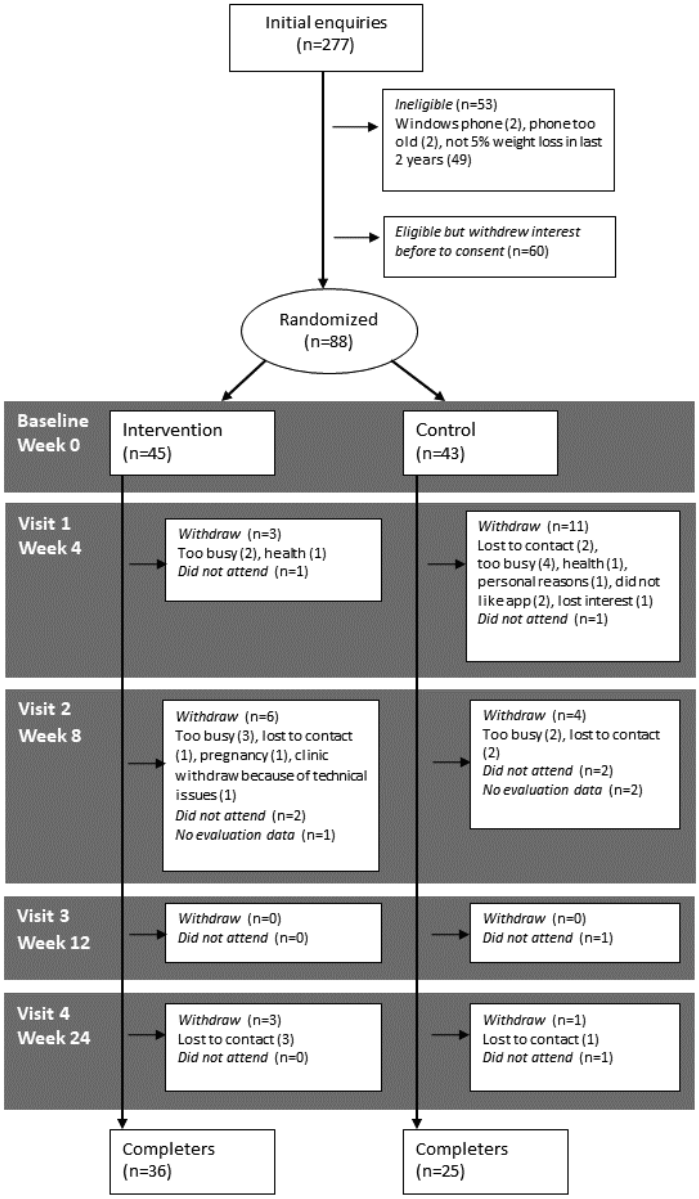


Table 2. Participants' demographics at trial commencement. No statistical differences were found between intervention and control groups for any categories.

Variables	Intervention (n=45)	Control (n=43)	Total
Sex (female), n (%)	33 (73)	33 (77)	66 (75)
Age, mean (SD)	44.5 (13.39)	45.8 (13.11)	45.13 (13.19)
Weight status, n (%)			
Normal	10 (22)	7 (16)	17 (19)
Overweight	20 (44)	15 (35)	35 (40)
Obese class 1	8 (18)	13 (30)	21 (24)
Obese class 2	1 (2)	4 (9)	5 (11)
Obese class 3	6 (13)	4 (9)	10 (22)
iPhone ownership (vs Android), n (%)	31 (69)	31 (72)	62 (71)

The sample was between the ages 20 and 67 years and mostly female (66/88; 75%; Table 2). A majority owned an iPhone (62/88; 71%) rather than an Android handset. The group's starting weight ranged from 53.4 to 170.4 kg with a mean of 85.8 kg (SD 22.08). Body mass index was between 20.9 and 60.8 kg/m^2.

In terms of their weight maintenance, most of the sample (76/88; 86%) was currently below the heaviest weight they had been, but above the lowest weight they had ever been. At the start of the trial, participants reported being between 5.0% and 45.2% lighter than their maximum ever weight with 64% (56/88) maintaining at least a 10% loss from their maximum weight. The remaining 11 people in the sample were at or within 1% of their lowest weight when they started the trial.

Primary Outcomes

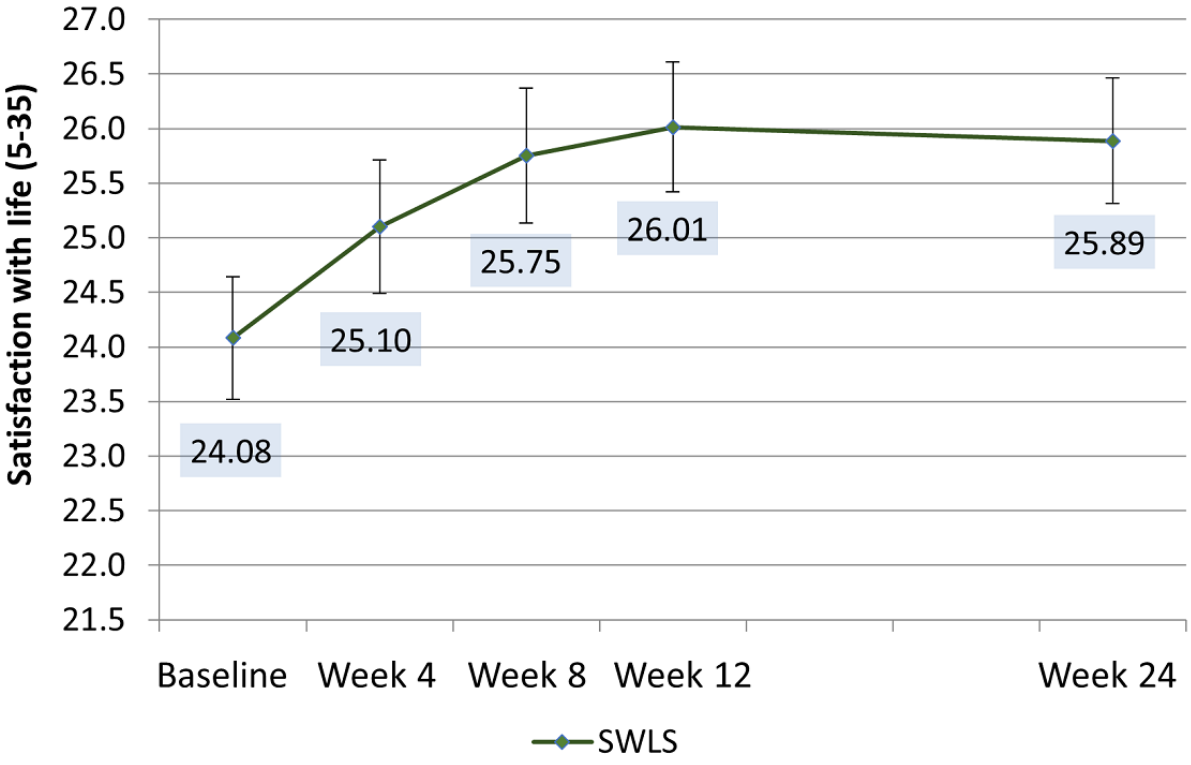
Subjective Well-Being

Satisfaction with life did not vary by app condition; however, there was a significant effect of time for the pooled means between groups (Figure 4). The differences between baseline and week 8 (P=.046), week 12 (P=.01), and week 24 (P=.01) were all significant. Means indicated that, for all study participants, life satisfaction improved over the 24-week study period.

Neither depression nor stress varied significantly by week or app condition (Multimedia Appendix 2). The interaction between app condition and week was significant for anxiety scores. Posthoc comparisons revealed no differences between the app conditions at any time point. The only significant pairwise comparison was for the large decrease in anxiety between baseline and week 4 for the control group (P=.02).

The interaction between app condition and time was also significant for negative affect. Scores for the control and intervention groups were significantly different at baseline (P=.02) and week 24 (P=.01), with the control group starting with significantly higher levels of negative affect and finishing the trial significantly lower. Although the control group had a significant reduction in negative affect between baseline and week 24 (P<.001) and week 12 and week 24 (P=.01), the intervention group had no differences in their negative affect levels throughout the trial (Multimedia Appendix 2). Positive affect, happiness, and the proportion of the time spent happy and unhappy did not vary by the app condition or week throughout the trial (Multimedia Appendix 2).

Figure 4. Adjusted means for Satisfaction With Life Score (SWLS) over the study period pooled for both intervention groups. Significant comparisons: baseline to week 8 (*P*=.046); baseline to week 12 (*P*<.01); and baseline to week 24 (*P*=.01). Means adjusted for participants' sex and age, neuroticism, self-esteem, and dispositional optimism.



Weight

Most of the final sample (32/61; 53%) remained within ±2% of their starting weight at 24 weeks, with an average shift of less than 0.1% between baseline and week 24. At week 24, 41% (25/61) of participants who attended their final visit ended at the same weight or with a net loss (0% to −8.54%). The remaining 4 participants gained between 0.17% and 10.32% of their starting weight. There were no significant differences between the different app versions or over time for the percentage of weight change from baseline (Multimedia Appendix 2).

Other Outcomes

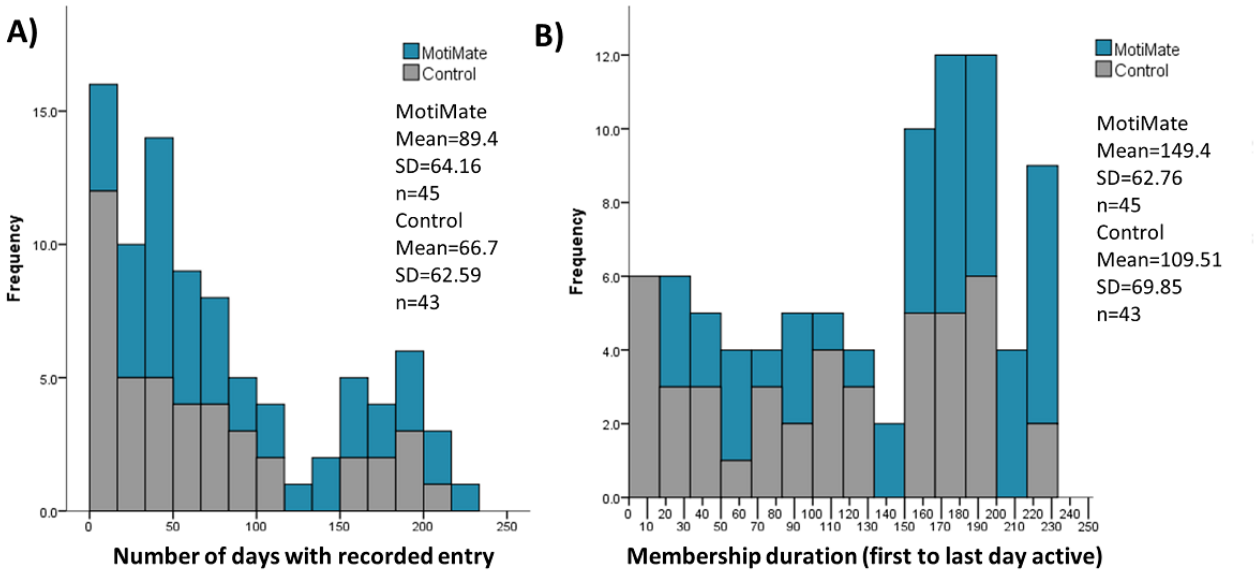
App Usage

Those with the intervention app remained active users of the app for significantly longer than the control group, with a mean difference of almost 50 days (Table 3). Interactions with the app ranged between 0 and 168 days of the trial, with some users continuing their usage beyond their final visit (up to 223 days; Figure 5). The single user who had 0 days of membership was in the control group and received the app but never opened it before dropping out before their second visit.

Table 3. Adjusted means based on negative binomial models for food and exercise entries, number of days data were entered, and total days remaining active (membership days). Models adjusted for participants' sex and age.

Variables	Intervention (n=45)		Control (n=43)		Wald chi-square	*P* value
	Mean	SE	Mean	SE		
Food entries	67.05	10.13	47.51	7.37	2.9	.09
Exercise entries	40.94	7.68	23.38	4.45	5.1	.02
Days data entered	87.04	11.86	62.69	8.95	3.2	.07
Membership days	151.24	16.91	105.42	12.43	5.7	.02

Figure 5. (A) Number of days data were recorded by the participants. (B) Participants' overall membership duration (the amount of time between when the user started using the app and the final time they used it).



Data Entries

In week 1, users created food entries on 5.8 out of a possible 7 days. By week 12, 3.3 entries, on average, were made per person, and during the free-living periods (weeks 12-24), between 2 and 3 entries per user were recorded per week (Figure 6).

The median food entry was 54 for the intervention group and 34 for the control. Both the intervention and control groups had 5 users contributing over 150 food entries. However, only 4 (8.9%) of the 45 intervention users made less than 10 food entries compared with 13 (30.2%) of the 43 control group users. Days that food entries were made only trended toward significance between the groups (Table 4).

Those receiving the intervention app made significantly more exercise entries relative to those receiving the control app (Table 3). The median exercise entry was 22 for intervention group and 12 for the control group. Exercise recording fell more sharply than food entry recording (Figure 5). Walking was by far the most popular exercise recorded (1836 entries) followed by weights/fitness classes (657 entries).

Figure 6. Number of data entries made per person each week over the study period for exercise, food, and mood (intervention group only).

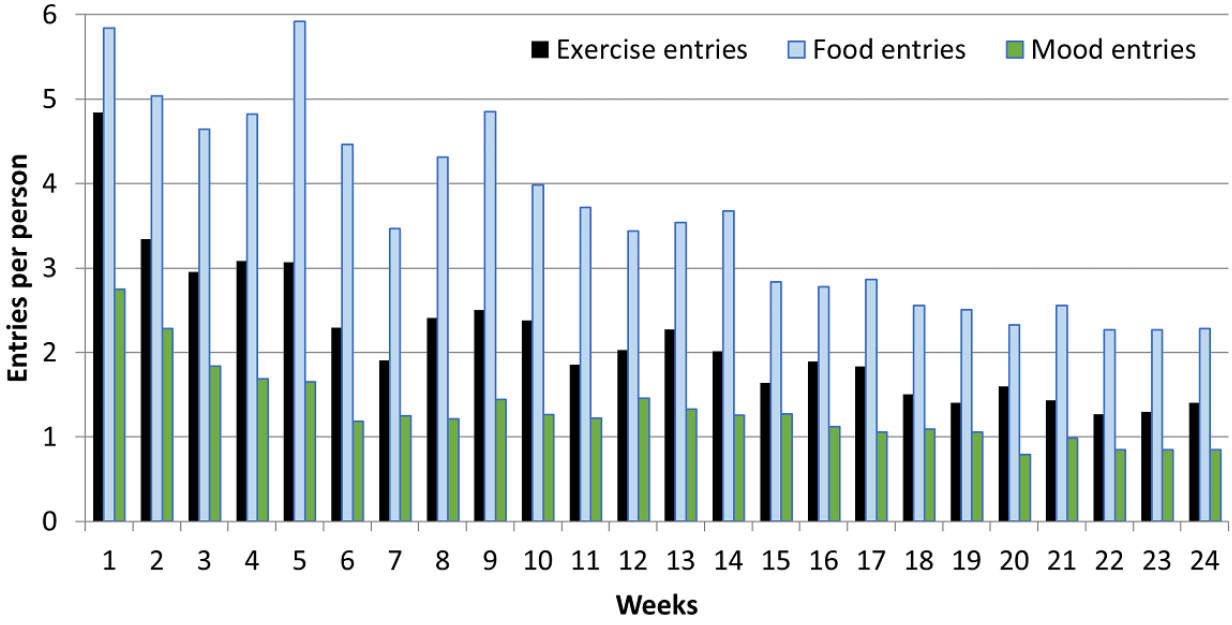


Table 4. Means for evaluation ratings of the app presented by the intervention and control app groups.

Variables	Intervention		Control		*P* value
	n	Mean (SD)	n	Mean (SD)	
Week 4					
The app helped me control my weight	39	3.41 (0.906)	31	3.16 (1.003)	.28
I have enjoyed using the app	39	3.67 (0.898)	31	3.42 (1.148)	.32
Ease of use of app (TAM[a])	39	3.69 (0.919)	31	3.94 (0.859)	.25
Perceived usefulness of app (TAM)	39	3.69 (0.919)	31	3.94 (0.859)	.23
Attitude to app (TAM)[b]	39	3.96 (0.818)	31	3.53 (0.890)	.04
Week 12					
The app helped me to be more aware of my eating	34	5.44 (1.501)	29	5.03 (1.592)	.30
The app helped me to be more aware of my exercise	34	5.24 (1.671)	29	4.55 (1.703)	.11
The app helped me to be more aware of my weight[b]	34	5.62 (1.415)	29	4.72 (1.709)	.03
The app has supported me[b]	34	5.38 (1.518)	29	4.17 (1.872)	.01
Week 24					
What score out of 10 would you give the app?[b]	32	6.38 (2.012)	28	5.04 (2.333)	.02

[a]TAM: Technology Acceptance Model.

[b]Significant differences.

In the intervention group, there was a wide range of interactions with the mood monitoring feature with a median of 20 mood entries per person over the 24 weeks. In total, 2346 mood entries were made. More than 50% of the intervention sample (23/45) had less than 20 mood entries throughout the study period. There was a small, but very active, group of users (7/45, 16%) who made more than 100 mood entries. Overall, mood recording was used less than the traditional food and exercise monitoring tools with 2.75 entries per person in week 1 and a steep decline in entries even over the first few weeks (Figure 5). The most commonly entered mood was happy (924/2346, 39.38%), followed by relaxed (n=500/2346, 21.31%) and positive (497/2346, 21.18%).

Use of Persuasive Features (Intervention Group Only)

Only 2 users were contacted by a provisionally, registered psychologist, who was part of the wider study team, because of entering a pattern of highly negative moods—one reported having a relationship break up, and the other reported suffering from posttraumatic stress disorder.

Furthermore, 5 users were contacted by a dietitian regarding weight gains. These were largely around the Christmas holiday period, and participants generally did not want specific help or advice, generally saying, “they knew what they needed to do.”

In addition, 25 (56%) of the 45 participants in the intervention group received a diet action plan (147 action plans generated). Only 43 of the 147 plans (29.3%) were marked as completed by users. The most common plans triggered were those relating to underconsumption (87/147). This may be because people had not entered food data for these days. This was followed by messages regarding excessive discretionary foods (n=25) and not enough fruits and vegetables consumption (n=18).

Only 3.87% (91/2346) of all moods entered had an associated workshop entry recorded. Hassles could be entered if users selected “Tell us more” to enter the coping workshop. In 2254 cases, users answered “Dismiss” to this question. A total of 92 coping workshop entries were made. The most frequent hassle was “Nothing in particular” (22%, 20/92), followed by “People problems” (20%, 18/92).

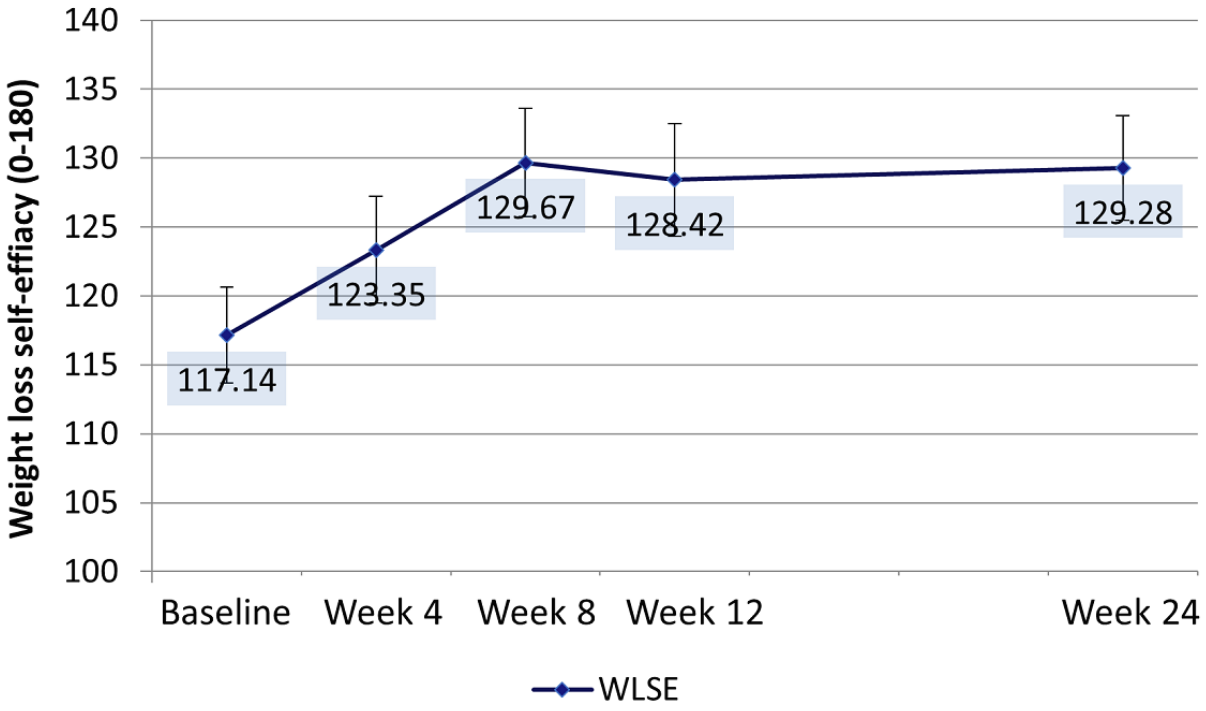
Self-Efficacy, Resilience, Coping, Diet, and Activity

Weight loss self-efficacy only differed significantly by week (Multimedia Appendix 3). Increases between baseline and week 8 (P=.001), week 12 (P<.01), and week 24 (P<.001) were significant (Figure 7).

There was a significant interaction effect between app condition and week for resilience (Multimedia Appendix 3). The strongest difference between apps at any time point was at week 12, where the intervention group had higher resilience than the control group. However, this failed to reach significance (P=.08). The interaction effect appeared to be driven by significant differences between weeks within app condition. The control group had an initial improvement in resilience with a significant difference between values at baseline and week 8 (P=.04). The intervention group had a significant decrease in resilience in the free-living period (from weeks 12 to 24; P=.02). These were the only significant pairwise comparisons.

There were no differences for coping styles, diet quality, and physical activity between app condition, time, or the interaction between the 2 (Multimedia Appendix 3).

Figure 7. Adjusted means for weight loss self-efficacy (WLSE) over the study period. Significant comparisons baseline and week 8 (*P*=.001), week 12 (*P*<.01), and week 24 (*P*<.001).

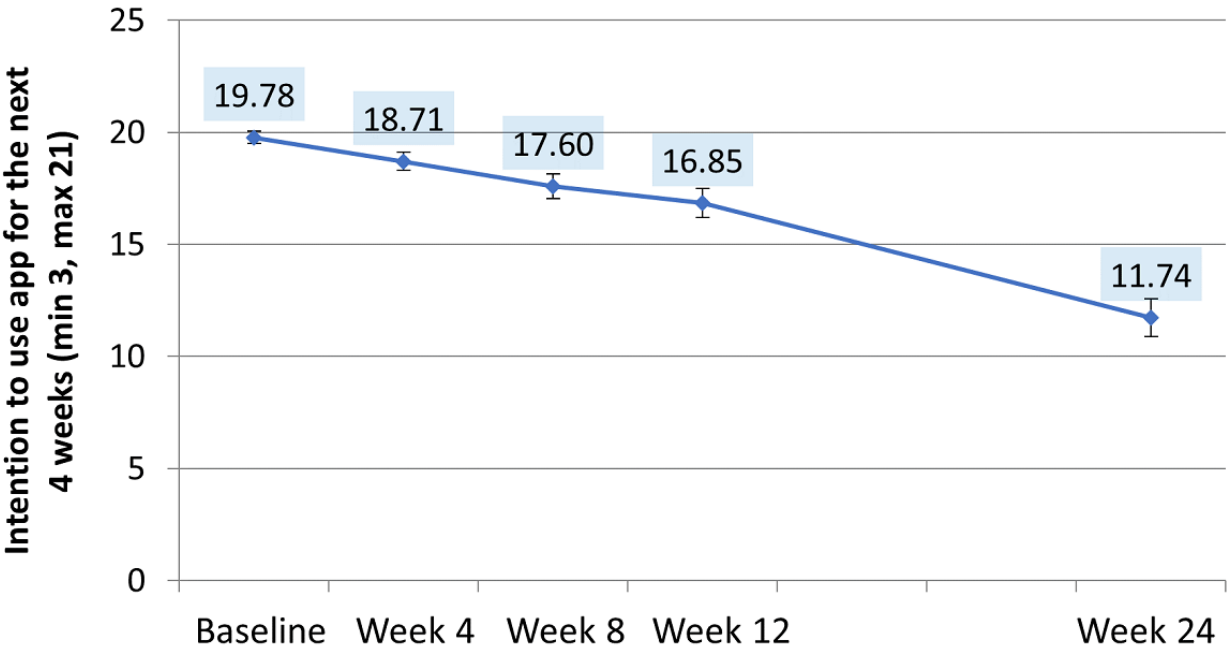


App Feedback

At week 4, most of the intervention app users would have recommended the app to a friend (32/39, 82% *yes*; 4/39, 10% *no*; and 3/39, 8% *other*). In contrast, less than half of the participants agreed that they would recommend the control app (14/31, 45 *yes*; n=12/31, 39% *no*; 5/31, 16 *other*). This difference was significant (χ^2_2=10.7, *P*<.01). A mixed models comparison of the average scores across the study showed that the intention to continue using the app had fallen significantly by the end of 24 weeks ($F_{4,\ 66.9}$=22.74, *P*<.001; Figure 8). Pairwise comparisons revealed that changes from baseline to week 8 onward were all significant (all *P*<.01). Changes did not differ by the app conditions (Figure 8).

There were differences between the intervention groups for attitudes toward the app, how much users felt the app supported them, and how the app assisted users in being aware of their weight (Table 4). At the end of the study, the overall rating given to the intervention app was also significantly higher.

Figure 8. Intention to use app for the next 4 weeks. Adjusted means based on mixed models, pooled across app condition. Changes from baseline to week 8 onward were all significant (all *P* ≤.005). min: minimum; max: maximum.



Discussion

Summary of Results

The MotiMate app was designed to provide support for those undertaking weight loss maintenance. It was designed to be cost-effective with minimal personal contact and support people in tracking their weight, food intake, exercise, and moods. In our highly controlled, 6-month clinical trial of the MotiMate app, we were unable to show any additional benefits of persuasive features and mood monitoring in terms of psychological well-being and weight maintenance for participants. This is despite observations of longer engagement with the app, more exercise entries, and more positive rating of the intervention app by users. There were improvements in weight loss self-efficacy and life satisfaction throughout the trial in both groups. These are important constructs for well-being and weight maintenance. Interaction effects were observed for anxiety and negative affect. However, posthoc analyses revealed that these may have been driven by baseline differences and not the intervention *per se*. A significant interaction effect for changes in resilience was also observed with the intervention group having significant falls in the free-living period, whereas the control group did not. There were minimal differences observed in lifestyle behaviors and other subjective well-being constructs.

Weight Management

Most participants maintained their weight regardless of app condition over 24 weeks, with more than half remaining within 2% of their starting weight. On face value, this seems like a positive outcome, especially for weight loss maintenance, which is notoriously challenging. This result supports a previous study by Wing et al, which indicated that 53% of their control group maintained weight at 6 months [8]. Their control group was much less active than the one used in this study, as their participants only received newsletters, whereas our control participants received a monitoring app.

At the start of our trail, participants were within different ranges from their lowest ever weights, and they each had different time frames with which they had been maintaining their weight, as well as different experiences with weight loss programs before starting the study. It would have been interesting to explore how these factors may have altered weight outcomes; however, our ability to do this was limited because of the sample size. The choice to recruit people with a range of weight management experiences was a purposeful one to assess if the MotiMate design could be effective in a real-world setting, where people have had a variety of weight loss experiences; however, this may have also added increased variability to the outcomes. Close to 40% of people continued to lose weight while on this trial, although they had no support specifically directed toward weight loss from the app. There were no differences over time or between apps for diet quality and exercise, which suggest that these people also did not significantly change their lifestyle practices. Therefore, some of these people may have been still engaging in active weight loss efforts to try to overcome a plateau in weight loss rather than maintain an existing weight. Unfortunately, we did not explicitly capture this intention at study commencement. However, recruitment materials and study information all focused on maintaining weight loss rather than losing weight.

Capturing people in the small window between weight loss and regain was more difficult than anticipated. This may be avoided by first placing people into a weight loss program in the future. Although we did not hit our recruitment target, the final sample size still provided adequate power to detect moderate effects.

Mood Features and Psychological Changes

Engagement with the app features related to mood was low. This is likely to explain the absence of differences between the 2 groups. Even more so, given that the control group also received an app with monitoring features. Therefore, these participants received a more active intervention than a standard usual care model in which they may only have been given once-off advice or static information such as newsletters and pamphlets. The low engagement with mood features may be partly because of the study design and the desire to blind participants as to their allocation. No mention was made regarding mood monitoring in recruitment. Qualitative feedback (not reported) indicated that some people were not receptive to tracking their mood. Moreover, 1 participant even indicated that they only ever had *one mood*, and there was *no need to record it*. Indeed, a review of emotion research suggested individual variability in emotional granularity [46]. Trialing the app in an uncontrolled sample would allow us to target a potentially more appropriate market in the future.

It is unclear why those in the intervention group had a significant fall in resilience in the free-living period. Although, it is important to note that this change was observed within this group, and the difference between resilience scores was not significant between the apps. We observed improvements in measures of well-being throughout the trial that have not been documented in many previous studies. Yet, recent studies reinforce the idea that well-being is a critical factor for weight loss maintenance [47], and apps using behavior change techniques relating to problem solving and stress reduction are needed [48]. However, it may also be that weight change alone may be crucial for changes in coping and problem-focused coping [49]. Therefore, the ability of simple behavioral therapy techniques may not be able to add value to weight loss alone. That is not to underplay the potential importance of behavioral therapies for improving adherence to lifestyle programs [50] and the potential benefits of improving behavioral skills before engaging in a weight management program [51].

Since this study started, recent evidence has emerged that suggests that resource depletion theory may not be as strong as has been previously thought [52]. More recent studies have failed to replicate observations consistent with ego depletion [53,54] and have called in to question the presence of the described effects. Ego depletion is a relatively new theory, and further studies may be needed to better understand ego depletion and its relationship to eating habits and weight management. Emotion regulation strategies may benefit those prone to emotional eating more observably than other groups.

MotiMate App Ratings and Engagement

The participants had significantly more positive attitudes to the intervention app relative to the control with 82.1% agreeing that they would be happy to recommend the full version of MotiMate to a friend. Intervention users also felt that the app helped them to be more aware of their weight and felt more supported than people using the control app. Nonetheless, motivation to engage with the app fell for both groups by the end of the trial. However, taken with usage data, those with the intervention app continued engaging with the app longer than those with the control app. Engagement with the intervention app features was also higher relative to the control group. Despite previous papers suggesting that app use is associated with better weight loss results [55] and that extended contact through mobile phone improves weight loss maintenance [56], better app use did not translate to better weight loss maintenance in this instance. This once again may be related to how long participants had been successfully maintaining their weight before the trial.

Engagement with the app and intention to continue using fell over 6 months for both apps. Aside from early drops in usage, there was a visible decrease in motivation at week 8. This is an observation that we have made in similar trials [26]. To improve the testing of app-based programs in the future, alternative methods of evaluation may be needed including adaptive intervention designs [57]. It is also important to note that although participants may not be recording their behaviors into the app, this does not necessarily mean that they have not performed these behaviors. It is likely that there is a point where behaviors such as diet monitoring become habitual, and there is no need to rely on tools for assistance. In a real-world translation of MotiMate, usage could be tracked, and just-in-time contact could be made with users when their interactions fall in an effort to understand why they have stopped using the app or to prompt them to keep using the app. This may help to mitigate disengagement because of perceived failure (eg, not entering weights because there has been a gain).

Conclusions

Although some aspects of the MotiMate app showed promise, there were few observable effects of using the full intervention app relative to the basic tracker. Future evaluation of the app may need to be implemented using more progressive research methods or targeting a larger or more specific population to better understand the utility of the coping interface.

Acknowledgments

The authors wish to thank staff members of the Clinical Research Unit for their management of the participants, Kathryn Bastiaans for preparing the data for analysis, and Kylie Lange for her advice on the statistical methods used. Bupa Health Foundation is a Principal Partner for the development and evaluation of the MotiMate app. They had no direct involvement with the development and evaluation of MotiMate.

Authors' Contributions

EB, GH, and JF were involved in the design of the MotiMate and/or development of app components. EB, GH and MN designed the study. EB performed the analysis of all behavioral and psychological data. JF prepared and analyzed app usage data. EB was responsible for the preparation of the manuscript. All authors contributed to and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Factor analysis of COPE inventory.
[PDF File (Adobe PDF File), 56KB - mhealth_v7i9e12882_app1.pdf]

Multimedia Appendix 2
Adjusted means and standard errors (SE) for primary outcomes presented by week and app condition.
[PDF File (Adobe PDF File), 84KB - mhealth_v7i9e12882_app2.pdf]

Multimedia Appendix 3
Adjusted means and standard errors (SE) for secondary study outcomes.
[PDF File (Adobe PDF File), 79KB - mhealth_v7i9e12882_app3.pdf]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 102KB - mhealth_v7i9e12882_fig.pdf]

References

1. World Health Organisation. 2018. Obesity and Overweight: Fact Sheet URL: http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [WebCite Cache ID 71yhwREPC]
2. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
3. Wing RR, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Dec;82(1 Suppl):222S-225S. [doi: 10.1093/ajcn/82.1.222S] [Medline: 16002825]
4. Klem ML, Wing RR, Lang W, McGuire MT, Hill JO. Does weight loss maintenance become easier over time? Obes Res 2000 Sep;8(6):438-444 [FREE Full text] [doi: 10.1038/oby.2000.54] [Medline: 11011910]
5. Ahern AL, Wheeler GM, Aveyard P, Boyland EJ, Halford JC, Mander AP, et al. Extended and standard duration weight-loss programme referrals for adults in primary care (WRAP): a randomised controlled trial. Lancet 2017 Jun 3;389(10085):2214-2225 [FREE Full text] [doi: 10.1016/S0140-6736(17)30647-5] [Medline: 28478041]
6. National Instiute of Health and Clinical Excellence. 2014. Weight Management: Lifestyle Services for Overweight or Obese Adults URL: https://www.nice.org.uk/guidance/ph53 [accessed 2017-05-01]
7. Gerlach G, Herpertz S, Loeber S. Personality traits and obesity: a systematic review. Obes Rev 2015 Jan;16(1):32-63. [doi: 10.1111/obr.12235] [Medline: 25470329]
8. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/NEJMoa061883] [Medline: 17035649]
9. Schippers M, Adam PC, Smolenski DJ, Wong HT, de Wit JB. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Dec;18(4):450-459. [doi: 10.1111/obr.12492] [Medline: 28187246]
10. Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare 2014 Sep;20(6):339-349. [doi: 10.1177/1357633X14537777] [Medline: 24875928]
11. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28(Suppl 2):1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
12. 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]
13. Hartmann-Boyce J, Boylan AM, Jebb SA, Aveyard P. Experiences of self-monitoring in self-directed weight loss and weight loss maintenance: systematic review of qualitative studies. Qual Health Res 2019 Dec;29(1):124-134. [doi: 10.1177/1049732318784815] [Medline: 29984630]
14. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 Dec 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
15. Elfhag K, Rössner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. Obes Rev 2005 Feb;6(1):67-85. [doi: 10.1111/j.1467-789X.2005.00170.x] [Medline: 15655039]
16. Dibb-Smith AE, Brindal E, Chapman J, Noakes M. A mixed-methods investigation of psychological factors relevant to weight maintenance. J Health Psychol 2019 Mar;24(4):440-452. [doi: 10.1177/1359105316678053] [Medline: 27899447]
17. Baumeister RF, Heatherton TF. Self-regulation failure: an overview. Psychol Inq 1996 Jan;7(1):1-15. [doi: 10.1207/s15327965pli0701_1]
18. Hobfoll SE. Social and psychological resources and adaptation. Rev Gen Psychol 2002 Dec;6(4):307-324. [doi: 10.1037/1089-2680.6.4.307]
19. Folkman S, Lazarus RS. If it changes it must be a process: study of emotion and coping during three stages of a college examination. J Pers Soc Psychol 1985 Jan;48(1):150-170. [doi: 10.2307/2136617] [Medline: 2980281]
20. Fredrickson BL, Branigan C. Positive emotions broaden the scope of attention and thought-action repertoires. Cogn Emot 2005 May 1;19(3):313-332 [FREE Full text] [doi: 10.1080/02699930441000238] [Medline: 21852891]
21. Tugade MM, Fredrickson BL. Resilient individuals use positive emotions to bounce back from negative emotional experiences. J Pers Soc Psychol 2004 Feb;86(2):320-333 [FREE Full text] [doi: 10.1037/0022-3514.86.2.320] [Medline: 14769087]
22. Tice DM, Baumeister R, Shmueli D, Muraven M. Restoring the self: positive affect helps improve self-regulation following ego depletion. J Exp Soc Psychol 2007 May;43(3):379-384. [doi: 10.1016/j.jesp.2006.05.007]
23. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
24. 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]

25. 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]
26. Brindal E, Hendrie G, Freyne J, Coombe M, Berkovsky S, Noakes M. Design and pilot results of a mobile phone weight-loss application for women starting a meal replacement programme. J Telemed Telecare 2013 Apr;19(3):166-174. [doi: 10.1177/1357633X13479702] [Medline: 23520213]
27. Brindal E, Hendrie GA, Taylor P, Freyne J, Noakes M. Cohort analysis of a 24-week randomized controlled trial to assess the efficacy of a novel, partial meal replacement program targeting weight loss and risk factor reduction in overweight/obese adults. Nutrients 2016 May 4;8(5):pii: E265 [FREE Full text] [doi: 10.3390/nu8050265] [Medline: 27153085]
28. Tang J, Abraham C, Greaves C, Yates T. Self-directed interventions to promote weight loss: a systematic review of reviews. J Med Internet Res 2014 Feb 19;16(2):e58 [FREE Full text] [doi: 10.2196/jmir.2857] [Medline: 24554464]
29. Brindal E, Hendrie G, Freyne J. Combining persuasive technology with behavioral theory to support weight maintenance through a mobile phone app: protocol for the MotiMate app. JMIR Res Protoc 2016 Jan 8;5(1):e5 [FREE Full text] [doi: 10.2196/resprot.4664] [Medline: 26747725]
30. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
31. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Second Edition. Sydney: Psychology Foundation; 1995.
32. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063] [Medline: 3397865]
33. Fordyce MW. Development of a program to increase personal happiness. J Couns Psychol 1977;24(6):511-521. [doi: 10.1037/0022-0167.24.6.511]
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
35. Schwarzer R, Renner B. The General Self - Userpage - Freie Universität Berlin. 2005. Health-Specific Self-Efficacy Scales URL: https://userpage.fu-berlin.de/health/healself.pdf
36. Clark MM, Abrams DB, Niaura RS, Eaton CA, Rossi JS. Self-efficacy in weight management. J Consult Clin Psychol 1991 Oct;59(5):739-744. [doi: 10.1037/0022-006X.59.5.739] [Medline: 1955608]
37. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
38. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989 Feb;56(2):267-283. [doi: 10.2307/351428] [Medline: 2926629]
39. 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]
40. Ainsworth BE, Bassett Jr DR, Strath SJ, Swartz AM, O'Brien WL, Thompson RW, et al. Comparison of three methods for measuring the time spent in physical activity. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S457-S464. [doi: 10.1097/00005768-200009001-00004] [Medline: 10993415]
41. Paradise AW, Kernis MH. Self-esteem and psychological well-being: implications of fragile self-esteem. J Soc Clin Psychol 2002 Oct;21(4):345-361. [doi: 10.1521/jscp.21.4.345.22598]
42. Rosenberg M. Society and the Adolescent Self-Image. New Jersey: Princeton University Press; 1965.
43. Westenhoefer J, Stunkard AJ, Pudel V. Validation of the flexible and rigid control dimensions of dietary restraint. Int J Eat Disord 1999 Jul;26(1):53-64. [doi: 10.1002/(sici)1098-108x(199907)26:1<53::aid-eat7>3.0.co;2-n] [Medline: 10349584]
44. Francis LJ, Brown LB, Philipchalk R. The development of an abbreviated form of the revised Eysenck personality questionnaire (EPQR-A): its use among students in England, Canada, the USA and Australia. Pers Individ Dif 1992 Apr;13(4):443-449. [doi: 10.1016/0191-8869(92)90073-X]
45. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [doi: 10.1037/t01038-000] [Medline: 7815302]
46. Barrett LF. Solving the emotion paradox: categorization and the experience of emotion. Pers Soc Psychol Rev 2006;10(1):20-46. [doi: 10.1207/s15327957pspr1001_2] [Medline: 16430327]
47. Robertson S, Davies M, Winefield H. Positive psychological correlates of successful weight maintenance in Australia. Clin Psychol 2015 Jul 7;21(3):236-244. [doi: 10.1111/cp.12073]
48. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
49. Rydén A, Karlsson J, Sullivan M, Torgerson JS, Taft C. Coping and distress: what happens after intervention? A 2-year follow-up from the Swedish obese subjects (SOS) study. Psychosom Med 2003;65(3):435-442. [doi: 10.1097/01] [Medline: 12764217]

50. Burgess E, Hassmén P, Welvaert M, Pumpa KL. Behavioural treatment strategies improve adherence to lifestyle intervention programmes in adults with obesity: a systematic review and meta-analysis. Clin Obes 2017 Apr;7(2):105-114. [doi: 10.1111/cob.12180] [Medline: 28199047]
51. Kiernan M, Brown SD, Schoffman DE, Lee K, King AC, Taylor CB, et al. Promoting healthy weight with 'stability skills first': a randomized trial. J Consult Clin Psychol 2013 Apr;81(2):336-346 [FREE Full text] [doi: 10.1037/a0030544] [Medline: 23106759]
52. Svaldi J, Tuschen-Caffier B, Lackner H, Zimmermann S, Naumann E. The effects of emotion regulation on the desire to overeat in restrained eaters. Appetite 2012 Oct;59(2):256-263. [doi: 10.1016/j.appet.2012.04.016] [Medline: 22580220]
53. Lurquin JH, Michaelson LE, Barker JE, Gustavson DE, von Bastian CC, Carruth NP, et al. No evidence of the ego-depletion effect across task characteristics and individual differences: a pre-registered study. PLoS One 2016;11(2):e0147770 [FREE Full text] [doi: 10.1371/journal.pone.0147770] [Medline: 26863227]
54. Lurquin JH, Miyake A. Challenges to ego-depletion research go beyond the replication crisis: a need for tackling the conceptual crisis. Front Psychol 2017;8:568 [FREE Full text] [doi: 10.3389/fpsyg.2017.00568] [Medline: 28458647]
55. Carter MC, Burley VJ, Cade JE. Weight loss associated with different patterns of self-monitoring using the mobile phone app My Meal Mate. JMIR Mhealth Uhealth 2017 Feb 2;5(2):e8 [FREE Full text] [doi: 10.2196/mhealth.4520] [Medline: 28153814]
56. Spark LC, Fjeldsoe BS, Eakin EG, Reeves MM. Efficacy of a text message-delivered extended contact intervention on maintenance of weight loss, physical activity, and dietary behavior change. JMIR Mhealth Uhealth 2015 Sep 15;3(3):e88 [FREE Full text] [doi: 10.2196/mhealth.4114] [Medline: 26373696]
57. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A 'SMART' design for building individualized treatment sequences. Annu Rev Clin Psychol 2012;8:21-48 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032511-143152] [Medline: 22224838]

Abbreviations

BMI: body mass index
CSIRO: Commonwealth Scientific and Industrial Research Organisation
MET: metabolic equivalent
SD: standard deviation
SWLS: Satisfaction With Life Scale
TAM: Technology Acceptance Model
WLSE: Weight Loss Self-Efficacy

Edited by L Buis; submitted 22.11.18; peer-reviewed by R Kadri, C Eisenhauer; comments to author 07.01.19; revised version received 29.04.19; accepted 05.06.19; published 04.09.19.

Please cite as:
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;7(9):e12882
URL: https://mhealth.jmir.org/2019/9/e12882/
doi:10.2196/12882
PMID:31486407

Original Paper

Comparing the Efficacy of an Identical, Tailored Smoking Cessation Intervention Delivered by Mobile Text Messaging Versus Email: Randomized Controlled Trial

Inger Torhild Gram[1,2], MD, PhD; Dillys Larbi[1], MSc, MPH; Silje Camilla Wangberg[3], PsyD, PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[2]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[3]Department of Health and Caring Sciences, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Inger Torhild Gram, MD, PhD
Norwegian Centre for E-health Research
University Hospital of North Norway
SIVA Innovation Centre Tromsø
Sykehusvn 23
Tromsø, 9038
Norway
Phone: 47 92401177
Email: inger.gram@ehealthresearch.no

Abstract

Background: There is a need to deliver smoking cessation support at a population level, both in developed and developing countries. Studies on internet-based and mobile phone–based smoking cessation interventions have shown that these methods can be as effective as other methods of support, and they can have a wider reach at a lower cost.

Objective: This randomized controlled trial (RCT) aimed to compare, on a population level, the efficacy of an identical, tailored smoking cessation intervention delivered by mobile text messaging versus email.

Methods: We conducted a nationwide 2-arm, double-blinded, fully automated RCT, close to a real-world setting, in Norway. We did not offer incentives to increase participation and adherence or to decrease loss to follow-up. We recruited users of the website, slutta.no, an open, free, multi-component Norwegian internet-based smoking cessation program, from May 2010 until October 2012. Enrolled smokers were considered as having completed a time point regardless of their response status if it was 1, 3, 6, or 12 months post cessation. We assessed 7315 participants using the following inclusion criteria: knowledge of the Norwegian language, age 16 years or older, ownership of a Norwegian cell phone, having an email account, current cigarette smoker, willingness to set a cessation date within 14 days (mandatory), and completion of a baseline questionnaire for tailoring algorithms. Altogether, 6137 participants were eligible for the study and 4378 participants (71.33%) provided informed consent to participate in the smoking cessation trial. We calculated the response rates for participants at the completed 1, 3, 6, and 12 months post cessation. For each arm, we conducted an intention-to-treat (ITT) analysis for each completed time point. The main outcome was 7-day self-reported point prevalence abstinence (PPA) at the completed 6 months post cessation. We calculated effect size of the 7-day self-reported PPA in the text message arm compared with the email arm as odds ratios (ORs) with 95% CIs for the 4 time points post cessation.

Results: At 6 months follow-up, 21.06% (384/1823) of participants in the text message arm and 18.62% (333/1788) in the email arm responded (P=.07) to the surveys. In the ITT analysis, 11.46% (209/1823) of participants in the text message arm compared with 10.96% (196/1788) in the email arm (OR 1.05, 95% CI 0.86-1.30) reported to have achieved 7 days PPA.

Conclusions: This nationwide, double-blinded, large, fully automated RCT found that 1 in 9 enrolled smokers reported 7-day PPA in both arms, 6 months post cessation. Our study found that identical smoking cessation interventions delivered by mobile text messaging and email may be equally successful at a population level.

Trial Registration: ClinicalTrials.gov NCT01103427; https://clinicaltrials.gov/ct2/show/NCT01103427

(JMIR Mhealth Uhealth 2019;7(9):e12137) doi:10.2196/12137

KEYWORDS

eHealth; electronic mail; mHealth; mobile phones; randomized controlled trial; smoking cessation; text-messaging

Introduction

Background

Tobacco use is, and has been for many years, one of the leading preventable causes of disease and death. The number of diseases that are established to be smoking related continues to increase [1,2]. Tobacco consumption is decreasing in the developed countries but increasing in the developing countries [3,4]. Although a high proportion of smokers will try to quit, only 2% to 3% will be successful each year [5].

There is a need to deliver smoking cessation support at a population level, both in developed and developing countries.

Studies on internet-based and mobile phone–based smoking cessation interventions have shown that these methods can be as effective as other methods of support, and they can have a wider reach at a lower cost [6-14]. The randomized controlled trials (RCTs) included in the most recent Cochrane reviews on internet-based [13] and mobile phone–based [12] smoking cessation interventions most frequently compare the effect of the intervention with the comparing condition at 6 months post cessation.

Several of these RCTs had different incentives to increase participation and decrease the loss to follow-up. This could be multiple follow-ups using the internet, email, or mobile phone if users did not respond [14-17], by payment for mobile phone use [18,19], by free Nicotine Replacement Therapy [15,20], by gift certificates [14], and by internet-based counseling from nurses [21] or tobacco treatment specialists [22]. However, as pointed out by Eysenbach, electronic health (eHealth) research studies with a high dropout or high loss to follow-up should not be looked upon as failures but rather a natural and typical feature of eHealth interventions that should be expected [23].

In a previous smoking cessation intervention RCT, we compared tailored with nontailored cessation support delivered by email. At 12 months follow-up, 11.2% of the 419 participants who had received the tailored email reported smoking cessation with similar results in the nontailored email arm [24]. This inspired us to design another RCT, at a population level, that would be fully automated, close to a real-world setting, and have high privacy protection. We decided to follow the recommendations from the Society for Research on Nicotine and Tobacco subcommittee on biochemical verification; that large-scale population studies are one of few settings for which biochemical verification is not required and may not be desirable [25]. We wanted to compare the efficacy of tailored smoking cessation support delivered by 2 modalities: text messages and emails.

Both email and mobile phone text message deliveries are easy to set up. The 2 methods have different strengths and weaknesses. Emails are easily deployable, inexpensive, and can deliver long, complex messages. Text messages may have some special advantages for delivering health behavioral interventions compared with emails. Mobile phones are now considered essential, everyday items and are owned by most adults. The *always-with-you* nature of the mobile phone and the intrusiveness or *push* factor of text messages makes this a simple, low-commitment way to receive smoking cessation support. One disadvantage with text messages is that there may be a fee. Another disadvantage may be that they are limited to 160 characters of text. However, in a previous study about diabetes education, the participants reported that they perceived the text messages as urgent and that the shorter format made the messages easier to understand and remember [26]. As we were not sure whether a short message was an advantage or a disadvantage, we decided to use the same tailored messages in the 2 arms.

Objective

This RCT aimed to compare, on a population level, the efficacy of an identical, tailored smoking cessation intervention delivered by mobile text messaging versus email. We hypothesized that smokers who were allocated to the text message arm compared with the email arm would be more or equally successful at achieving 7-day self-reported point prevalence abstinence (PPA) at the completed 6 months post cessation.

Methods

Trial Design

We conducted a nationwide 2-arm, double-blinded, fully automated RCT, close to a *real-world setting*, in Norway. We did not offer incentives such as free medication, other gifts, or personal counseling to increase the participation and adherence rate and to decrease the loss to follow-up. We did not request biochemical verification of smoking cessation.

Recruitment

We recruited from smokers using an open, free, multi-component Norwegian internet-based smoking cessation program, from May 2010 until October 2012. This website was a part of the Norwegian Directorate of Health's program to promote smoking cessation. The Directorate promoted the website through newspapers, internet, radio, and television (public service announcements) 3 times during the trial period. Enrolled smokers were counted as having completed a time point regardless of their response status if it was 1, 3, 6, or 12 months post cessation. The Regional Committee for Medical Research Ethics approved the study.

Participants

At the start of enrollment, an estimated 94% of the adult population had access to the internet in their homes and 96% owned a mobile phone [27]. We assessed 7315 participants using the following inclusion criteria: knowledge of the Norwegian language, aged 16 years or older, ownership of a Norwegian cell phone, having an email account, current cigarette smoker, willingness to set a cessation date within 14 days (mandatory), and completion of a baseline questionnaire for tailoring algorithms. Altogether 4378 of the 6137 participants

(71.33%) who were eligible for the study provided informed consent for the smoking cessation trial.

Randomization and Blinding

A Web-based online random number generator [28] automatically assigned the participants to a text message arm or email arm. The study was double-blinded at enrollment, so neither the participants nor researchers knew in which arm the participants were enrolled. We do not know if any of the participants discovered during the trial that the purpose of the study was to compare the efficacy of tailored messages delivered by text versus email or about their allocation.

We excluded 53 participants (29 consent withdrawn and 24 missing information and double allocation). The remaining 4335 participants took part in the RCT with 2188 (50.47%) participants in the text message arm.

Implementation

We used Drupal version 6 [29], an open source content management system, to create an automated system that performed all the study procedures (informed consent, registration, randomization, baseline and follow-up questionnaires, and intervention messages). To protect privacy, the data management system consisted of 2 dispatcher servers (A and B).

Dispatcher A sent emails containing a hyperlink to a baseline questionnaire on day 1 and follow-up questionnaires at completed 1, 3, 6, and 12 months post cessation. If there was no response to the questionnaires after 7 days, Dispatcher A sent one reminder email with the same hyperlink.

Dispatcher B applied an algorithm to the responses from the baseline questionnaire and created tailored smoking cessation advice that was delivered either by text message or email. We have described the tailoring algorithm in detail elsewhere [16]. Table 1 shows examples of messages related to personalization and cessation date from the intervention.

Table 1. Examples from the intervention.

Type of message and time of delivery	Question	Answer	Responses
Personalization			
At enrollment	What would you like us to call you?	Jane	—[a]
180 days after cessation date	—	—	Congratulations, Jane. Today you have been smoke free for a half year!
Cessation date			
At enrollment	When do you intend to stop smoking?	Exact date	—
5 days after cessation date	—	—	There is no longer nicotine present in your body
Step down			
10 days before cessation	Would you like to do a step-down of your smoking?	Yes	Create a smoke free zone in your home
Descriptive			
2 days before cessation date	Are you currently working?	Yes, working full time	Consider which situations at work that is tempting you to smoke
Social pressure			
58 days after cessation	Do your friends smoke?	Yes, all of them	Watch out! Some of them might like it if you fail. It could make them feel better

[a]Not applicable.

Dispatcher B created a maximum of 150 individually tailored messages. It delivered the first message 14 days before and the last message 12 months after the stated cessation date. Dispatcher B sent daily messages in the beginning, then the frequency decreased gradually during the first 3 months with a substantial fall in frequency after that. More than half of the messages had been sent to the participants 3 months after the cessation date.

The participants in both arms could read the tailored advice directly without logging on to the website. All users had a personal profile on the website showing their progress, that is, days abstained from smoking, money saved, number of cigarettes not smoked, days since last cigarette, today's advice, cessation calendar with previous advice, and an overview of the social network features. The users could participate in social networking with discussion forums, post questions and advice, and read questions and answers from other users. The users could invite friends (smokers and nonsmokers) to post encouraging messages to them.

Baseline Registration and Data Collection

The baseline questionnaire asked about sex, age, education in years (0-9, 10-12, 13-16, >16), occupational status (8 categories), number of previous cessation attempts, motivation to cessation (4-point scale), and nicotine dependence as measured by the Fagerstrom Test for Nicotine Dependence (6 items, 10-point scale) [30]. It was optional to answer the descriptive background questions. We had less than 5% missing data for the different questions, except age. A technical error caused the system to not record the age variable correctly at enrollment. We re-introduced this variable as a mandatory question in the baseline questionnaire. Age as an inclusion criterion was not disturbed by this technical error. For each user, the program automatically gathered the total number of log-ins to the website, use of the forum (yes, no), posting new topics (yes, no), replies (yes, no), diary entries (yes, no), and number of entries in another person's guestbook.

Outcomes

We calculated response rates and 7-day self-reported PPA for enrolled smokers at completed 1, 3, 6, and 12 months post cessation. The main outcome was the 7-day self-reported PPA at 6 months post cessation. PPA is an assessment of cessation status at a particular point in time when these questions are asked. It is independent of previous answers from the participants. We used the following 2 questions: *Are you currently smoking?* and *Have you been smoking, even as little as one single puff during the past 7 days?* Those who answered "No" to both questions had achieved 7-day self-reported PPA at that specific time point.

Sample Size

A total number of 540 participants were needed per arm at 12 months post cessation to detect a difference of 5% for 7-day PPA (ie, 15% vs 10%) based on a significance level of 5% and 80% power. We did an interim analysis, almost 2 years into the study. The results showed that the enrollment of smokers had been much slower, and the difference between the 2 arms was smaller than anticipated. We therefore extended the enrollment period by 6 months until October 2012. We also changed the time point for the main outcome to 6 instead of 12 months post cessation so we would have a larger sample size and more power to detect a real difference between the 2 arms.

Statistical Methods

We recruited smokers continuously, so the number of enrolled smokers in the study and the number of participants who had completed each time point varied throughout the study period. For each arm, we calculated the response rate for the 4 (1, 3, 6, and 12 months) post cessation time points, as the number of participants who had responded to the email questionnaire at that time point divided by all enrolled smokers who had completed that time point. For each arm, we conducted an intention-to-treat (ITT) analysis and calculated the 7-day self-reported PPA for the completed time points. We calculated the number of participants who reported to have achieved 7-day PPA divided by all enrolled smokers who had completed that time point. This means that all nonresponders, who had completed a time point, were counted as smokers. We used chi-square test statistics to compare, by arm, the selected characteristics at baseline and the time point–specific response rates. We calculated effect size of the 7-day self-reported PPA in the text message arm compared with the email arm as odds ratios (ORs) with 95% CIs for the 4 time points post cessation. A 2-sided *P* value of <.05 was considered statistically significant. All analyses were conducted using IBM SPSS Statistics version 21.

Results

Study Population

Baseline data were available for 4335 (50.5% text message arm) smokers. At enrollment, the median age was 39 years for both arms. In the text message arm (n=334), the age range was from 16 to 72 years, and in the email arm (n=338), it was from 16 to 71 years.

Table 2 shows that more than 70% of the participants were females, more than 60% reported to have at least 13 years of education, and the majority was employed full time. The table shows that the distribution of the selected characteristics at baseline did not vary according to study arm (all *P* values >.13), confirming that the randomization process had worked as it should.

The use of the website's guestbook, diary, and forum and number of log-ins were low and did not differ between the 2 arms (data not shown).

Consolidated Standards of Reporting Trials' Diagram

Figure 1 shows that the response rates were higher in the text message arm compared with the email arm at 1 and 3 months (both *P* values <.05) but not at 6 months (*P*=.07; Figure 1). At 12 months post cessation, the response rate was 25.3% (238/941) in the text message arm and 22.7% (210/927) in the email arm for participants who had completed that time point (*P*=.18).

Table 2. Distribution of selected characteristics at baseline (N=4335) by study arm.

Selected characteristics[a]	Total, n (%)	Text message arm, n (%)	Email arm, n (%)	*P* value[b]
Sex (N=4335)		2188 (50.47)	2147 (49.53)	.41
Male	1214 (28.00)	626 (28.00)	588 (27.38)	
Female	3121 (71.99)	1562 (72.00)	1559 (72.61)	
Education (years; N=4322)[c]		2180 (50.44)	2142 (49.56)	.56
0-9	305 (7.1[d])	161 (7.38)	144 (6.7)	
0-12	1446 (33.46)	722 (33.11)	724 (33.8)	
13-16	1707 (39.50)	848 (38.89)	859 (40.1)	
>16	864 (20.0)	449 (20.59)	415 (19.4)	
Occupation (N=4334)		2188 (50.49)	2146 (49.95)	.70
Employed, full-time	2423 (55.91[d])	1221 (55.91)	1202 (56.01)	
Employed, part-time	543 (12.52)	263 (12.02)	280 (13.0)	
Retired	32 (0.73)	13 (<1)	19 (<1)	
Home keeper	69 (1.59)	35 (2)	34 (2)	
Student	490 (11.30)	265 (12.1)	225 (10.5)	
Disability	266 (6.1)	133 (6.1)	133 (6.2)	
Rehabilitation	267 (6.2)	138 (6.3)	129 (6.0)	
Unemployed	244 (5.6)	120 (5.5)	124 (5.8)	
Cessation attempts (N=4335)		2188 (50.47)	2147 (49.53)	.47
Never	651 (15.0)	324 (14.80)	327 (15.2)	
Once	728 (16.8)	370 (16.91)	358 (16.7)	
Twice	992 (22.9)	505 (23.08)	487 (22.7)	
3 times	646 (14.9)	309 (14.1)	337 (15.7)	
>3 times	1318 (30.40)	680 (31.1)	638 (29.7)	
Motivation score (N=4319)[c]		2181(50.50)	2138(49.50)	.14
1 (very weak)	74 (2)	46 (2)	28 (1)	
2 (pretty weak)	483 (11.2)	235 (10.8)	248 (11.6)	
3 (pretty strong)	2691 (62.31)	1374 (63.00)	1317 (61.60)	
4 (very strong)	1071 (24.80)	526 (24.1)	545 (25.5)	
Fagerstrom Test for Nicotine Dependence score (N=4237)[c]		2138 (50.47)	2098 (49.53)	.72
0-3 Low	1245 (29.39)	622 (29.1)	623 (29.7)	
4-6 Medium	2725 (64.33)	1386 (64.83)	1339 (63.82)	
7-10 High	266 (6.3)	130 (6.1)	136 (6.5)	

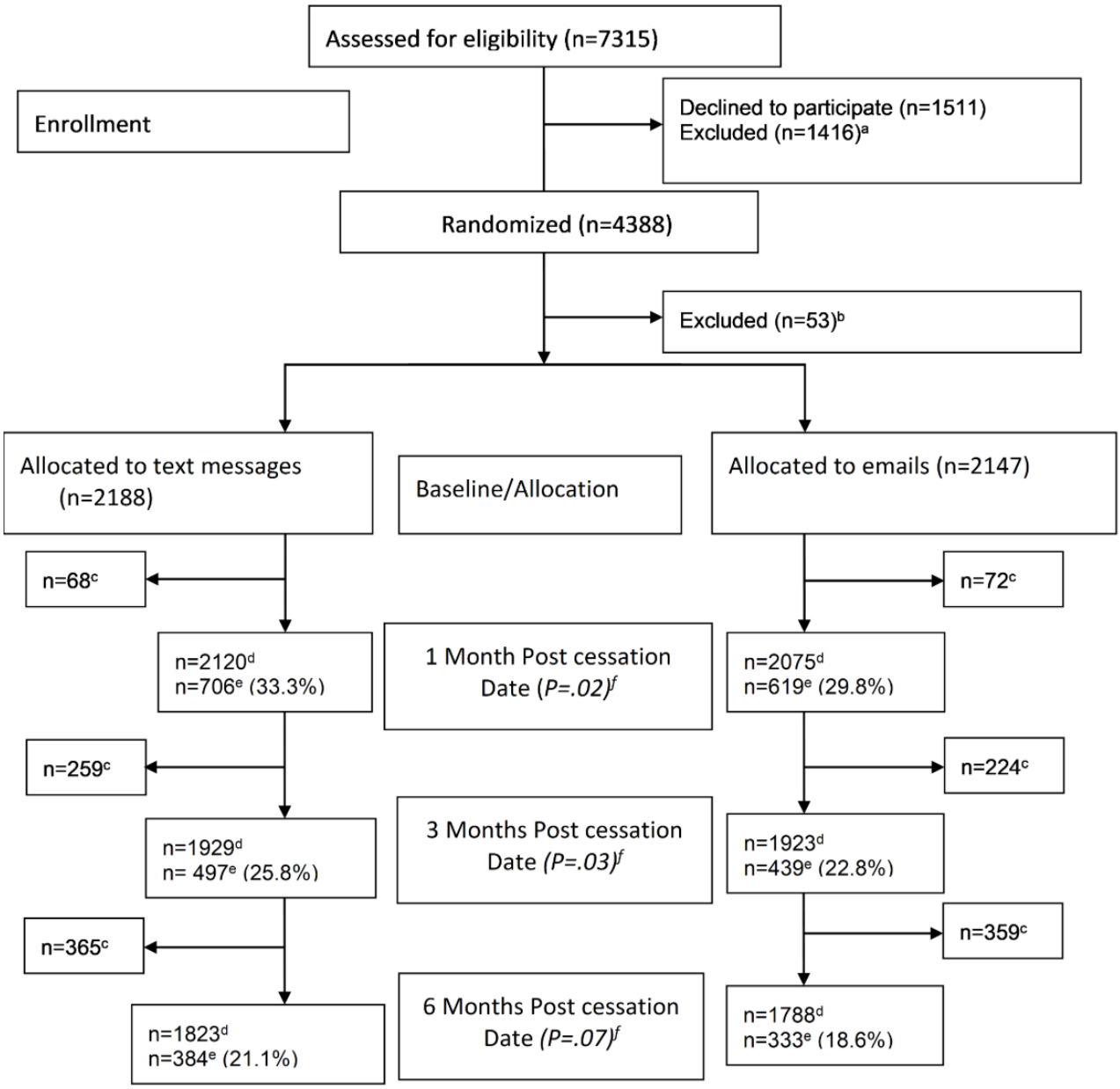
[a]Given as numbers (%).

[b]Chi-square statistics; *P* value for difference between the 2 arms according to sex, education, occupation, motivation score and Fagerstrom Test for Nicotine Dependence score.

[c]Some numbers vary owing to missing values.

[d]Some percentages add up to more than 100 owing to rounding.

Figure 1. Consolidated Standards of Reporting Trials' diagram. Randomized controlled trial, Norway 2010-2012 (N= 4335). a – already stopped smoking (n=631); did not complete baseline registration (n=517); not smoking cigarettes (n=20); referred to substudy (n=248). b – text message/email arm; consent withdrawn n=29 (17/12); missing/double allocation n=24 (12/12). c – participants that had not completed the next follow-up time point. d – Participants included in the analysis. e – responders to follow-up email questionnaire. f – Chi-square statistics; P value for difference between the 2 arms.



Smoking Cessation

Table 3 shows that 11.46% (209/1823) in the text message arm compared with 10.96% (196/1788) in the email arm reported 7-day PPA at 6 months post cessation (OR 1.05, 95% CI 0.86-1.30). A similar ITT analysis for the 12-month post cessation time point revealed that the 7-day self-reported PPA was 12.2% (115/941) in the text message arm and 13.6% (126/927) in the email arm (OR .89, 95% CI 0.68-1.16).

Table 3. Seven-day self-reported point prevalence abstinence (PPA), among enrolled smokers who had completed the 1-, 3-, and 6-month post cessation time point by arm and the corresponding likelihood odds ratio (95% CI) comparing the text message arm with the email arm.

Completed time point post cessation	7-day self-reported PPA (%) in text message arm divided by total enrolled[a] (N=2188), n (%)	7-day self-reported PPA (%) in email arm divided by total enrolled[a] (N=2147), n (%)	Email arm reference at the corresponding time point	Likelihood odds ratio (OR) with 95% CI) In text message arm compared with email arm
1 month	2120 (19.1)	2075 (19.0)	1.00	1.01 (0.86-1.18)
3 months	1929 (14.6)	1923 (14.6)	1.00	1.00 (0.84-1.20)
6 months	1823 (11.5)	1788 (11.0)	1.00	1.05 (0.86-1.30)

[a]Intention-to-treat analyses.

Discussion

Principal Findings

The main result from this large, nationwide, double-blinded RCT was that for those who had completed the time point at 6 months post cessation, the identical program delivered by text messages and emails was equally effective at supporting smoking cessation. In both arms, 1 in 9 enrolled smokers had achieved 7-day self-reported PPA at 6 months post cessation. Similarly, the response rate to the program was 1 in 5 enrolled smokers in both arms at this time point. Furthermore, this RCT conducted at a population level, close to a real-world setting, found that smokers may successfully achieve 7-day PPA at 6 months post cessation without having received incentives such as free medication, other gifts, or personal counseling. Another finding was that a smoking cessation intervention RCT on a population level scale can be fully automated. The program also had a long-term component of 12 months which very few smoking intervention RCTs have.

We find it promising that the tailored interventions delivered by text messages were equally successful as those delivered by email at both 6 and 12 months post cessation.

Comparison With Past Work

To our knowledge, only the UK txt2stop RCT [19], with close to 3000 participants in the intervention arm that received smoking cessation text messages, is larger than our comparable text message arm. In the UK trial, at 6 months post cessation, the ITT analyses revealed that the smoking cessation rate was doubled in the intervention arm (9%) compared with the control arm (4%). The latter received text messages unrelated to quitting [19]. This RCT used continuous smoking abstinence that had biochemical verification.

The previously referred Cochrane review on mobile phone interventions included a total of 12 RCTs. The RCTs varied in how they measured the smoking abstinence outcome from how the UK trial provided the 6 months post cessation outcome to how we measured it. The overall result from the Cochrane meta-analysis showed that 1 in 11 (9%) smokers with support from text messages and 1 in 18 smokers with no program support managed to be abstinent at 6 months post cessation. In total, 9 of the 12 RCT studies enrolled less than 500 persons in each arm and all the studies stopped at 6 months post cessation [9].

In another recent review on mobile phone interventions, 17 of the 20 studies (85%) included had follow-up that was shorter than 6 months post cessation [11].

Our RCT had, in each arm, more than 300 responders at 6 months and more than 200 responders at the 12-month post cessation time point, with close to 1000 participants in each arm that had completed the 12-month post cessation time point. We find it motivating that neither the loss to follow-up nor the achieved 7-day self-reported PPA declined from the completed 6- to 12-month post cessation time point in our trial.

Muench et al have discussed the beneficial features of mobile phone text messages as a tool for smoking interventions. They find that text messages are perceived as more of a personal form of communication and are more likely to be read quickly, to be understood, and responded to upon receipt, compared with emails that are often not viewed by individuals upon receipt [31]. Some participants in the UK RCT reported that text messages about smoking in the intervention arm did stimulate craving [32]. In our study, both arms received smoking cessation advice and could see encouraging messages if they logged on to the website, according to their smoking cessation status.

The anticipated beneficial features of the mobile phone text messages compared with emails did neither result in a different response rate nor a different achieved 7-day self-reported PPA at 6 months post cessation.

Strengths

The main strengths are that our RCT is nationwide, double-blinded, large, fully automated and conducted close to a real-world setting. We believe that these features are important requirements for any smoking cessation intervention at a population level. We were able to show that a large RCT could be fully automated so that the researchers did not have to interact with the participants.

All our efficacy comparisons are from ITT analyses, and the results should be considered to be conservative measures of the effect of the smoking cessation intervention [23]. We have a high internal validity for comparing the 2 different delivery methods, as the messages in the text message and email arms were identical. We also consider as strengths the computerized randomization and the 2 dispatcher servers for privacy protection.

Limitations

Our study has several limitations. One limitation is the loss to follow-up and another the low website adherence. However, neither of these differed by arm, making it unlikely that the comparison results are biased. In addition, we use ITT analyses to avoid this bias. The ITT approach reduces the power to detect differences between the 2 arms, therefore increasing the likelihood of not revealing a true difference. Furthermore, we did not utilize the email capabilities of longer and more complex smoking cessation messages as we did not know if this was an advantage or not. It can be argued that this creates an artificial situation that sacrifices external validity for internal validity. It will always be a trade-off between internal (control) and external (allowing real-world applications) validity. In this study, we decided to have the 2 arms as similar as possible and to focus on the delivery methods.

We experienced a technical error during our trial, as they did in the study by Westmaas et al [14]. We consider continuous technical monitoring and support to be crucial, so technical errors can be discovered and fixed when they occur.

During the last part of our trial, smartphones with email functionality that the normal mobile phone did not have, became more common in Norway. This converging of technologies may have blurred the distinction between the emails and text messages during the last part of the study.

Implications for Future Research

Our study included only Norwegians, of which the majority had more than a high school level of education. Norway has had a good and strict tobacco control policy for many years [33]. Thus, we do not know if our results may be generalized to other racial and ethnic groups, to those with less education, or to those living in countries with no or limited tobacco control policy. In developed countries, most smokers have both a mobile phone and an email account, but this may not be the case in developing countries. Our results are promising, as text messaging is used by most adults in both the developed and developing countries. We encourage the further study of mobile phone–based smoking cessation interventions in low- and middle-income countries.

Conclusions

This nationwide, double-blinded, large, fully automated RCT found that 1 in 9 enrolled smokers reported 7-day PPA in both arms, 6 months post cessation. Our study found that identical smoking cessation interventions delivered by mobile text messaging and emails may be equally successful at a population level.

Acknowledgments

The project was funded by a grant from the Northern Norway Regional Health Authority (TFP743-08) and by the Norwegian Center for E-health Research. The authors would like to thank the Norwegian Directorate of Health's Quitline (*Røyketelefonen*) for their contributions in administering and promoting the website during the trial period. Today, the Norwegian Directorate of Health owns the website, and runs and operates the mobile app, slutta. Lastly, the authors would like to thank all the smokers that enrolled in the study.

Authors' Contributions

ITG and SCW developed the concept and design of the study. All authors contributed substantially to the interpretation of the data, drafting, and revising the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File)2098 KB - mhealth_v7i9e12137_app1.pdf]

References

1. International Agency for Research on Cancer. A Review of Human Carcinogens: Personal Habits and Indoor Combustions. Volume 100E - IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon, France: International Agency for Research on Cancer; 2012.
2. US Department of Health and Human Services. Centers for Disease Control and Prevention. 2014. 2014 Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress URL: https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/ [accessed 2019-08-28]
3. Giovino GA, Mirza SA, Samet JM, Gupta PC, Jarvis MJ, Bhala N, GATS Collaborative Group. Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. Lancet 2012 Aug 18;380(9842):668-679. [doi: 10.1016/S0140-6736(12)61085-X] [Medline: 22901888]
4. Eriksen MP, Mackay J, Schluger N, Islami F, Drope J. The Tobacco Atlas. Fifth Edition. Atlanta, GA: American Cancer Society; 2015.

5. Aveyard P, West R. Managing smoking cessation. Br Med J 2007 Jul 7;335(7609):37-41 [FREE Full text] [doi: 10.1136/bmj.39252.591806.47] [Medline: 17615224]
6. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2009 Oct 7;4:CD006611. [doi: 10.1002/14651858.CD006611.pub2] [Medline: 19821377]
7. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010 Sep 8(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
8. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
9. Civljak M, Stead LF, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013 Jul 10(7):CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
10. 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]
11. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
12. 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]
13. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 4;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
14. Westmaas JL, Bontemps-Jones J, Hendricks PS, Kim J, Abroms LC. Randomised controlled trial of stand-alone tailored emails for smoking cessation. Tob Control 2018 Mar;27(2):136-146. [doi: 10.1136/tobaccocontrol-2016-053056] [Medline: 28522745]
15. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-84; discussion 485. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
16. Brendryen H, Drozd F, Kraft P. A digital smoking cessation program delivered through internet and cell phone without nicotine replacement (happy ending): randomized controlled trial. J Med Internet Res 2008 Nov 28;10(5):e51 [FREE Full text] [doi: 10.2196/jmir.1005] [Medline: 19087949]
17. Borland R, Balmford J, Benda P. Population-level effects of automated smoking cessation help programs: a randomized controlled trial. Addiction 2013 Mar;108(3):618-628. [doi: 10.1111/j.1360-0443.2012.04091.x] [Medline: 22994457]
18. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
19. 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]
20. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS. Improving adherence to smoking cessation treatment: smoking outcomes in a web-based randomized trial. Ann Behav Med 2018 Mar 15;52(4):331-341 [FREE Full text] [doi: 10.1093/abm/kax023] [Medline: 29878062]
21. Smit ES, Candel MJ, Hoving C, de Vries H. Results of the PAS study: a randomized controlled trial evaluating the effectiveness of a web-based multiple tailored smoking cessation program combined with tailored counseling by practice nurses. Health Commun 2016 Sep;31(9):1165-1173. [doi: 10.1080/10410236.2015.1049727] [Medline: 26934538]
22. Houston TK, Sadasivam RS, Allison JJ, Ash AS, Ray MN, English TM, et al. Evaluating the QUIT-PRIMO clinical practice ePortal to increase smoker engagement with online cessation interventions: a national hybrid type 2 implementation study. Implement Sci 2015 Nov 2;10:154 [FREE Full text] [doi: 10.1186/s13012-015-0336-8] [Medline: 26525410]
23. 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]
24. Wangberg SC, Nilsen O, Antypas K, Gram IT. Effect of tailoring in an internet-based intervention for smoking cessation: randomized controlled trial. J Med Internet Res 2011 Dec 15;13(4):e121 [FREE Full text] [doi: 10.2196/jmir.1605] [Medline: 22169631]
25. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
26. Wangberg SC, Arsand E, Andersson N. Diabetes education via mobile text messaging. J Telemed Telecare 2006;12(Suppl 1):55-56. [doi: 10.1258/135763306777978515] [Medline: 16884582]
27. Statistisk Sentralbyrå - Statistics Norway. 2013. ICT Usage in Households, 2015, 2nd Quarter URL: https://www.ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus/aar/2015-10-01 [accessed 2019-05-01]
28. Random.org. 2008. Random Sequence Generator URL: https://www.random.org/sequences/ [accessed 2008-01-01]
29. Drupal development group. Drupal (Version 6.0). 2008. URL: https://www.drupal.org/ [accessed 2019-08-28]

30. 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]
31. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]
32. Douglas N, Free C. 'Someone batting in my corner': experiences of smoking-cessation support via text message. Br J Gen Pract 2013 Nov;63(616):e768-e776 [FREE Full text] [doi: 10.3399/bjgp13X674459] [Medline: 24267860]
33. Government of Norway. 2013. A Tobacco-Free Future: National Strategy for Tobacco Control - 2013–2016 URL: https://www.regjeringen.no/contentassets/818bac68f5994a9181a0251032a8685a/national_strategy_tobacco.pdf [accessed 2019-05-01]

Abbreviations

eHealth: electronic health
ITT: intention-to-treat
OR: odds ratio
PPA: point prevalence abstinence
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.09.18; peer-reviewed by L Scott-Sheldon, B Bock, S Moore, R Mercado, A Graham; comments to author 01.04.19; revised version received 23.05.19; accepted 28.06.19; published 27.09.19.

Please cite as:
Gram IT, Larbi D, Wangberg SC
Comparing the Efficacy of an Identical, Tailored Smoking Cessation Intervention Delivered by Mobile Text Messaging Versus Email: Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(9):e12137
URL: https://mhealth.jmir.org/2019/9/e12137
doi:10.2196/12137
PMID:31573935

Original Paper

Using the Unified Theory of Acceptance and Use of Technology (UTAUT) to Investigate the Intention to Use Physical Activity Apps: Cross-Sectional Survey

Di Liu[1,2*], MMedSc; Remina Maimaitijiang[1*], MMedSc; Jing Gu[1,3], PhD; Shuyi Zhong[1], BMedSc; Mengping Zhou[1], BMedSc; Ziyue Wu[1], BMedSc; Ao Luo[1], BMedSc; Cong Lu[1], BMedSc; Yuantao Hao[1,3], PhD

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]School of Public Health, The University of Hong Kong, Hong Kong Special Administrative Region, China

[3]Sun Yat-sen Global Health Institute, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Jing Gu, PhD
School of Public Health
Sun Yat-sen University
No. 74 Zhongshan Road II
Guangzhou,
China
Phone: 86 020 87335413
Email: gujing5@mail.sysu.edu.cn

Abstract

Background: Many university students are lacking adequate physical exercise and are failing to develop physical activity (PA) behaviors in China. PA app use could improve this situation.

Objective: The aim of this study was to use the unified theory of acceptance and use of technology (UTAUT) to investigate the intention to use PA apps among university students in Guangzhou, China, and how body mass index (BMI) moderates the effects of UTAUT in explaining PA app use intention.

Methods: A cross-sectional study was conducted among 1704 university students from different universities in Guangzhou, China. The UTAUT model was used to measure the determinants of intention to use PA apps.

Results: Of the participants, 41.8% (611/1461) intended to use PA apps. All three UTAUT-related scales (performance expectancy, effort expectancy, and social influence) were positively associated with the intention to use PA apps after adjusting for background variables (adjusted odds ratio 1.10-1.31, $P<.001$). The performance expectancy scale had stronger associations with the intention to use PA apps among those whose BMI were beyond normal range compared with those whose BMI were within normal range ($P<.001$).

Conclusions: UTAUT is useful for understanding university students' intention to use PA apps. Potential moderating effects should be kept in mind when designing UTAUT-based interventions to improve PA via app use.

(JMIR Mhealth Uhealth 2019;7(9):e13127) doi:10.2196/13127

KEYWORDS

intention; physical activity apps; university students; UTAUT

Introduction

Physical inactivity is one of the biggest public health issues of the 21st century and has been identified as the fourth leading risk factor of global mortality by the World Health Organization [1,2]. Evidence has shown that regular physical activity (PA) helps balance energy, control weight, and reduce the risk of noncommunicable diseases (eg, diabetes and hypertension) and mental illness (eg, depression) [1]. Physical inactivity can lead to a worsening health condition and deterioration of quality of life [3,4]. Hallal et al [5] collected PA level data of a population aged 15 years or older worldwide and demonstrated that 31.1% were physically inactive and inactivity increased with age in all regions included. Similarly, several studies indicated an

age-related decline in the level of PA throughout life and that the PA pattern in adolescence usually affected the pattern in adulthood [6-9].

Despite the importance of regular PA, many university students are now living in an environment with increased barriers to PA, resulting in a lack of adequate physical exercise and a failure to develop PA behaviors [10-13]. The 2014 National Physique Monitoring Bulletin released by the General Administration of Sport of China indicated that Chinese university student physical fitness has continued to decline [14].

Effective approaches are urgently warranted to improve this situation. Review studies showed that school-based education programs and interclass exercises could effectively promote PA and fitness among younger adolescents [15]. However, such educational programs could be less effective among university students as their PA was usually less regulated by universities. Furthermore, university students have unique characteristics (ie, they are in a transitional period between adolescence and adulthood). Therefore, innovative approaches are vital to engaging more university students in PA programs. Given the widespread use of smartphones by university students, interventions using this technology may provide a viable opportunity to reach this population and deliver interventions. One benefit of mobile health (mHealth) approaches over traditional methods is that interventions can be provided anywhere and at any time, making them potentially more accessible and feasible [16].

International Telecommunication Union reported that by 2015 Chinese people's mobile phone ownership and internet use reached 92.18% and 50.30%, respectively [17]. University students are often early adopters of new consumer technologies such as smartphones and PA apps [18]. The number of smartphone health and fitness apps has dramatically increased in recent years, with more than 17,000 having been developed for the public [19,20]. The mHealth approaches delivered through PA apps can make PA promotion interventions more attractive and interesting [16] by incorporating strategies such as gamification [21], personalization [22], and creating social network and peer support [23]. Meanwhile, the effectiveness of using PA apps to promote PA has been examined [24,25]. A review and meta-analysis demonstrated the positive effects of PA apps on increasing PA and promoting weight loss [26]. As use behavior is directly affected by use intention according to theories in the social sciences and related domains (eg, theory of planned behavior, technology acceptance model, and unified theory of acceptance and use of technology [UTAUT]), promoting use intention before implementing interventions could facilitate intervention promotion [27-29]. Thus, assessing use intention is important.

UTAUT has been used to investigate behavioral intention to use technology and its influencing factors [30]. To the best of our knowledge, there is no published study applying UTAUT to the investigation of PA app use intention in China. As PA app use is a promising measure in health promotion, in this study, we investigated associations between UTAUT-related variables and PA app use intention in Chinese university students. It is noted that body mass index (BMI) has been associated with PA app use in previous studies, so BMI may have an interaction effect with UTAUT-related variables [26]. The hypothesis that BMI would moderate the effects of UTAUT in explaining PA app use intention was also tested.

Methods

Study Design

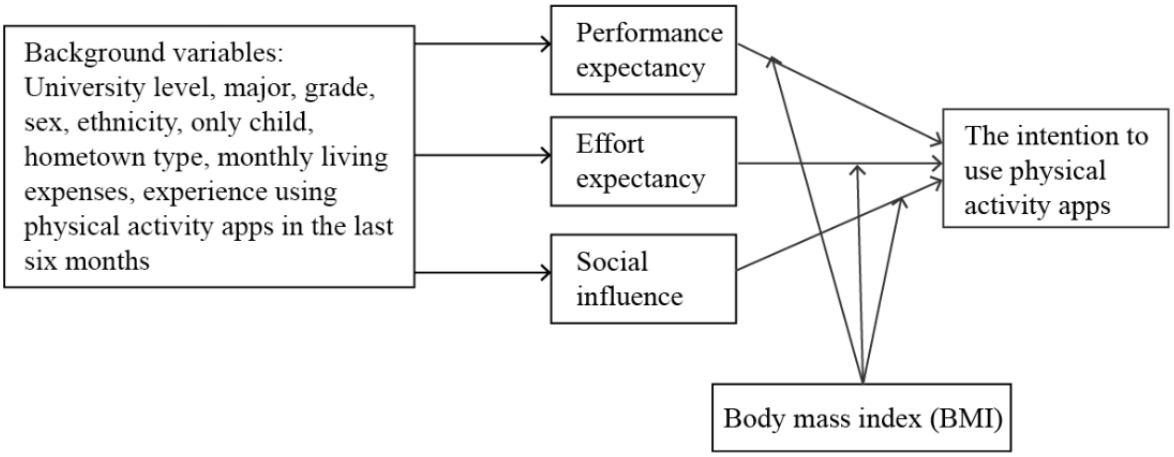
This cross-sectional survey was conducted among university students in Guangzhou, China, from March 1, 2016, to April 20, 2016. A multistage stratified cluster sampling method was used. Selection criteria included all full-time students of universities located in Guangzhou admitted from 2013 to 2015 but excluded students whose majors were sports-related.

Universities in Guangzhou were divided into first and second class according to the Education Examinations Authority of Guangdong Province [31]. Two of the first-class universities and three of the second-class universities were selected using purposive sampling. Student majors from the universities were divided into five categories (natural science, agricultural science, medical science, humanities and social science, and engineering and technology science) according to China's National Classification and Code Disciplines [32]. For each major category, we recruited at least 3 classes from each grade (students admitted in 2015, 2014, and 2013) separately. Contact persons were recruited from the selected universities. After training on study purpose, procedure, data collection, and quality control, they served as helpers to approach different classes and collect data. These contact persons distributed the questionnaires among their classmates, collected the answered questionnaires, and performed a preliminary verification of the quality of all answered questions. Financial reimbursement of 25 yuan (US $4) per hour was provided to them as compensation for their time.

Theoretical Framework

According to UTAUT, a commonly used theory for identifying determinants of intention to use health technologies [29], direct determinants of intention are performance expectancy (PE, the degree to which using a technology will provide benefits in performing certain activities), effort expectancy (EE, the degree of ease associated with the use of the technology), and social influence (SI, the degree to which an individual perceives that important others believe he or she should use the system). Associations between PE, EE, SI, and intention are moderated by variables such as age and voluntariness of use. As PA app use is voluntary and age range does not vary much among university students, moderators were replaced by BMI (Figure 1).

Figure 1. Modified model of unified theory of acceptance and use of technology.



Data Collection

The questionnaire was anonymous and self-administered. Background variables including university, major, grade, sex, ethnicity, being the only child or not, hometown type, monthly living expenses (using 1000 yuan as a cutoff based on living expenses of university students in Guangzhou), and BMI were collected. BMI was calculated by self-reported height and weight, using 18.5 and 24 kg/m^2 as cutoffs. BMI between 18.5 and 24 was defined as normal and BMI outside the range as abnormal (ie, BMI <18.5 means underweight and BMI ≥24 means overweight or obese, as defined by China's Obesity Working Group). We measured intention to use PA apps with one question: Will you use PA apps in the coming 6 months? The experience of having used PA apps before was defined as having used PA apps at least once in the last 6 months. The questionnaire took about 5 to 8 minutes to finish.

PE, EE, and SI were measured with items generated by the research team according to literature and qualitative interviews. An item pool of questions used to measure UTAUT according to literature review was formed, and items were screened by the research group. We interviewed some university students to understand their thoughts on exercise app use according to UTAUT and collect their suggestions on the screened UTAUT items. A revised version of the UTAUT-related questionnaire measuring determinants PE, EE, and SI was developed after the research group discussion.

The determinants were assessed using multi-item measures scored by summing relevant item scores. Each item was scored on a 5-point Likert-type scale (1 = extremely disagree, 5 = extremely agree). Using a principal component analysis, factors were identified for the PE, EE, and SI scales, explaining 80.7%, 82.2%, and 76.3% of total variance, respectively. Cronbach alphas were 0.88, 0.89, and 0.68, respectively.

Statistical Analyses

Univariate logistic regressions were performed to measure the associations between background variables, experience using PA apps in the last 6 months, and intention to use PA apps in the coming 6 months. Background variables with P<.10 in the univariate analyses were selected by a multivariate model using a stepwise method except for experience using PA apps in the last 6 months. Both univariate and multivariate analyses (adjusting for significant background variables and experience using PA apps in the last 6 months) were performed to calculate the association between the UTAUT construct and intention to use PA apps in the coming 6 months.

To identify the interaction effects of BMI on the associations between UTAUT constructs and intention to use PA apps, we performed multivariate logistic regression models, adjusting for significant background variables.

Results

Participant Characteristics

Among the students (from 55 classes) contacted, the response rate was 94.1% (1603/1704), and the effective response rate was 85.7% (1461/1704). Of all participants, 64.8% (947/1461) were from first-tier universities, 50.3% (735/1461) were male, 61.5% (899/1461) were not the only child, 48.9% (714/1461) were from a town or rural area, 67.8% (991/1461) had over 1000-yuan (US $150) monthly living expenses, and 33.3% (487/1461) had a BMI beyond the normal range (24.1%, 352/1461, were lower than the normal range while 9.2%, 135/1461, were overweight or obese; Table 1).

Of the participants, 41.8% (611/1461) intended to use PA apps. In univariate analysis, all background variables except for students' grades and ethnicity were significantly associated with the intention to use PA apps (Table 1). In the multivariate analysis, students who were female (odds ratio [OR] 1.36, 95% CI 1.10-1.68) and from a capital city or municipality (OR 1.43, 95% CI 1.11-1.86) with monthly living expenses were over 1000 yuan (OR 1.44, 95% CI 1.14-1.83) and BMI beyond the normal range (OR 0.71, 95% CI 0.56-0.89) were more likely than others to intend to use PA apps (Table 1). Meanwhile, compared with those not having used PA apps in the last 6 months, participants having used apps in the last 6 months were more likely to intend to use PA apps in the coming 6 months (OR 4.16, P<.001; Table 1).

Table 1. Associations between background variables and the intention to use physical activity apps.

Characteristics	Statistical descriptive, n (%)	Intention to app physical activity apps				
		Yes, n (%)	OR_u[a]	*P* $value_u$	OR_m[b] (95% CI)	*P* $value_m$
University level				**<.001**	**NS**[c]	**NS**
Second-tier	514 (35.2)	181 (35.2)	1.00			
First-tier	947 (64.8)	430 (45.4)	1.53			
Major					**NS**	**NS**
Natural science	378 (25.9)	180 (47.6)	1.00	Ref		
Agricultural science	257 (17.6)	100 (38.9)	0.70	.03		
Medical science	241 (16.5)	95 (39.4)	0.72	.046		
Humanities and social science	157 (10.7)	79 (50.3)	1.11	.57		
Engineering and technology science	428 (29.3)	157 (36.7)	0.64	.002		
Grade					**—**[d]	**—**
Freshman	532 (36.4)	230 (43.2)	1.00	Ref		
Sophomore	513 (35.1)	211 (41.1)	0.92	.49		
Junior	416 (28.5)	170 (40.9)	0.91	.46		
Gender				**.004**		**.005**
Male	735 (50.3)	280 (38.1)	1.00		1.00	
Female	726 (49.7)	331 (45.6)	1.36		1.36 (1.10-1.68)	
Ethnicity				**.75**	**—**	**—**
Han	1369 (93.7)	574 (41.9)	1.00			
Others	92 (6.3)	37 (40.2)	0.93			
Only child				**.75**	**NS**	**NS**
No	899 (61.5)	356 (39.6)	1.00			
Yes	562 (38.5)	255 (45.4)	1.27			
Hometown type						
Town or rural area	714 (48.9)	267 (37.4)	1.00	Ref	1.00	Ref
Noncapital city	377 (25.8)	168 (44.6)	1.35	.02	1.23 (0.95-1.60)	.11
Capital city or municipality	370 (25.3)	176 (47.6)	1.52	.001	1.43 (1.11-1.86)	.007
Monthly living expenses (yuan/month)				**<.001**		**.002**
≤1000	470 (32.2)	161 (34.3)	1.00		1.00	
>1000	991 (67.8)	450 (45.4)	1.60		1.44 (1.14-1.83)	
Body mass index				**.006**		**.003**
18.5-24 kg/m^2	974 (66.7)	435 (44.7)	1.00		1.00	
Beyond 18.5-24 kg/m^2	487 (33.3)	176 (36.1)	0.73		0.71 (0.56-0.89)	
Experience of using physical activity apps in the last 6 months				**<.001**	**NA**[e]	**NA**
No	924 (63.2)	271 (29.3)	1.00			
Yes	537 (36.8)	340 (63.3)	4.16			

[a]Refers to univariate anlyses.

[b]Refers to multivariate analyses.

[c]NS: nonsignificant. Denotes variables with *P*<.10 in the univariate analyses that were not significant in the multivariate analyses.

[d]Denotes variables with *P*>.10 in the univariate analyses that were not used in the subsequent multivariate analyses.

[e]N/A: not applicable. Indicates that the experience of using physical activity apps in the last 6 months was not included in the multivariate analyses.

Associations Between UTAUT-Related Variables and Intention to Use Physical Activity Apps

In the univariate analyses, all 3 UTAUT-related scales were significantly associated with intention to use PA apps (OR_u 1.22-1.49, P<.001). Such associations remained significant after adjusting for significant background variables (gender, hometown type, monthly living expenses, and BMI), and experience of using PA apps in the last 6 months (OR_a 1.10-1.31, P<.001).

In the associations between each UTAUT-related item and the studied outcome, all 8 items were significantly associated with the intention to use PA apps (OR_u 2.16-2.97; Table 2).

Moderating Effects of Body Mass Index on the Associations Between UTAUT and Intention to Use Physical Activity Apps

One out of the three models considered presented statistically significant interaction effects: interaction between the BMI and the performance expectancy scale (beta 0.10, P<.001; Table 3). Higher scores on the PE scale (x-axis) were associated with higher log odds for intention to use PA apps (y-axis), but the strength of associations depended on the BMI, as seen by the slopes of the straight lines (Figure 2). The significant moderating effect indicated stronger associations between the PE scale and the intention to use PA apps among those whose BMI was beyond normal range (BMI≥24), as compared with those whose BMI was within normal range.

Table 2. Associations between UTAUT-related scales and the intention to use physical activity apps.

Scale	OR_u[a] (95% CI)	P $value_u$	OR_a[b] (95% CI)	P $value_a$
Scale 1. Performance expectancy scale	1.26 (1.20-1.32)	<.001	1.16 (1.11-1.22)	<.001
Term 1.1. Using physical activity apps could inspire you to keep doing physical activity.	2.55 (2.06-3.16)			
Term 1.2. Using physical activity apps could contribute to maintaining physical fitness.	2.60 (2.10-3.23)			
Term 1.3. Using physical activity apps could contribute to maintaining good mental health.	2.16 (1.74-2.68)			
Scale 2. Effort expectancy scale	1.22 (1.16-1.27)	<.001	1.10 (1.04-1.15)	<.001
Term 2.1. You can quickly master how to use physical activity apps.	2.57 (2.00-3.30)			
Term 2.2. You can be proficient with using physical activity apps.	2.57 (2.04-3.24)			
Term 2.3. Using physical activity apps is not difficult for you.	2.64 (2.05-3.41)			
Scale 3. Social influence scale	1.49 (1.39-1.60)	<.001	1.31 (1.21-1.42)	<.001
Term 3.1. Your good friends are in favor of your using physical activity apps.	2.32 (1.88-2.87)			
Term 3.2. Many of your friends are using physical activity apps.	2.97 (2.39-3.69)			

[a]Refers to univariate analyses.

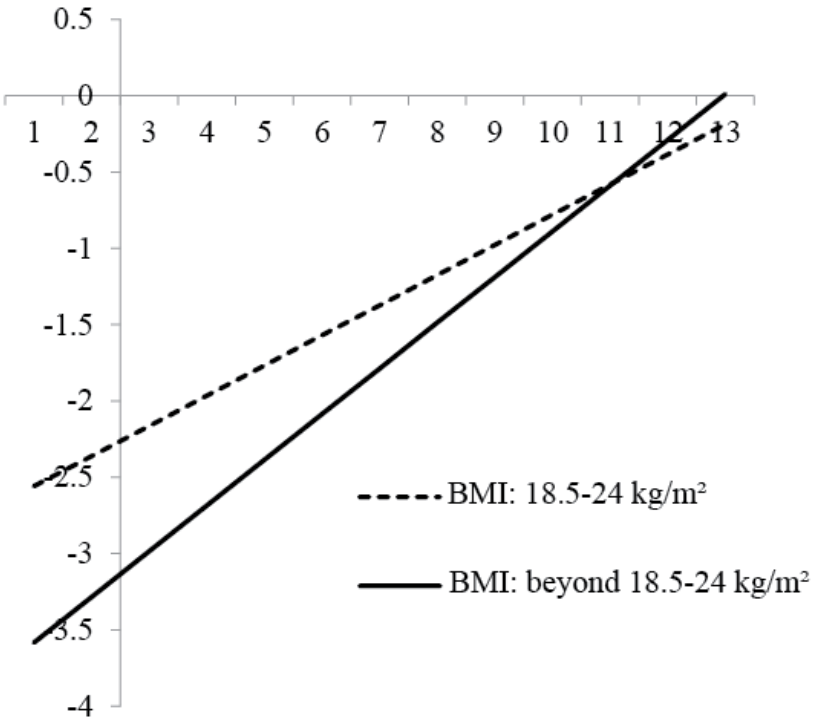
[b]Refers to adjustment for gender, hometown type, monthly living expenses, experience of using physical activity apps in the last 6 months, and body mass index.

Table 3. Summary of logistic regression models testing significance of main and interaction effects of UTAUT-related scales and body mass index.

Model	Beta	SE (beta)	*P* value
Model 1			
Performance expectancy scale	0.20	0.03	.05
BMI[a]	–1.33	0.53	.01
BMI × performance expectancy scale	0.10	0.05	<.001
Model 2			
Effort expectancy scale	0.17	0.03	<.001
BMI	–0.81	0.59	.17
BMI × effort expectancy scale	0.04	0.05	.44
Model 3			
Social influence scale	0.35	0.04	<.001
BMI	–0.88	0.56	.12
BMI × social influence scale	0.08	0.08	.29

[a]BMI: body mass index. BMI was divided into two levels: 0 = normal range of Chinese people (18.5-24 kg/m^2) and 1 = beyond normal range.

Figure 2. Interaction effect between body mass index and performance expectancy scale. BMI: body mass index.



Discussion

Principal Findings

To the best of our knowledge, this study is among the first to examine university students' intention to use PA apps based on a new technology use–related model in China. Overall, university students showed a relatively high level (611/1461, 41.8%) of intention to use PA apps for increasing PA. Moreover, the number of participants having the intention to use PA apps in the coming 6 months was higher than the number of participants having used PA apps in the last 6 months (611 vs 537), which indicated that the intention to use PA apps of university students in Guangzhou showed an upward trend. This relatively strong intention suggests that future efforts to increase PA among university students with PA apps will be promising.

Of all the participants, 24.1% (352/1461) had a BMI lower than the normal range and 9.2% (135/1461) were overweight or obese according to their BMI. The rates of lower body weight and overweight or obese were both higher than the results of research conducted in Henan Province (low body weight: 14.7%;

overweight or obese: 8.5%) [33] and northern Anhui Province (low body weight: 12.6%; overweight or obese: 8.3%) [34] among university students. But the rate of overweight or obesity in our study was a little lower than that found in one study conducted on university students in Zhejiang (low body weight: 8.2%; overweight or obese: 10.9%) [35]. Although there are regional differences in BMI, our results still indicated that the situation of university students' BMI in Guangzhou being beyond the normal range was relatively serious. The physical health status of university students in Guangzhou is thus in urgent need of improvement.

In our study, PE, EE, and SI were all positively related to PA app use intention after adjusting for significant background variables. UTAUT can therefore be used to develop implementation interventions to increase the use of apps designed for improving PA among university students. SI showed a relatively strong association with the intention to use PA apps among the three scales with OR 1.49 (95% CI 1.39-1.60). This may be due to the community lifestyle of university students in China. University students live on campus, spend most of their time in class or in the dormitory, have close contact with their peers (classmates or roommates), and are easily influenced by their peers [36]. A survey conducted in European universities reached a similar conclusion that students' alcohol use behavior was affected by their peers' alcohol use behavior [37]. Social influence mainly refers to the influence from surrounding people and environment, which can explain why social influence plays a more important role in affecting the intention to use PA apps. To promote university students' intention to use PA apps, schools or society may be able to achieve their goals with the aid of peer influence.

UTAUT-based interventions seem to be useful among all students regardless of their BMI. However, the strength of associations between UTAUT and PA app use intention may depend on other contextual factors. UTAUT-based interventions may have a better chance of success if they pay more attention to university students whose BMI are beyond the normal range. Such findings suggest that to improve physical activities among university students via app use, for students with BMI beyond normal range interventions enhancing PE (eg, peers sharing of benefits) could be considered, but for students with normal or low BMI intervention targeting PE alone may be less effective.

Limitations

This study has some limitations. First, the observational design may not establish a causal relationship between independent variables and outcome. Although we collected data of use intention in the coming 6 months and use experience in the last 6 months, which guaranteed the time sequence, participants' intention might still be based mainly on the time point of the survey. It was assumed that cognitive situations would be quite stable in the coming 6 months.

Second, although anonymity and privacy were guaranteed, a reporting bias due to social desirability and self-expectation may still exist. For example, as university students, the participants might give a high score to the EE scale. There were at least two items in each scale to try to avoid this problem, and Cronbach alpha was high for each scale.

Finally, we cannot assume that the results can be extrapolated widely without further research. A purposive sampling method was used to recruit participants, which may weaken the external validity of our sample. In addition, our study involved only one city and this city's economic level is higher than most other cities in China, which might affect participants' acceptability of new technology and health consciousness.

Conclusions

This study, based on a theoretical approach to technology use, indicated which factors will need to be addressed to design an effective implementation intervention for the use of PA apps to increase PA among university students. Our findings indicated that university students' intention to use PA apps was influenced by UTAUT-related constructs, but potential moderating effects of BMI should be kept in mind when UTAUT-based interventions are being developed. Different intervention strategies should be considered for students within and beyond the normal range of BMI.

Acknowledgments

The authors received no financial support for the research, authorship, and publication of this article.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2010. Global recommendations on physical activity for health URL: http://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2018-01-19]
2. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. Br J Sports Med 2009 Jan;43(1):1-2. [Medline: 19136507]
3. Hancox RJ, Milne BJ, Poulton R. Association between child and adolescent television viewing and adult health: a longitudinal birth cohort study. Lancet 2004;364(9430):257-262. [doi: 10.1016/S0140-6736(04)16675-0] [Medline: 15262103]
4. Kimm SYS, Glynn NW, Obarzanek E, Kriska AM, Daniels SR, Barton BA, et al. Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. Lancet 2005;366(9482):301-307 [FREE Full text] [doi: 10.1016/S0140-6736(05)66837-7] [Medline: 16039332]

5. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. 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. Pate RR, Stevens J, Webber LS, Dowda M, Murray DM, Young DR. Age-related change in physical activity in adolescent girls. J Adolesc Health 2009 Mar;44(3):275-282 [FREE Full text] [doi: 10.1016/j.jadohealth.2008.07.003] [Medline: 19237114]
7. Sallis JF. Age-related decline in physical activity: a synthesis of human and animal studies. Med Sci Sports Exerc 2000 Sep;32(9):1598-1600. [Medline: 10994911]
8. Kimm SYS, Glynn NW, Kriska AM, Barton BA, Kronsberg SS, Daniels SR, et al. Decline in physical activity in black girls and white girls during adolescence. N Engl J Med 2002 Sep 05;347(10):709-715. [doi: 10.1056/NEJMoa003277] [Medline: 12213941]
9. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195. [doi: 10.1159/000222244] [Medline: 20054224]
10. American College Health Association–National College Health Assessment II: Thompson Rivers University Executive Summary Spring 2013. Hanover: American College Health Association; 2013. URL: https://www.tru.ca/__shared/assets/NCHA_II_Spring_201329926.pdf [accessed 2018-01-20]
11. Bongaarts J. U.S. health in international perspective: shorter lives, poorer health. Popul Dev Rev 2013;39(1):165-167. [Medline: 24006554]
12. Balon R, Beresin EV, Coverdale JH, Louie AK, Roberts LW. College mental health: a vulnerable population in an environment with systemic deficiencies. Acad Psychiatry 2015 Oct;39(5):495-497. [doi: 10.1007/s40596-015-0390-1] [Medline: 26327172]
13. Anxious students strain college mental health centers. New York: New York Times; 2015 May 27. URL: https://tinyurl.com/y5jz23ac [accessed 2018-01-23]
14. General Administration of Sport in China. [National Students' Physical Health Survey results]. Chin J School Health 2015;12:4. [doi: 10.16835/j.cnki.1000-9817.2015.12.001]
15. Dobbins M, Husson H, DeCorby K, LaRocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev 2013;2:CD007651. [doi: 10.1002/14651858.CD007651.pub2] [Medline: 23450577]
16. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
17. International Telecommunication Union. 2016. Percentage of individuals using the internet URL: https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2016/Individuals_Internet_2000-2015.xls [accessed 2018-01-21]
18. Ma C. Media use and media literacy of urban and rural youth in the digital media era: an empirical investigation of the youth in S Province. J Sichuan Univ Sci Eng Soc Sci Ed 2018;33(5):79-100. [doi: 10.11965/xbew20180506]
19. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
20. Direito A, Jiang Y, Whittaker R, Maddison R. Smartphone apps to improve fitness and increase physical activity among young people: protocol of the Apps for IMproving FITness (AIMFIT) randomized controlled trial. BMC Public Health 2015;15(1):635 [FREE Full text] [doi: 10.1186/s12889-015-1968-y] [Medline: 26159834]
21. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]
22. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
23. 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]
24. 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 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
25. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via facebook app: randomized controlled trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
26. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
27. Compeau DR, Higgins CA. Application of social cognitive theory to training for computer skills. Inform Syst Res 1995 Jun;6(2):118-143. [doi: 10.1287/isre.6.2.118]

28. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
29. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
30. Liu L, Miguel CA, Rios RA, Buttar V, Ranson Q, Goertzen D. What factors determine therapists' acceptance of new technologies for rehabilitation—a study using the Unified Theory of Acceptance and Use of Technology (UTAUT). Disabil Rehabil 2015;37(5):447-455. [doi: 10.3109/09638288.2014.923529] [Medline: 24901351]
31. Education Examinations Authority of Guangdong Province. URL: http://eea.gd.gov.cn [accessed 2019-07-16]
32. National Technical Supervision Bureau. 2006. [Classification and code disciplines of China] URL: http://www.ynnubs.com/newsimges/xls/2010043443.doc [accessed 2018-02-15]
33. Yang ML, Lou XM, Peng YL. [BMI and health-related physical fitness in Henan undergraduate students]. Chin J School Health 2013;34(9):1093-1095. [doi: 10.16835/j.cnki.1000-9817.2013.09.026]
34. Zhao W, Zhou C, Li S. [The influence of study of nutriology on undergraduate' nutritional knowledge-attitude-practice and BMI]. J Molecular Imag 2015;38(4):409-412. [doi: 10.3969/j.issn.1674-4500.2015.04.32]
35. Xiang J, Song Z, Tian C. [Research on the step count and its correlation with BMI and WHR of college students]. Zhejiang Sport Sci 2016;38(2):104-109 [FREE Full text]
36. Choukas-Bradley S, Giletta M, Cohen G, Prinstein M. Peer influence, peer status, and prosocial behavior: an experimental investigation of peer socialization of adolescents' intentions to volunteer. J Youth Adolesc 2015;44(12):2197-2210. [Medline: 26525387]
37. Helmer S, Sebena R, McAlaney J, Petkeviciene J, Salonna F, Lukacs A. Perception of high alcohol use of peers is associated with high personal alcohol use in first-year university students in three central and eastern European countries. Subst Use Misuse 2016;51(9):1224-1231. [Medline: 27219473]

Abbreviations

BMI: body mass index
EE: effort expectancy
PA: physical activity
PE: performance expectancy
OR: odds ratio
SI: social influence
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 13.12.18; peer-reviewed by CJ Ng, G Signorelli, J Alvarez Pitti; comments to author 21.03.19; revised version received 15.05.19; accepted 29.05.19; published 10.09.19.

Please cite as:
Liu D, Maimaitijiang R, Gu J, Zhong S, Zhou M, Wu Z, Luo A, Lu C, Hao Y
Using the Unified Theory of Acceptance and Use of Technology (UTAUT) to Investigate the Intention to Use Physical Activity Apps: Cross-Sectional Survey
JMIR Mhealth Uhealth 2019;7(9):e13127
URL: https://mhealth.jmir.org/2019/9/e13127
doi:10.2196/13127
PMID:31507269

Original Paper

Effects of Mobile Health Prompts on Self-Monitoring and Exercise Behaviors Following a Diabetes Prevention Program: Secondary Analysis From a Randomized Controlled Trial

Megan M MacPherson[1], MSc; Kohle J Merry[2], MSc; Sean R Locke[1], PhD; Mary E Jung[1], PhD

[1]School of Health and Exercise Sciences, University of British Columbia, Kelowna, BC, Canada

[2]Mechatronic Systems Engineering, School of Engineering Science, Simon Fraser University, Surrey, BC, Canada

Corresponding Author:
Mary E Jung, PhD
School of Health and Exercise Sciences
University of British Columbia
Okanagan Campus
3333 University Way
Kelowna, BC, V1V 1V7
Canada
Phone: 1 2508079670
Fax: 1 2508078085
Email: mary.jung@ubc.ca

Abstract

Background: A number of mobile health (mHealth) apps exist that focus specifically on promoting exercise behavior. To increase user engagement, prompts, such as text messages, emails, or push notifications, are often used. To date, little research has been done to understand whether, and for how long, these prompts influence exercise behavior.

Objective: This study aimed to assess the impact of prompts on mHealth self-monitoring and self-reported exercise in the days following a prompt and whether these effects differ based on exercise modality.

Methods: Of the possible 99 adults at risk for developing type II diabetes who participated in a diabetes prevention program, 69 were included in this secondary analysis. Participants were randomly assigned to 1 of the following 2 exercise conditions: high-intensity interval training or moderate-intensity continuous training. In the year following a brief, community-based diabetes prevention program involving counseling and supervised exercise sessions, all participants self-monitored their daily exercise behaviors on an mHealth app in which they were sent personalized prompts at varying frequencies. mHealth self-monitoring and self-reported exercise data from the app were averaged over 1, 3, 5, and 7 days preceding and following a prompt and subsequently compared using *t* tests.

Results: In the year following the diabetes prevention program, self-monitoring (t_{68}=6.82; *P*<.001; d=0.46) and self-reported exercise (t_{68}=2.16; *P*=.03; d=0.38) significantly increased in the 3 days following a prompt compared with the 3 days preceding. Prompts were most effective in the first half of the year, and there were no differences in self-monitoring or self-reported exercise behaviors between exercise modalities (*P* values >.05). In the first half of the year, self-monitoring was significant in the 3 days following a prompt (t_{68}=8.61; *P*<.001; d=0.60), and self-reported exercise was significant in the 3 days (t_{68}=3.7; *P*<.001; d=0.37), 5 days (t_{67}=2.15; *P*=.04; d=0.14), and 7 days (t_{68}=2.46; *P*=.02; d=0.15) following a prompt, whereas no significant changes were found in the second half of the year.

Conclusions: This study provides preliminary evidence regarding the potential influence of prompts on mHealth self-monitoring and self-reported exercise and the duration for which prompts may be effective as exercise behavior change tools. Future studies should determine the optimal prompting frequency for influencing self-reported exercise behaviors. Optimizing prompt frequency can potentially reduce intervention costs and promote user engagement. Furthermore, it can encourage consumers to self-monitor using mHealth technology while ensuring prompts are sent when necessary and effective.

International Registered Report Identifier (IRRID): DERR2-10.2196/11226

(JMIR Mhealth Uhealth 2019;7(9):e12956) doi:10.2196/12956

KEYWORDS

self-monitoring; health behavior; prompts; mHealth; mobile apps; exercise; high-intensity interval training; reminder system

Introduction

Background

Mobile phones are ubiquitous and becoming an integral part of daily life. In 2015, global subscriptions of mobile phones were approximately 7 billion; this constitutes a substantial increase from 738 million subscriptions in 2000 [1]. In addition, 95% of the global population resides in areas covered by cellular networks, the majority of which has the opportunity to access the internet through their mobile devices, as mobile broadband networks (3G or above) reach approximately 84% of the global population [2]. Smartphones are internet-enabled mobile phones that possess a multitude of capabilities through the use of electronic apps, which are specifically developed to be used on a handheld device for various purposes. In fact, a survey by Bender et al [3] examining mobile phone usage among white, Filipinos, Koreans, and Latino Americans found that individuals are more likely to access the internet through mobile phones when compared with computers, and that mobile phone usage did not significantly differ between these groups.

As the widespread adoption of mobile phones increases, so too does the opportunity for the development and implementation of theory-driven, cost-effective, evidence-based mobile phone apps (ie, mobile health [mHealth] app) used to influence health behaviors. An mHealth app is any mobile phone app, which is used for tracking, guiding, teaching, or enabling individuals in any health-related behaviors and can range from tracking diet and exercise to guided meditation or monitoring of diabetic sugar level. The accessibility of these apps is also advantageous for researchers who can monitor consumer behaviors remotely, provide real-time feedback, and aggregate data so as to improve monitoring systems [4]. Despite the rapidly growing number of mHealth apps on the market and the advantages they may afford to consumers and researchers alike, there is a profound lack of theory-driven, evidence-based mHealth apps [5-8].

Mobile Health and Behavior Change Techniques

This lack of evidence-based mHealth apps may be because of the time-consuming nature of conventional methods of evaluation, such as randomized controlled trials (RCTs), being unable to keep up with the dynamic nature of mHealth app development, and the rapid advancement of mobile technologies [9]. One approach to address this issue has been to research the irreducible, replicable, and observable components—known as behavior change techniques (BCTs)—of mHealth interventions [10,11]. BCTs most frequently used within physical activity (PA) mHealth apps include self-monitoring of behavior, feedback on behavior, and prompts or cues [12,13].

Self-monitoring is a commonly used and robust BCT, which often involves participants logging target behaviors [14]. A meta-regression by Michie et al [15] found that interventions, which included self-monitoring, were more effective at improving PA than those that did not. Within mHealth literature, self-monitoring has been shown to improve PA and dietary behaviors [16,17]. Carels et al [18] posit that daily self-monitoring may allow individuals to increase their awareness of the target behavior, thus allowing them to implement strategies to resume a behavior when they become aware that they are not engaging in the target behavior. In support of this, studies have shown that adherence to daily self-monitoring is associated with increased weight loss [19], and self-monitoring of daily exercise is associated with increased PA and weight loss [18]. Furthermore, research has suggested that self-monitoring and adherence to PA goals may be bolstered through the use of personalized prompts or feedback [20]. Specifically, 1 study found that individuals who received personalized goal setting prompts logged significantly more PA than their counterparts who received generic prompts [21].

Prompts within mHealth apps promote individual-app interaction (eg, text messages, multimedia message services, and mobile phone push notifications). There is a growing body of evidence to support the use of prompts as either stand-alone interventions or supplementary features to increase the effectiveness of health interventions [22,23]. Specifically, reviews have shown that prompts may be effective in enhancing diet or weight loss, PA behaviors, and smoking cessation behaviors [24-27]. That said, few interventions parse out and examine the influence of prompts. Prompt interventions targeting health behaviors are often short in duration, lasting less than 14 weeks on average [22,28,29], and vary in the frequency of prompts delivered from daily to weekly or monthly messages [26,30]. This variability in design, coupled with the fact that few studies have reported on or assessed the effectiveness of individual intervention characteristics [26,31], demonstrates that informative research is required to understand ideal message frequency targeting behavior modification.

Purpose

This paper analyzes mobile phone prompt data to promote exercise adherence for 1 year following a diabetes prevention program research study. Program participants were randomized to perform 1 of the following 2 exercise modalities: high-intensity interval training (HIIT) or moderate-intensity continuous training (MICT). HIIT has garnered attention as an exercise program primarily because of its shorter duration and similar cardiometabolic health effects when compared with MICT [32,33]. There may be a differential impact of prompts on cuing the engagement of time-efficient HIIT compared with MICT. Previous studies highlighting the positive impact prompts have on promoting PA have primarily used MICT to examine outcomes such as walking behaviors, daily step count, and sedentary behaviors [22] but have yet to examine the impact on HIIT engagement.

The main objective of this study was to examine whether mHealth prompts influence self-monitoring and self-reported exercise in 1, 3, 5, or 7 days following a prompt. Prompts are meant to provide brief effects; therefore, we hypothesized that there would be initial increases in both mHealth self-monitoring and self-reported exercise behaviors. No specific hypotheses

on whether prompt effects would last 1, 3, 5, or 7 days were made. Given the short follow-up durations of previous research and lack of literature addressing the impact of prompts on exercise prescriptions, we wanted to explore whether the effects of a prompt were consistent in the first and second half of the year following a diabetes prevention program and whether the impact of prompts differed between those randomized to HIIT or MICT.

Methods

Overview

This paper presents a secondary analysis examining the effect of personalized mHealth prompts on self-monitoring and self-reported exercise behaviors within a diabetes prevention program. Complete details regarding the study design, methods, and procedures have been previously published [34]. The program was a 2-week lifestyle modification program aimed at reducing type II diabetes risk (ClinicalTrials.gov; NCT02164474). This program consisted of 7 one-on-one sessions with a trained exercise counselor focusing on brief counseling, self-regulatory skills development, and exercise. A total of 99 individuals (69 of which were included in this secondary analysis) participated in a diabetes prevention program and were randomly assigned to 1 of 2 exercise conditions: HIIT or MICT. HIIT involves alternating bursts of vigorous-intensity exercise with a recovery period of light exercise, whereas MICT encompasses exercising at a steady pace for a longer duration. Following the program completion, all participants were prescribed 3 days of exercise and permitted up to 4 rest days per week to be used at the discretion of the participant. Participants in the MICT group were prescribed 150 min of weekly moderate-intensity exercise (50 min 3 times per week), whereas HIIT participants were prescribed 75 min of vigorous interval exercise (25 min of intervals 3 times per week). To promote exercise adherence in free-living conditions 1 year following the diabetes prevention program, participants were provided with an mHealth app (or paper logbook if they opted not to use the app) to encourage exercise self-monitoring for 1 year immediately following completion of the intervention.

Mobile Health App

The theory-based mHealth app used in the diabetes prevention program was designed using principles from social cognitive theory to help participants self-monitor their exercise behaviors [35]. Participants were encouraged to self-monitor their exercise behaviors (including days in which they did not exercise) through the mHealth app. The app was designed to allow each participant to self-monitor their daily exercise behaviors; possible responses included “yes I exercised today,” “rest day,” and “no I did not exercise today.” If a participant responded with “yes I exercised today,” they were asked additional questions regarding the type, duration, and intensity of their exercise session. Participants were rewarded with points on the app for continual self-monitoring and exercise engagement. Feasibility testing of self-monitoring through this app has demonstrated increased self-monitoring and PA behaviors over an 8-week period for those who used the app when compared with a control group [35].

The messaging platform within this mHealth app allowed for 2-way messaging between participants and their exercise counselors. Participants received personalized messages that encompassed counselors sending name-specific prompts using a series of message templates (Table 1). These messages were based on social cognitive theory and modeled off of those used by Voth et al, which targeted self-monitoring, verbal persuasion, and performance accomplishment [35]. Exercise counselors sent their participants 1 message per month and would respond to participants’ messages with social or instrumental support to reinforce the aforementioned behavior change concepts. Participants were also sent a reminder message to self-monitor if they failed to self-monitor for 3 consecutive days. A *prompt* was defined as any of the above message types in which there was a minimum of 6 days preceding it with no other message. This definition was determined to exclude any subsequent conversation resulting from a prompt in the analyses.

Table 1. Example messages to participants.

Message type	Example message
Reminder to self-monitor	Hi (insert name), I noticed that you have not checked in with your exercise for the past few days. Is everything ok? Let me know how I can help you. You can do this!
Use of verbal persuasion and self-set rewards	Hi (insert name)! Just dropping a note to say how proud I am of all of your hard work over these past few months. Wow - you've been working hard towards being a regular exerciser for over half a year! Very impressive. Few things in life come easy - motivating yourself to exercise consistently is no exception. We are so proud of you for each and every exercise bout you do - because we know firsthand how difficult it is. So keep up the good work! And while you're at it - start acknowledging all of your hard work and REWARD yourself! A bath, a glass of wine, 10 minutes of peace and quiet, whatever it is - give it yourself after you complete your next exercise bout. You deserve it.
Importance of self-monitoring	Wow – how time flies when you're doing fabulously! (insert name) – really impressed with your exercise behaviour, but equally impressed by your faithful check-ins. Keeping tabs on what you're doing keeps you honest, so make sure you continue to self-monitor here. And remember, self-monitoring is most important when you miss a day – so report that if it happens! You're human!
Performance accomplishment	Hey (insert name). I have been watching your progress for the last few weeks and wanted to say congratulations on what an awesome job you have been doing! You should be really proud of yourself – you've been sticking with your exercise plan over the past month! Keep up this fantastic effort and I'll be right here watching your fabulous achievements.
Response to participant (providing social support)	I love your attitude (insert name), and your perseverance! I'm glad you can recognize the changes you have achieved, but also strive for more. Keep pushing through and you will get there!
Response to participant (providing instrumental support)	Hi (insert name), we are having some trouble with the system. I have unlocked yesterday for you so hopefully you can re-enter your exercise and it works! Let me know.

Participants

Of 99 adults participated in a diabetes prevention program, 69 (51 females, 17 males, and 1 missing; mean age 50.7 years, SD 9.4) were included in this analysis. Participants were eligible to participate if they were between the ages of 30 and 65 years, were inactive (defined as engaging in <3 bouts of moderate or vigorous aerobic exercise per week in the past 6 months), had a body mass index between 24 and 40 kg/m^2, and were cleared to engage in vigorous exercise using Canadian Society for Exercise Physiology Physical Activity Readiness Questionnaire-Plus [36]. Participants were asked to provide demographic information including age, ethnicity, highest level of education completed, and current occupational status. Only individuals who chose to self-monitor through the mHealth app were included in this analysis; an additional 30 participants were not included because of the use of paper logbook (n=7), using the app for less than 2 months throughout the 1-year follow-up period (n=14; self-monitored an average of 40 days), and data error (n=9).

Outcome Measures

Outcome measures include frequency of both mHealth self-monitoring and self-reported exercise in the week before and after a prompt. mHealth self-monitoring was defined as any day in which a participant logged on the mHealth app; this includes days in which they engaged in purposeful exercise, rest days, and days in which they did not exercise and exceeded their number of rest days. Self-reported exercise was defined as only those days in which participants logged on the mHealth app that they engaged in purposeful exercise. Specifically, when a participant self-monitored "yes I exercised today," they were able to type in the details of their exercise; however, for the purpose of this study, the level or type of logged exercise was not examined.

Procedures

During the one-on-one counseling sessions, the exercise counselor created each participant's profile on the mHealth app and taught participants how to self-monitor their exercise to ensure participants were confident in their ability to monitor their exercise . Throughout the free-living 1-year follow-up period, participants were sent personalized messages delivered through the app messaging system from their exercise counselor at a variable frequency. A *prompt* was defined as any message in which there was a minimum of 6 days preceding it with no other message. This means that any subsequent conversation resulting from a prompt was not included in the analyses.

Data Acquisition

The following procedures were completed using MATLAB (MathWorks, Inc) in to extract outcome measures from app data regarding daily activity; these measures include mean mHealth self-monitoring and self-reported exercise in the week before and after a prompt. Participants' daily activity on the mHealth app was initially coded as (1) logged "yes I exercised today," (2) logged "no I did not exercise today," (3) logged "rest day," and (4) did not log anything. Following this, participants logging was dichotomously categorized: *mHealth self-monitoring* (1-3=yes and 4=no) and *self-reported exercise* (1=yes and 2-4= no).

To determine if self-monitoring behaviors increased in the days following a prompt, the average number of days self-monitored in 1, 3, 5, and 7 days preceding and following a prompt was calculated. These days were selected to facilitate analysis on how the brief effects of a prompt on self-monitoring and self-reported exercise may vary over the week. Days 2, 4, and 6 were excluded to decrease the number of *t* tests being run in an attempt to decrease type I error. Once averages for individual prompts were calculated, weekly averages were established for

the whole year, months 1 to 6, and months 7 to 12; these time points are in line with the overall research study, which assessed all main outcomes at 6 and 12 months. The same procedures were followed to identify the average number of days for self-reported exercise.

Analysis

Paired samples *t* tests were conducted to determine whether self-monitoring and self-reported exercise differed (1) in the day following a prompt compared with the day preceding a prompt, (2) in the 3 days following a prompt compared with the 3 days preceding a prompt, (3) in the 5 days following a prompt compared with the 5 days preceding a prompt, and (4) in the 7 days following a prompt compared with the 7 days preceding it. This analytic procedure was chosen as it aligns with the nature of our hypotheses examining differences before and after a prompt. Change scores for mHealth self-monitoring and self-reported exercise were calculated by taking the difference between the days before and after a prompt. Independent samples *t* tests were conducted to compare change scores between those randomized to HIIT and MICT. Analyses were completed independently for the whole year (months 1-12); the first half of the year (months 1-6) and the latter half of the year (months 7-12). All data were analyzed using SPSS statistics for Windows (version 21, SPSS Inc). Significance level was set at $P<.05$. Effect sizes were calculated using Cohen *d*.

Results

Months 1 to 12

Baseline measurements and demographics of the 69 inactive and overweight adults (mean age 50.7 years, SD 9.40) whose data were included in this study are reported in Table 2. A total of 32 participants were randomized to HIIT, and 37 were randomized to MICT. During the free-living 1-year follow-up period, a total of 369 prompts were sent to the HIIT group (mean 10.25 per participant, SD 3.05), and 465 prompts were sent to the MICT group (mean 10.11 per participant, SD 4.29).

In the year following a diabetes prevention program, there were no significant increases in mHealth self-monitoring or self-reported exercise in 1, 5, and 7 days following a prompt compared with the days preceding a prompt. Both mHealth self-monitoring and self-reported exercise did significantly increase in the 3 days following a prompt compared with the 3 days preceding it.

There were no significant differences between exercise conditions (HIIT and MICT) for both mHealth self-monitoring and self-reported exercise in 1, 3, 5, and 7 days following a prompt. Descriptive statistics and inferential statistics are given in Tables 3 and 4, respectively.

Table 2. Descriptive statistics for individuals who took part in the intervention.

Characteristics	All (N=69)	HIIT[a] (n=32)	MICT[b] (n=37)
Age (years), mean (SD)	50.70 (9.40)	50.72 (9.01)	50.61 (9.87)
Gender, n (%)			
Male	17 (25)	5 (16)	12 (32)
Female	51 (74)	26 (81)	25 (68)
Did not answer	1 (1)	1 (3)	0 (0)
Body mass (kg), mean (SD)	87.92 (19.87)	87.54 (22.35)	88.26 (17.70)
Waist circumference (cm), mean (SD)	107.02 (14.31)	106.66 (14.80)	107.34 (14.05)
VO_2 relative (mL/kg/min)[c], mean (SD)	22.77 (5.88)	22.23 (4.92)	23.20 (6.64)
Ethnic origin, n (%)			
Caucasian	60 (87)	27 (85)	33 (85)
Latin American	2 (3)	1 (3)	1 (3)
Asian	2 (3)	1 (3)	1 (3)
Aboriginal	1 (1)	1 (3)	1 (3)
Other	2 (3)	1 (3)	1 (3)
Missing	2 (3)	1 (3)	1 (3)
Annual income (Can $), n (%)			
0-24,999	1 (2)	0 (0)	1 (3)
25,000-49,999	5 (7)	2 (6)	3 (8)
50,000-74,999	14 (20)	8 (25)	6 (16)
75,000-99,999	14 (20)	9 (28)	5 (13)
>100,000	33 (48)	12 (38)	21 (57)
Missing	2 (3)	1 (3)	1 (3)
Education, n (%)			
High school	9 (13)	5 (15)	4 (11)
College diploma	22 (32)	13 (41)	9 (24)
Bachelor’s degree	20 (29)	6 (19)	14 (38)
Postgraduate degree	15 (22)	6 (19)	9 (24)
Missing	3 (4)	2 (6)	1 (3)
Marital status, n (%)			
Single	6 (9)	3 (10)	3 (8)
Married	49 (72)	25 (78)	24 (65)
Common law	5 (7)	1 (3)	4 (11)
Divorced	5 (7)	1 (3)	4 (11)
Widowed	1 (1)	1 (3)	0 (0)
Missing	3 (4)	1 (3)	2 (5)

[a]HIIT: high-intensity interval training.

[b]MICT: moderate-intensity continuous training.

[c]Cardiorespiratory fitness was the primary outcome of the diabetes prevention program. Participants completed a maximal cardiorespiratory fitness (VO_{2peak}) test to exhaustion on a cycle ergometer at baseline and 6- and 12-month follow-ups.

Table 3. Average number of days participants self-monitored and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 1 to 12.

Days[a]	Total, mean (SD)		HIIT[b], mean (SD)		MICT[c], mean (SD)	
	Before	After	Before	After	Before	After
1						
SM[d]	0.86 (0.15)	0.87 (0.16)	0.88 (0.11)	0.86 (0.14)	0.85 (0.17)	0.87 (0.17)
Exercise	0.44 (0.20)	0.45 (21)	0.40 (0.16)	0.42 (0.21)	0.48 (0.23)	0.49 (0.21)
3						
SM	2.41 (0.42)	2.60 (0.41)	2.44 (0.34)	2.59 (0.36)	2.38 (0.48)	2.61 (0.46)
Exercise	1.21 (0.48)	1.40 (0.53)	1.16 (0.38)	1.27 (0.45)	1.44 (0.53)	1.51 (0.57)
5						
SM	4.23 (0.73)	4.25 (0.74)	4.27 (0.61)	4.29 (0.59)	4.20 (0.83)	4.21 (0.85)
Exercise	2.22 (0.79)	2.26 (0.85)	1.94 (0.64)	2.03 (0.73)	2.46 (0.83)	2.46 (0.91)
7						
SM	5.94 (0.99)	5.99 (1.00)	5.99 (0.82)	6.03 (0.81)	5.90 (1.13)	5.96 (1.16)
Exercise	3.12 (1.09)	3.16 (1.19)	2.75 (0.89)	2.82 (0.96)	3.42 (1.15)	3.46 (1.30)

[a]Comparisons were made between 1 day before and after a prompt, 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Table 4. *T* test, *P* values, and effect size of self-monitoring and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 1 to 12.

Days[a]	Total			HIIT[b] versus MICT[c]		
	t test (df)	*P* value	Cohen *d*	*t* test (df)	*P* value	Cohen *d*
1						
SM[d]	0.16 (68)	.87	0.06	1.87 (67)	.07	0.46
Exercise	0.73 (68)	.47	0.05	0.09 (67)	.93	0.02
3						
SM	6.82 (68)	<.001	0.46	1.25 (67)	.22	0.31
Exercise	2.16 (68)	.03	0.38	0.46 (67)	.65	0.11
5						
SM	0.46 (68)	.65	0.03	0.05 (67)	.97	0.01
Exercise	0.89 (68)	.38	0.05	1.03 (67)	.31	0.25
7						
SM	1.18 (68)	.24	0.09	0.17 (67)	.86	0.04
Exercise	0.99 (68)	.33	0.04	0.18 (67)	.86	0.04

[a]Comparisons were made between 1 day before and after a prompt, 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Months 1 to 6

In months 1 to 6, a total of 226 prompts were sent to the HIIT group (mean 6.28 per participant, SD 1.77), and 283 prompts were sent to the MICT group (mean 6.15 per participant, SD 1.85).

In the first 6 months following the program, mHealth self-monitoring significantly increased in the 3 days following a prompt compared with the 3 days preceding it but did not significantly differ in 1, 5, or 7 days following a prompt compared with preceding days. Self-reported exercise did not significantly increase in the day following a prompt compared with the day preceding it; however, it did significantly increase in 3, 5, and 7 days following a prompt compared with the respective preceding days.

In 1, 5, and 7 days following a prompt compared with the days preceding it, there were no significant differences between HIIT and MICT groups for both mHealth self-monitoring and self-reported exercise. In the 3 days following a prompt compared with the 3 days preceding it, independent samples *t* tests conducted on change scores suggest that there was a significantly larger change in self-monitoring for those randomized to MICT compared with those randomized to HIIT (t_{67}=2.2; *P*=.03; *d*=0.54), but no significant group differences for self-reported exercise (t_{67}=0.05; *P*=.96; *d*=0.012). When looking at HIIT and MICT independently, both groups demonstrated significant increases between the 3 days before and after a prompt in mHealth self-monitoring (HIIT: t_{31}=4.44; *P*<.001; *d*=0.64; MICT: t_{36}=7.94; *P*<.001; *d*=0.90). Additional information regarding descriptive statistics and inferential statistics are given in Tables 5 and 6, respectively.

Table 5. Average number of days participants self-monitored and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 1 to 6.

Days[a]	Total, mean (SD)		HIIT[b], mean (SD)		MICT[c], mean (SD)	
	Before	After	Before	After	Before	After
1						
SM[d]	0.90 (0.17)	0.89 (0.18)	0.93 (0.12)	0.89 (0.18)	0.87 (0.20)	0.89 (0.17)
Exercise	0.44 (0.26)	0.49 (0.26)	0.42 (0.22)	0.47 (0.26)	0.46 (0.29)	0.51 (0.27)
3						
SM	2.40 (0.39)	2.70 (0.37)	2.46 (0.32)	2.68 (0.37)	2.35 (0.44)	2.72 (0.38)
Exercise	1.31 (0.53)	1.52 (0.59)	1.20 (0.47)	1.41 (0.54)	1.40 (0.57)	1.62 (0.62)
5						
SM	4.41 (0.68)	4.40 (0.75)	4.49 (0.59)	4.45 (0.65)	4.35 (0.75)	4.36 (0.83)
Exercise	2.32 (0.86)	3.45 (0.98)	2.10 (0.83)	2.24 (0.89)	2.51 (0.86)	2.64 (1.01)
7						
SM	6.21 (0.90)	6.21 (1.00)	6.31 (0.78)	6.28 (0.83)	6.12 (1.00)	6.15 (1.13)
Exercise	3.26 (1.21)	3.45 (1.27)	2.98 (1.12)	3.11 (1.04)	3.50 (1.24)	3.74 (1.39)

[a]Comparisons were made between 1 day before and after a prompt, 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Table 6. *T* test, *P* values, and effect size of self-monitoring and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 1 to 6.

Days[a]	Total			HIIT[b] versus MICT[c]		
	t test (df)	*P* value	Cohen *d*	*t* test (df)	*P* value	Cohen *d*
1						
SM[d]	0.61 (68)	.05	0.06	1.53 (67)	.13	0.37
Exercise	1.40 (68)	.17	0.19	0.09 (67)	.93	0.02
3						
SM	8.61 (68)	<.001	0.60	2.20 (67)	.03	0.54
Exercise	3.70 (68)	<.001	0.37	0.05 (67)	.96	0.01
5						
SM	0.38 (68)	.71	0.01	0.51 (67)	.62	0.12
Exercise	2.15 (68)	.04	0.14	0.09 (67)	.93	0.02
7						
SM	0.06 (68)	.95	0.01	0.52 (67)	.60	0.13
Exercise	2.46 (68)	.02	0.15	0.80 (67)	.43	0.20

[a]Comparisons were made between 1 day before and after a prompt, 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Months 7 to 12

In months 7 to 12, a total of 143 prompts were sent to the HIIT group (mean 4.47 per participant, SD 1.5), and 182 prompts were sent to the MICT group (mean 4.92 per participant, SD 2.41).

In the second half of the year following the program, there were no significant differences in either mHealth self-monitoring or self-reported exercise in 1, 3, 5, and 7 days following a prompt compared with the days preceding a prompt. There were no significant differences between exercise conditions (HIIT and MICT) for both mHealth self-monitoring and self-reported exercise in 1, 3, 5, and 7 days following a prompt in the second half of the year. Descriptive statistics and inferential statistics are presented in Tables 7 and 8, respectively.

Table 7. Average number of days participants self-monitored and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 7 to 12.

Days[a]	Total, mean (SD)		HIIT[b], mean (SD)		MICT[c], mean (SD)	
	Before	After	Before	After	Before	After
1						
SM[d]	0.82 (0.23)	0.84 (0.24)	0.84 (0.20)	0.84 (0.23)	0.80 (0.25)	0.84 (0.25)
Exercise	0.44 (0.28)	0.43 (0.29)	0.38 (0.25)	0.35 (0.29)	0.49 (0.29)	0.50 (0.28)
3						
SM	2.41 (0.66)	2.45 (0.69)	2.40 (0.62)	2.45 (0.61)	2.42 (0.71)	2.46 (0.76)
Exercise	1.31 (0.71)	1.24 (0.65)	1.11 (0.58)	1.06 (0.60)	1.49 (0.76)	1.39 (0.67)
5						
SM	4.00 (1.09)	4.06 (1.03)	4.02 (1.01)	4.08 (0.96)	3.98 (1.16)	4.04 (1.09)
Exercise	2.12 (1.02)	2.06 (0.96)	1.76 (0.81)	1.78 (0.84)	2.42 (1.09)	2.32 (1.00)
7						
SM	5.58 (1.50)	5.71 (1.43)	5.57 (1.40)	5.69 (1.38)	5.59 (1.61)	5.72 (1.49)
Exercise	2.93 (1.42)	2.87 (1.38)	2.46 (1.18)	2.46 (1.23)	3.35 (1.49)	3.22 (1.42)

[a]Comparisons were made between 1 day before and after a prompt, 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Table 8. *T* test, *P* values, and effect size of self-monitoring and self-reported exercise in a mobile health app in 1, 3, 5, and 7 days before and after a prompt in months 7 to 12.

Days[a]	Total			HIIT[b] versus MICT[c]		
	t test (df)	*P* value	Cohen *d*	*t* test (df)	*P* value	Cohen *d*
1						
SM[d]	1.06 (68)	.29	0.09	0.81 (67)	.42	0.20
Exercise	0.25 (68)	.80	0.04	0.56 (67)	.58	0.13
3						
SM	0.96 (68)	.34	0.06	0.20 (67)	.85	0.05
Exercise	1.22 (68)	.23	0.10	0.43 (67)	.67	0.11
5						
SM	1.05 (68)	.29	0.06	0.08 (67)	.94	0.02
Exercise	0.62 (68)	.05	0.06	0.72 (67)	.47	0.18
7						
SM	1.57 (68)	.12	0.09	0.07 (67)	.95	0.02
Exercise	0.72 (68)	.48	0.04	0.74 (67)	.46	0.18

[a]Comparisons were made between the 1 day before and after a prompt, the 3 days before and after a prompt, 5 days before and after a prompt, and 7 days before and after a prompt for mHealth self-monitoring and self-reported exercise.

[b]HIIT: high-intensity interval training.

[c]MICT: moderate-intensity continuous training.

[d]SM: mHealth self-monitoring.

Discussion

Principal Findings

The primary objective of this secondary data analysis was to assess changes in mHealth self-monitoring and self-reported exercise in the days preceding and following a prompt. Secondary objectives of this research were to examine whether results differed based on exercise modality (HIIT vs MICT) and the differences between the first and second half of the year following a diabetes prevention program. Overall results suggest that both self-monitoring and self-reported exercise behaviors significantly increase in the 3 days following a prompt when compared with the 3 days preceding it, the greatest changes were observed in the first half of the year, and there were no differences between exercise modality.

Months 1 to 12

In the year following a diabetes prevention program, the observed differences in self-monitoring and self-reported exercise behaviors were most potent in the 3 days following a prompt, whereas there were no significant changes in 1, 5, or 7 days following a prompt. Exercise is a complex behavior that requires self-regulation such as scheduling and planning [37]. As such, it might be unrealistic to expect to observe changes in self-reported exercise behavior in a singular day or in the day immediately following a prompt. The changes in behaviors before and after a prompt in the year following a diabetes prevention program may be most effective somewhere between 1 and 3 days, as an individual begins to self-regulate to schedule exercise to get back on track.

First and Second Half of the Year

In the first half of the year, self-reported exercise behavior significantly increased in 3, 5, and 7 days following a prompt, but not the day immediately following a prompt, whereas no significant changes were observed in the second half of the year. Reasons why prompts appeared to have no observed change on behaviors in the second half of the year are unknown and should be the focus of future research. In the first half of the year, a total of 226 prompts were sent, whereas in the second half of the year, 143 prompts were sent. Although the difference in frequency of prompts sent between the first and second half of the year could have influenced the observed changes, additional research is required to examine the role of prompt frequency in changing self-monitoring and self-reported exercise behavior.

Similar to self-reported exercise, the impact of prompts on mHealth self-monitoring was only observed in the first half of the year. Specifically, in the first half of the year following a diabetes prevention program, mHealth self-monitoring significantly increased in the 3 days following a prompt compared with the 3 days preceding a prompt, but not in 1, 5, or 7 days following a prompt. Individuals included in this analysis self-monitored an average of 286 days in the 12-month follow-up period. It may be the case that prompts are not needed for individuals who regularly self-monitor. However, it is difficult to discern the impact of a prompt on daily self-monitoring behaviors, given that the majority of participants were self-monitoring on a daily basis, and there was no control group.

High-Intensity Interval Training Versus Moderate-Intensity Continuous Training

There is a growing body of evidence suggesting HIIT may be a viable exercise alternative to MICT [32,33]. There were no differences in behaviors preceding and following a prompt between the 2 exercise modalities. This may suggest that certain self-regulatory skills and cognitions may not appreciably differ between HIIT and MICT. Although there are compelling arguments for HIIT being a more time efficient and easier to self-manage alternative to MICT [38-42], our results suggest that the impact of prompts on self-monitoring and self-report exercise did not differ between HIIT and MICT.

Strengths and Limitations

Systematic reviews of the literature have shown that individuals who received prompts had greater weight loss and increased PA compared with nonprompt controls [26,31]. Despite the overall positive effects of prompts, design of these studies has varied significantly, and few studies have reported on or assessed the effectiveness of individual intervention characteristics [26,31]. Within these reviews of the literature, it has been recommended that future research focus on the impact of specific prompt delivery characteristics such as prompt frequency, timing, and intervention duration. A primary strength of this analysis is that it looked at the immediate effect of a prompt on self-monitoring and self-reported exercise behaviors. Another strength of this study was that we examined participants randomly assigned to different exercise conditions (HIIT and MICT), which allowed us to examine differences in prompt effectiveness between exercise modalities, which has not been addressed in previous studies. Finally, the 12-month follow-up period in which participants self-monitored on an mHealth app is longer than previous prompt studies, which often last less than 14 weeks [22,28,29]. Although examining 12 months of self-monitoring was a strength, people may not need to self-monitor in this way. Once people establish a regular behavioral pattern of exercise, self-monitoring through an app may not be needed to facilitate this regular exercise engagement.

Despite these strengths, this study represents a secondary analysis of prompt data, and the primary objectives of this RCT did not relate to mHealth prompts. Limitations of this paper include a lack of a control condition (ie, not receiving prompts), no a priori sample size calculation, and conducting multiple *t* tests without adjustment. All participants using the mHealth app received prompts, and there was no experimental manipulation of prompts. Another limitation was that there had been no validation of this mHealth app as an exercise measure. However, the information participants report on the mHealth app is similar to the information contained in validated measures (eg, Godin-Shephard Leisure-Time Physical Activity Questionnaire [43]). Although we recognize this measure has not been validated, our research question and outcomes concern engagement or nonengagement in exercise. As such, we are less concerned with the validation impacting results as we are simply looking whether or not individuals logged exercise.

One final limitation regarding the criteria for prompts to be included in the analyses. Our analyses examined the effects of a prompt on self-monitoring and self-reported exercise behaviors and did not include any subsequent conversation resulting from a prompt in the analysis. There is a possibility that the amount of interaction between a participant and their exercise counselor on the mHealth app influenced their behaviors.

Despite this preliminary evidence that prompts may influence self-monitoring and self-reported exercise behaviors, future research is needed to examine the causal impact of prompt frequency on self-monitoring and self-reported exercise behavior in an attempt to elucidate an optimal prompt frequency for behavior change.

Future Directions

These analyses used only those participants who were engaging with the app and individuals self-monitored approximately 80% of the time. In addition, we were unable to analyze whether the level of virtual interaction between exercise counselors and participants influenced the effect of a prompt on self-monitoring and self-reported exercise behaviors. Future studies should address the impact of prompts on less consistent self-monitors while also examining the role that social interaction may play on self-monitoring and self-reported exercise.

The duration of prompts' impact on self-reported exercise behaviors was relatively short (in 3, 5, and 7 days following a prompt, only in the first half of the year). Future studies should examine the optimal prompt frequency and timing for cueing self-reported exercise behavior following behavior change programs. Utilization of optimization trials or n-of-1 trials may be 1 possible means to examine dose-response relationship between app-delivered prompts and exercise.

Conclusions

Within this analysis, we provided evidence regarding the observed changes in self-monitoring and self-reported exercise behavior following a prompt and the duration for which prompts may be effective as exercise behavior change tools. Future studies assessing prompts should examine causal factors relating to the observed decrease in prompt effectiveness on self-reported exercise behaviors in the 7 to 12 months following an exercise behavior change program. Understanding how to optimally intervene through prompts can decrease intervention cost and time as researchers may limit unnecessary prompts while continually encouraging consumers to use mHealth technology to change health behaviors.

Acknowledgments

This research is supported with funds from the WorkSafeBC research program in addition to MEJ's Michael Smith Foundation for Health Research Award #5917 and CIHR Foundation Grant #333266.

Authors' Contributions

The primary research project is part of MEJ's ongoing research program. The secondary analysis idea was conceived by MMM, MEJ, and SRL. MMM was responsible for data interpretation and writing and editing the manuscript. KJM was responsible for data analysis and writing and editing the manuscript. SRL and MEJ were responsible for overseeing all aspects of the study, contributing to data interpretation, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2MB - mhealth_v7i9e12956_app1.pdf]

References

1. International Telecommunication Union. 2015. ICT Facts and Figures: The World in 2015 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf [accessed 2019-08-07]
2. International Telecommunications Union Geneva. 2016. ICT Facts and Figures: 2016 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [accessed 2019-08-07]
3. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: the digital link to health survey. JMIR Mhealth Uhealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
4. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;46(Suppl 2):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
5. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]

6. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of Android and Apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
7. 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]
8. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behaviour 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. Mohr DC, Cheung K, Schueller SM, Brown CH, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
10. Conroy DE, Dubansky A, Remillard J, Murray R, Pellegrini CA, Phillips SM, et al. Using behavior change techniques to guide selections of mobile applications to promote fluid consumption. Urology 2017 Jan;99:33-37. [doi: 10.1016/j.urology.2016.09.015] [Medline: 27645525]
11. 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]
12. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
13. 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]
14. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
15. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
16. Munson SA, Consolvo S. Exploring Goal-Setting, Rewards, Self-Monitoring, and Sharing to Motivate Physical Activity. In: 6th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 2012 Presented at: PervasiveHealth'12; May 21-24, 2012; San Diego, CA, USA p. 25-32. [doi: 10.4108/icst.pervasivehealth.2012.248691]
17. 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 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
18. Carels RA, Darby LA, Rydin S, Douglass OM, Cacciapaglia HM, O'Brien WH. The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes. Ann Behav Med 2005 Dec;30(3):182-190. [doi: 10.1207/s15324796abm3003_2] [Medline: 16336069]
19. 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]
20. Conroy MB, Yang K, Elci OU, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574 [FREE Full text] [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]
21. Liu S, Willoughby JF. Do fitness apps need text reminders? An experiment testing goal-setting text message reminders to promote self-monitoring. J Health Commun 2018;23(4):379-386. [doi: 10.1080/10810730.2018.1455768] [Medline: 29601270]
22. 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]
23. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]
24. 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]
25. de Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
26. 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]
27. Head KJ, Noar SM, Iannarino NT, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]

28. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
29. Monroe CM, Thompson DL, Bassett Jr DR, Fitzhugh EC, Raynor HA. Usability of mobile phones in physical activity-related research. Med Sci Sports Exerc 2014;46(4):196-206. [doi: 10.1249/01.mss.0000495254.66691.dd]
30. Wei J, Hollin I, Kachnowski S. A review of the use of mobile phone text messaging in clinical and healthy behaviosr interventions. J Telemed Telecare 2011;17(1):41-48. [doi: 10.1258/jtt.2010.100322] [Medline: 21097565]
31. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: a literature review. Health Informatics J 2012 Dec;18(4):235-250 [FREE Full text] [doi: 10.1177/1460458212442422] [Medline: 23257055]
32. Guiraud T, Nigam A, Gremeaux V, Meyer P, Juneau M, Bosquet L. High-intensity interval training in cardiac rehabilitation. Sports Med 2012 Jul 1;42(7):587-605. [doi: 10.2165/11631910-000000000-00000] [Medline: 22694349]
33. Kessler HS, Sisson SB, Short KR. The potential for high-intensity interval training to reduce cardiometabolic disease risk. Sports Med 2012 Jun 1;42(6):489-509. [doi: 10.2165/11630910-000000000-00000] [Medline: 22587821]
34. Bourne JE, Little JP, Beauchamp MR, Barry J, Singer J, Jung ME. Brief exercise counseling and high-intensity interval training on physical activity adherence and cardiometabolic health in individuals at risk of type 2 diabetes: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Mar 26;8(3):e11226 [FREE Full text] [doi: 10.2196/11226] [Medline: 30912761]
35. Voth EC, Oelke ND, Jung ME. A theory-based exercise app to enhance exercise adherence: a pilot study. JMIR Mhealth Uhealth 2016 Jun 15;4(2):e62 [FREE Full text] [doi: 10.2196/mhealth.4997] [Medline: 27307134]
36. Warburton DE, Jamnik VK, Bredin SS, Gledhill N. The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). Health Fit J Can 2011;4(2):3-17. [doi: 10.14288/hfjc.v11i4.270]
37. Vohs KD, Baumeister RF, editors. Handbook of Self-Regulation: Research, Theory, and Applications. Second Edition. New York: Guilford Press; 2011.
38. Astorino TA, Thum JS. Response: commentary: why sprint interval training is inappropriate for a largely sedentary population. Front Psychol 2016;7:746 [FREE Full text] [doi: 10.3389/fpsyg.2016.00746] [Medline: 27242644]
39. Biddle SJ, Batterham AM. High-intensity interval exercise training for public health: a big HIT or shall we HIT it on the head? Int J Behav Nutr Phys Act 2015 Jul 18;12:95 [FREE Full text] [doi: 10.1186/s12966-015-0254-9] [Medline: 26187579]
40. del Vecchio FB, Gentil P, Coswig VS, Fukuda DH. Commentary: why sprint interval training is inappropriate for a largely sedentary population. Front Psychol 2015;6:1359 [FREE Full text] [doi: 10.3389/fpsyg.2015.01359] [Medline: 26441735]
41. Jung ME, Little JP, Batterham AM. Commentary: why sprint interval training is inappropriate for a largely sedentary population. Front Psychol 2015;6:1999 [FREE Full text] [doi: 10.3389/fpsyg.2015.01999] [Medline: 26779101]
42. Robertson-Wilson J, Eys M, Hazell TJ. Commentary: why sprint interval training is inappropriate for a largely sedentary population. Front Psychol 2017;8:1603 [FREE Full text] [doi: 10.3389/fpsyg.2017.01603] [Medline: 28955289]
43. Godin G. The Godin-Shephard leisure-time physical activity questionnaire. Health Fit J Can 2011;4(1):18-22 [FREE Full text]

Abbreviations

BCT: behavior change technique
HIIT: high-intensity interval training
mHealth: mobile health
MICT: moderate-intensity continuous training
PA: physical activity
RCT: randomized controlled trial
SM: mHealth self-monitoring

Edited by G Eysenbach; submitted 27.11.18; peer-reviewed by J Willoughby, B Sakakibara, J Job; comments to author 31.03.19; revised version received 08.05.19; accepted 21.07.19; published 05.09.19.

Please cite as:
MacPherson MM, Merry KJ, Locke SR, Jung ME
Effects of Mobile Health Prompts on Self-Monitoring and Exercise Behaviors Following a Diabetes Prevention Program: Secondary Analysis From a Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(9):e12956
URL: http://mhealth.jmir.org/2019/9/e12956/
doi:10.2196/12956
PMID:31489842

Original Paper

Mobile Phone Access and Preference for Technology-Assisted Aftercare Among Low-Income Caregivers of Teens Enrolled in Outpatient Substance Use Treatment: Questionnaire Study

Stacy R Ryan-Pettes[1*], PhD; Lindsay L Lange[1*], MA; Katherine I Magnuson[1*], BA

Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States

[*]all authors contributed equally

Corresponding Author:
Stacy R Ryan-Pettes, PhD
Department of Psychology and Neuroscience
Baylor University
One Bear Place 97334
Waco, TX, 76798-7334
United States
Phone: 1 2547102544
Email: stacy_ryan@baylor.edu

Abstract

Background: Improvements in parenting practices can positively mediate the outcomes of treatment for adolescent substance use disorder. Given the high rates of release among adolescents (ie, 60% within three months and 85% within one year), there is a critical need for interventions focused on helping parents achieve and maintain effective parenting practices posttreatment. Yet, research suggests that engaging parents in aftercare services is difficult, partly due to systemic-structural and personal barriers. One way to increase parent use of aftercare services may be to offer mobile health interventions, given the potential for wide availability and on-demand access. However, it remains unclear whether mobile phone–based aftercare support for caregivers of substance-using teens is feasible or desired. Therefore, formative work in this area is needed.

Objective: This study aims to determine the feasibility and acceptability of mobile phone–based aftercare support in a population of caregivers with teens in treatment for substance use.

Methods: Upon enrollment in a treatment program, 103 caregivers completed a mobile phone use survey, providing information about mobile phone ownership, access, and use. Caregivers also provided a response to items assessing desire for aftercare services, in general; desire for mobile phone–based aftercare services specifically; and desire for parenting specific content as part of aftercare services. Research assistants also monitored clinic calls made to caregivers' mobile phones to provide an objective measure of the reliability of phone service.

Results: Most participants were mothers (76.7%) and self-identified as Hispanic (73.8%). The average age was 42.60 (SD 9.28) years. A total of 94% of caregivers owned a mobile phone. Most had pay-as-you-go phone service (67%), and objective data suggest this did not impede accessibility. Older caregivers more frequently had a yearly mobile contract. Further, older caregivers and caregivers of adolescent girls had fewer disconnections. Bilingual caregivers used text messaging less often; however, caregivers of adolescent girls used text messaging more often. Although 72% of caregivers reported that aftercare was needed, 91% of caregivers endorsed a desire for mobile phone–based aftercare support in parenting areas that are targets of evidence-based treatments.

Conclusions: The results suggest that mobile phones are feasible and desired to deliver treatments that provide support to caregivers of teens discharged from substance use treatment. Consideration should be given to the age of caregivers when designing these programs. Additional research is needed to better understand mobile phone use patterns based on a child's gender and among bilingual caregivers.

(JMIR Mhealth Uhealth 2019;7(9):e12407) doi:10.2196/12407

KEYWORDS

mobile phones; text messaging; substance use treatment; mhealth; parenting; aftercare

Introduction

Adolescent substance use is a major public health concern in the United States. The high rates of substance use coupled with significant unmet treatment needs and the alarming rates of recurrence of use are concerning. Recent estimates suggest that 2.0 million adolescents in the 12- to 17-year age group used illicit substances in 2017. Among these users, 1.0 million needed substance use treatment. Strikingly, among those who complete treatment, research suggests that 60%-70% will have a recurrence of use within 90 days after a treatment episode and 85%, within 1 year following treatment [1-3].

Improvement in parenting practices can positively mediate the outcomes of outpatient treatment for adolescent substance use disorder [4,5], and these findings extend to the aftercare period. Research suggests that evidence-based treatments for adolescent substance use are successful at improving parenting practices and that changes in monitoring and positive parenting behaviors mediate relations between treatment and adolescent substance use outcome [4,6]. There is evidence to suggest that parenting practices during the aftercare period also mediate adolescent substance use outcomes. To date, one study has examined parenting practices as a mechanism of change during aftercare [7]. Results of this study showed that the combination of continuing care and parent involvement is related to better adolescent substance use outcomes. Results of nonmediation studies of the posttreatment period suggest that these results may extend beyond parental involvement. For example, Stanger and colleagues [8] found that parental monitoring at the end of outpatient treatment was related to abstinence among adolescents after treatment. In a follow-up study of a low-income, higher-proportion, minority sample, Stanger and colleagues [8] included posttreatment booster sessions for parents, hypothesizing that aftercare for parents would be related to maintenance of effective parenting practices, resulting in higher adolescent rates of abstinence [9]. The investigators were unable to test this hypothesis due to poor parental attendance at the booster sessions. Results showed that parents attended an average of less than one session over a 3-month period and that the rates of adolescent relapse were significant. Caregiver participation in aftercare may be improved when aftercare is transitioned to the community; however, previous research on aftercare following outpatient services has not disentangled child and caregiver rates of participation [10].

Research suggests that aftercare services for parents of youth discharged from outpatient substance use treatment are important; however, offering clinic-based aftercare services may not be effective due to poor participation [9]. Parental participation in clinic-based aftercare may be poor, in part, due to significant barriers [11,12]. Common systemic-structural barriers to parent participation in mental health services include indirect financial cost (eg, lost wages for missed work) and lack of flexibility of appointments and settings (ie, clinic-based). Although moving face-to-face aftercare to the home setting may improve attendance among families [10], low-income caregivers may still find it difficult to fully engage in sessions [13,14]. Low-income caregivers may experience insurmountable barriers to clinic- and home-based services. For example, inadequate support, poor parental efficacy, low hourly wages, and significant daily hassles [15] may make it exceptionally difficult to pool resources or miss work due to the downstream effects.

Barriers to treatment participation is a major contributing factor in observed socioeconomic health disparities [16]. Research shows that while low-income adolescents are not more likely to use drugs [17], they are more likely to develop a problem and face the consequences of substance use due to differences in initiating treatment [16,18], engaging in treatment [16,18], and attending aftercare support [19]. One way to improve adolescent rates of abstinence and decrease socioeconomic health disparities in the area of adolescent substance use may be to increase the availability of aftercare services via cost-effective technologies that are far reaching, are on demand, and target effective parenting strategies.

Given that mobile phone ownership and usage are omnipresent in our society [20], one of the many advantages of mobile health (mHealth) interventions is their ability to provide widely available, far-reaching, and on-demand treatments to individuals facing significant barriers to receiving face-to-face services. Successful adoption of mobile phone–based aftercare for low-income parents is predicated upon access and reliability of mobile phone service, desire to engage in mobile phone–based aftercare support, and belief that a mobile phone–based program with parenting support is needed.

Although research has established that low-income populations have access to mobile phones [20-23] and that ownership may be significantly higher among low-income caregivers of children than national averages [24], there are at least two gaps in this research. First, our review of the literature failed to find reports of access to mobile phones among low-income caregivers of children involved in substance use treatment. Low-income caregivers of children involved in substance use treatment face more economic strain than low-income caregivers of children not involved in substance use [25]. This strain may impact access to mobile phone technology and reliability of service. Indeed, rates of mobile phone access among adult substance users is lower than the national averages [21-23]. Second, prior research has relied on participants' self-report of reliability of service [21-23]. Research has yet to include objective indicators of reliability of service in research characterizing mobile phone ownership in low-income populations. A better understanding of access and reliability of service among low-income caregivers of adolescents involved in substance use treatment is needed to better understand the feasibility of using mHealth interventions with this population.

Although research has established that low-income caregivers of children desire mHealth interventions [26,27], this research has been limited to caregivers of children with medical conditions. Preference for mHealth interventions among caregivers of substance-using teens is not available. Successful adoption of mobile phone–based interventions with low-income caregivers of teens exiting treatment for substance use requires an understanding of whether caregivers desire to engage in mobile phone–based support. Attitudes toward mHealth interventions for medical conditions may not be shared toward mHealth interventions for substance use. Two main reasons

weaken transferability of attitudes. First, interventions for medical conditions are viewed as less stigmatizing than interventions for substance use [28]. Second, caregivers may denounce the need for their own mHealth intervention because it is the child who is struggling with substance use—an attitude that is pervasive in standard substance use counseling for teens [29]. Therefore, it is critical to understand whether caregivers of substance-using teens desire mHealth interventions.

Currently, there is no research on whether caregivers of youth who use substances believe that a mobile phone–based program with parenting support would meet their needs. A critical review and empirical test of the mobile health service adoption model suggests that perceived usefulness is both directly related to behavioral intention to use mHealth services and indirectly related to behavioral intentions to use mHealth services through an individual's feelings about performing the target behavior [30]. Given that the goal of designing mHealth services is user adoption, it is critical to understand whether caregivers would find a mobile phone–based program with parenting support useful.

To address the current gaps in the literature and offer formative work for the design of mHealth interventions for parents of youth in substance use treatment, this study recruited caregivers of low-income status who were attending behavioral family therapy for adolescent substance use to characterize mobile phone ownership and use; assess self-report and objective data of reliability of mobile phone service; examine caregiver desire for mobile phone–based aftercare support; and examine specific content caregivers desire as part of a mobile phone–based aftercare program focused on parenting skills.

Methods

Participants and Study Overview

Caregivers of teenagers (N=103) enrolled in an outpatient substance abuse treatment program affiliated with an academic institution in the Southwestern region of the United States participated in this study, which concluded in July 2016. For the purpose of this study, a caregiver was defined as an adult who has the legal authority to make treatment decision for the enrolled teenager, an adult who makes decisions about the enrolled teenager, or an adult who makes sure the teenager is looked after every day. Eligible caregivers were associated with teens under the age of 18 years who were actively participating in an outpatient treatment program for a substance use disorder. The survey information reported herein was collected as part of procedures for the development of an outpatient clinic devoted to delivering substance use treatments to low-income teenagers and their caregivers and for development of mobile tools capable of enhancing the effects of treatment and sustaining treatment gains. Upon admission, all families were presented with a consent to treatment form that included information about the goals of the clinic, a rationale for surveys included in their admissions packet, and potential uses of their survey and clinic data. Participants had the opportunity to allow or disallow their survey and clinic data to be used in research. All families understood that they could receive services irrespective of their decision. The local institutional review board approved this study.

Procedure

Completion of a mobile phone ownership, usage, and preference survey via paper and pencil (n=19) or a computer (n=84) was part of admission to an adolescent outpatient substance use treatment program that provides service to families that are uninsured, receiving Medicaid benefits, or earning <US $20,000 per year. Upon discharge from treatment, research assistants completed a chart review to obtain survey results and code the outcome of outbound calls made to mobile phones by clinic staff. Data to describe the population, also collected via chart review, included demographic information and primary substance of choice for the enrolled teenager.

Measures

Client Information and Substance Use History

As part of the admissions process, caregivers completed a client information form and their respective teenager completed a substance use history questionnaire. For the purpose of this study, only data necessary to describe the sample was collected during chart review. Parents provided their age, relationship to the enrolled teenager, languages spoken, and preference for spoken language in an open-text field. Parents selected their race (white, black/African American, American Indian/Alaskan Native, Native Hawaiian/other Pacific Islander, Asian, more than one race, or other), ethnicity (Hispanic or non-Hispanic), and education (less than seventh grade, junior high, partial high school, high school graduate, partial college, college degree, or graduate/professional training) from a list of options. Enrolled teenagers provided the name of their primary substance of choice in an open-text field on a substance use history questionnaire.

Mobile Phone Ownership, Usage, and Preference Survey

The questionnaire consisted of 18 items replicating the surveys conducted by McClure and colleagues [21] and Milward and colleagues [22], which covered availability of a mobile phone, type of phone, service plan, and day-to-day use of mobile phone features. To extend previous surveys, additional questions were developed by the first author to assess preference for aftercare support, in general, and specifically via mobile phone and preference for the content of mobile phone-based support. Surveys were available in English or Spanish, and caregivers selected the version according to their primary language preference.

Access and Reliability of Service

In addition to self-report of accessibility and reliability of service, an objective measure of accessibility and reliability of service during treatment was obtained. This measure included a chart review of outbound calls made to caregivers who completed the survey. These caregivers reported that they owned their mobile phone and indicated that they intended to use the mobile phone as their primary means of communication with the clinic (100%). Clinic confirmation of appointment policy included contacting caregivers 3 days before a scheduled appointment and on the day of a scheduled appointment. All outgoing phone calls were entered into a telephone contact and

appointment log specifying the time, date, phone number used to make contact, and outcome by clinic staff. Clinic staff recorded the outcome of outgoing calls as either "confirmed," "left a message," "unable to leave a message, yet ringing," or "phone disconnected." Second author LL completed all chart reviews and created a data file for review by the first author SR-P. All data were checked and verified as correct by the first author.

Accessibility was operationalized as the percent of contact calls determined to be in-service (ie, outbound calls recorded as confirmed, left a message, or unable to leave a message yet ringing). Reliability of service was operationalized as the number of days between disconnection and when the mobile phone was able to accept clinic calls, and treatment staff was able to either leave a message or speak with the individual. Through this method, researchers were able to calculate the number of days until the clinic was able to either leave a message or talk with the client, after an unsuccessful attempt due to disconnection. Three call outcomes were missing and left as missing data.

Statistical Analyses

Data were analyzed using SPSS (Version 23. IBM Corp, Armonk, NY). Descriptive statistics were used to quantify demographic information and mobile phone characteristics, utilization, and accessibility and reliability of service. Consistent with previous research [21,23], exploratory binary logistic regression and linear regression analyses were conducted to examine the association between demographic variables and mobile phone characteristics, utilization, and accessibility and reliability of service.

For binary logistic regression analyses, select mobile phone characteristics, use preferences, and self-report of accessibility and reliability of service were regressed on parent age (continuous variable), ethnicity (non-Hispanic=0, Hispanic=1), bilingual language (nonbilingual=0, bilingual=1), and education (less than high school=0, more than high school=1) as well as child age (continuous variable) and gender (female=0, male=1) in separate regression models. The mobile phone characteristics and use preference variables were selected a priori as those that would provide developers information about accessibility and reliability of service as well as mobile features to consider incorporating when tailoring mHealth services: yearly contract (no=0), primarily use phone for text messaging (no=0), phone number changed one or more times in the past year (no=0), accesses to the internet mostly from the mobile phone (no=0), and experience with service connection issues (never or rarely=0). For linear regression analyses, continuous variables indicating inaccessibility and disruption of service (ie, number of times unable to leave message and number of times disconnected) were regressed on the same demographic variables as those used in the logistic regression analyses.

An alpha of .05 was maintained throughout. Considering these parameters, power analysis was conducted using the online estimation tool GPOWER (Version 31. Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany) [31]. Results showed that our sample size was adequate to detect results of the medium effect size for linear regression analyses.

Results

Participants

Demographic information for caregiver and teen participants is presented in Table 1. Caregivers had a mean age of 42.60 (SD 9.28) years. The majority of caregivers were mothers (76.7%; biological, step, or adoptive), identified as Hispanic (73.8%), and reported English as their primary spoken language (88.3%). Teens participating in treatment had an average age of 15.94 (SD 1.32) years. Most teens were male (65%), enrolled in high school (71.8%), and primarily used marijuana (94.2%).

Table 1. Caregiver and child demographics (N=103).

Variable	Value
Caregiver variables	
Age of parent, mean (SD)	42.60 (9.28)
Relationship to client, n (%)	
Biological/step/adoptive mother	79 (76.7)
Biological/step/adoptive father	14 (13.6)
Grandmother	5 (4.9)
Other (aunt, adult sibling)	5 (4.9)
Race, n (%)	
White	73 (70.9)
Black/African American	9 (8.7)
American Indian/Alaskan Native	1 (1.0)
Native Hawaiian/Other Pacific Islander	0 (0.0)
Asian	0 (0.0)
More than one race	6 (5.8)
Other, not specified	12 (11.7)
Ethnicity, n (%)	
Hispanic	76 (73.8)
Primary language, n (%)	
English	91 (88.3)
Spanish	12 (11.7)
Bilingual (English/Spanish)	49 (47.6)
Education[a], n (%)	
Less than seventh grade	2 (1.9)
Junior high	9 (8.9)
Partial high school	14 (13.9)
High school graduate	34 (33.7)
Partial college	29 (28.7)
4-year college degree	10 (9.9)
Graduate/professional training	3 (3.0)
Child variables	
Age of child, mean (SD)	15.94 (1.32)
Gender, n (%)	
Male	67 (65.0)
Female	36 (35.0)
Grade level, n (%)	
High school	74 (71.8)
Junior high	21 (20.4)
Graduated	2 (1.9)
Not in school	6 (5.8)
Primary substance, n (%)	
Marijuana	97 (94.2)
Synthetic marijuana	4 (3.9)

Variable	Value
Alcohol	1 (1.0)
Other	1 (1.0)

[a]N=101.

Mobile Phone Characteristics and Use

Full results for mobile phone ownership, characteristics, and use are presented in Table 2. Notably, the majority of caregivers used mobile phones at least once per week (92%), reported mobile phone ownership (94.2%) of smartphones (83.5%), had unlimited text messaging (92%), and reported text messaging as the most used mobile phone feature (59%).

Self-Report of Accessibility and Reliability of Mobile Phone Service

Full results for self-report of mobile phone accessibility and reliability of service are presented in Table 2. Notably, most caregivers maintained the same mobile phone number over the past year (64%) with infrequent connection disruptions (48% reported rare disruption and 28% reported no disruption). Additionally, positive outcomes of outbound calls from clinic staff to caregivers' mobile phones were high, with 97.2% of calls reaching a phone that was in service and able to receive calls or text messages, including 47.2% of clients who were reached, 45.5% of clients who received a voice mail, and 4.5% who showed a missed call from the clinic (ie, phone ringing but unable to leave a message because the client's mailbox was either full or no message had been setup), while only 2.7% of calls reached disconnected phones.

Table 2. Mobile phone accessibility, reliability, and use.

Caregiver variable	Value
Accessibility, n (%)	
Access (whole sample, n=103)	
Owns a mobile phone	97 (94.2)
Daily access, but do not own	3 (2.9)
Unreliable access, do not own	3 (2.9)
Mobile phone owners (n=97)	
Cell is primary phone	97 (100.0)
Smartphone device	81 (83.5)
Smartphone type	
iPhone	13 (16.0)
Android	57 (70.4)
Windows	4 (4.9)
Other	7 (8.6)
Service type (n=97)	
Pay-as-you-go	65 (67.0)
Yearly	32 (33.0)
Reliability (those with access, n=100), n (%)	
Mobile number changed last year	
Never	64 (64.0)
Once	21 (21.0)
Twice	8 (8.0)
More than thrice	7 (7.0)
Disruptions in mobile phone connections	
Never	28 (28.0)
Rarely	48 (48.0)
Sometimes	20 (20.0)
Often	2 (2.0)
Always	2 (2.0)
Use (those with access, n=100)	
Text message limit, n (%)	
No	92 (92.0)
Yes	3 (3.0)
Not sure	5 (5.0)
Use mobile to text, n (%)	97 (97.0)
Use mobile to email, n (%)	76 (76.0)
Use mobile to take pictures, n (%)	93 (93.0)
Use mobile to play music, n (%)	82 (82.0)
Use mobile to download mobile applications, n (%)	84 (84.0)
Use mobile to access internet, n (%)	92 (92.0)
Use mobile most for, n (%)	
Text	59 (59.0)
Email	8 (8.0)

Caregiver variable	Value
Pictures	3 (3.0)
Music	7 (7.0)
Apps	4 (4.0)
Internet	19 (19.0)
Regular internet use (at least once a week), n (%)	92 (92.0)
Access internet from which device most?, n (%)	
Cell	58 (58.0)
Other device	13 (13.0)
Both equally	21 (21.0)
Not sure	8 (8.0)
Outcome of outgoing clinic calls to caregivers' mobile phone (n=2776), n (%)	
Connected	2698 (97.2)
Caregiver reached	1311 (47.2)
Left message	1263 (45.5)
Unable to leave voice message	124 (4.5)
Disconnected	75 (2.7)
Unknown/missing details	3 (0.1)
Number of unreachable days, median (range)	
Number of disconnected days	14 (2)
Number of unable to leave voice message days	28 (2)

Preference for Mobile Phone–Based Services and Support

When queried, only 72% of caregivers endorsed the need for nonspecific aftercare support; however, 91% of caregivers endorsed the desire for text messaging–based aftercare support (Table 3). Caregivers reported that text messages with the following content would be helpful: ways for improving communication with their child (63%), reminders and encouragement to use consequences (62%), suggestions for getting their teen involved in positive activities (62%), and messages with tips for monitoring their teen's substance use (56%). Caregivers also reported the desire for additional counseling for the child (32%) and general family/caregiver support (26%). Overall, 70.3% of caregivers preferred receiving texts 1-3 times weekly, 22% preferred daily, and 7.7% preferred 4-5 times weekly.

Demographic and Mobile Phone Relationships

Regression results showing relationships of caregiver and teen demographics with mobile phone characteristics, accessibility, and use patterns are presented in Tables 4 and 5. Younger caregivers were significantly more likely to have pay-as-you-go mobile phone contracts (β=0.06, *P*=.03), have a higher number of phone disconnections during treatment (β=–0.03, *P*=.04), and use their phone to access the internet (β=–0.07, *P*=.009). In addition, bilingual caregivers were significantly less likely to use texting as their main mobile phone feature (β=–0.87, *P*=.04). Caregivers with male teens were significantly more likely to have fewer disconnections during treatment (β=–0.46, *P*=.04) and those with adolescent girls were more likely to use texting as their main mobile phone feature (β=–0.99, *P*=.03).

Table 3. Aftercare support and clinic calls made to mobile phones (caregiver variables).

Variable	Value, n (%)
Additional support needed (those with access, n=100)	
Yes	72 (72.0)
No	6 (6.0)
Do not know	22 (22.0)
Interested in receiving text message parenting support (those with access, n=100)	
Yes	91 (91.0)
Support focused on (those with access, n=100)	
Monitoring substance use	56 (56.0)
Using consequences	62 (62.2)
Positive activities	62 (62.0)
Communication	63 (63.0)
Requested text frequency (those interested, n=91)	
1-3 times weekly	64 (70.3)
4-5 times weekly	7 (7.7)
Daily	20 (22.0)

Table 4. Logistic regressions for relations between demographics and self-report of mobile phone use, characteristics, accessibility, and reliability of service. The first group of predictors listed served as the comparison group for the binary logistic regression analyses and were coded as 0, while the second group was coded as 1. All technology outcomes were coded dichotomously with no=0 and yes=1.

Variable	Yearly contract			Text messaging used most			Phone number change once or more than once			Internet on phone			Self-report of connection issues		
	β	SE	95% CI	β	SE	95% CI	β	SE	95% CI	β	SE	95% CI	β	SE	95% CI
Caregiver															
Age	0.06[a]	0.03	1.01-1.12	0.04	0.03	0.99-1.09	–0.04	0.03	0.92-1.02	–0.07[b]	0.03	0.88-0.98	–0.06	0.03	0.88-1.01
Non-Hispanic vs Hispanic	–0.27	0.49	0.29-2.01	–0.30	0.48	0.29-1.89	1.04	0.55	0.96-8.32	–0.08	0.57	0.30-2.84	0.39	0.56	0.49-4.48
Nonbilingual vs bilingual	–0.28	0.43	0.32-1.77	–0.87[a]	0.42	0.19-0.95	0.30	0.42	0.60-3.06	–0.22	0.49	0.31-2.10	–0.24	0.47	0.32-2.00
Less than versus more than high school education	0.82	0.44	0.95-5.41	–0.04	0.05	0.88-1.06	–0.01	0.01	0.97-1.02	–0.30	0.50	0.28-2.00	–0.01	0.01	0.96-1.02
Child															
Age	0.10	0.17	0.80-1.53	0.09	0.16	0.81-1.49	0.23	0.17	0.91-1.74	–0.18	0.19	0.57-1.22	0.09	0.18	0.77-1.55
Female vs male	–0.42	0.44	0.28-1.57	–0.99[a]	0.45	0.15-0.89	0.06	0.43	0.46-2.48	0.06	0.51	0.39-2.86	0.08	0.48	0.42-2.78

[a]$P<.05$

[b]$P<.01$.

Table 5. Linear regressions for relations between teen demographics and objective measures of reliability of mobile phone service. The first group of predictors listed served as the comparison group for the binary logistic regression analyses and were coded as 0, while the second group was coded as 1.

Variable	Number of times ULM[a]			Number of times disconnected		
	β	SE	95% CI	β	SE	95% CI
Caregiver						
Age	–0.01	0.01	–0.04 to 0.01	–0.03[b]	0.01	–0.05 to 0.00
Non-Hispanic vs Hispanic	0.29	0.27	–0.24 to 0.82	0.26	0.25	–0.24 to 0.76
Nonbilingual vs bilingual	0.00	0.23	–0.46 to 0.46	0.08	0.22	–0.35 to 0.52
Less than versus more than high school education	0.01	0.01	–0.01 to 0.02	0.00	0.01	–0.01 to 0.02
Child						
Age	–0.02	0.09	–0.20 to 0.15	–0.05	0.08	–0.22 to –0.11
Female vs male	0.02	0.24	–0.46 to 0.49	–0.46[b]	0.22	–0.90 to –0.02

[a]ULM: unable to leave voice message.

[b]$P<.05$.

Discussion

Overview

Adolescent rates of return to substance use following outpatient treatment are staggering [1-3]. Research suggests that participation in aftercare and effective parenting practices posttreatment positively mediates adolescent substance use outcomes [7]; however, parent participation in aftercare is poor [9], likely due to structural and personal barriers [11,12]. One way to overcome barriers to participation is to provide parents with mobile phone–based aftercare. mHealth services are cost-effective, far-reaching, and on-demand. Further, mHealth services for low-income parents could help address socioeconomic disparities in access to aftercare services [19]. Prior to designing mHealth aftercare interventions for low-income caregivers of teens exiting treatment for substance use, it is imperative to conduct formative work to address the gaps in current knowledge. This study is the first to report mobile phone ownership, use, and accessibility and reliability of service in a low-income sample of caregivers of teens exiting treatment for substance use. Further, this is the first study to report on whether low-income caregivers of teens involved in substance use desire mHealth aftercare services. Lastly, this study provides the first report of the parenting skills caregivers would perceive as helpful if included in an mHealth aftercare program.

Summary of Principle Results

Results of this study replicate and extend existing research. Consistent with the extant literature, our results indicate that a high percentage of low-income caregivers own mobile phones and are familiar with mobile communication technology. Extending this literature, results also demonstrate good subjective report and objective analysis of accessibility and reliability of service despite most of the sample having pay-as-you-go service. Interestingly, older caregivers were more likely to have yearly contracts and fewer mobile disconnections. Additionally, caregivers with girls in treatment used their mobile phones more often for texting and those with boys in treatment had fewer mobile service disconnections. Further extending the current literature, most caregivers reported that they would like to receive mobile phone–based support in the form of text messages following their teen's treatment for substance use and reported interest in receiving support in areas of parenting that are common targets in evidence-based treatments.

Technology Characteristics and Use

The rates of mobile phone ownership among caregivers in this study are comparable to the current rates in the general population [20], yet the rates were higher than those of previous reports on low-income adult populations participating in substance use treatment [21-23]. This finding suggests that among caregivers with teens participating in treatment, access to mobile phones better matches national rates. One possible explanation for these different results is that caregivers may be more likely to own mobile phones to ensure communication with their teenager and to have a broader window of availability for other authority figures that may need to reach them (eg, schools, doctors, and juvenile probation).

Similarly, the rate of smartphone ownership was higher (83.5%) in this sample than in previous reports of adults involved in substance use treatment (67% [23] and 57% [22]). Rates of smartphone ownership were also higher than the national averages (77% [20]). Two factors may account for these differences: Smartphone ownership was less common in 2014 [32], and smartphone ownership among caregivers may be higher than that of noncaregivers. High rates of smartphone ownership among caregivers may be due to greater exposure to this technology through their teens [33]. In addition, research suggests that smartphone ownership has taken the place of computer ownership [33,34]. Thus, smartphone ownership among caregivers with teens offers a way of monitoring teen activity online. In support of this explanation, overall use of smartphone technology was higher in this population than in previous studies. Specifically, 76% reported using their phone for email compared to 45% in earlier research [21]. Furthermore,

92% reported accessing the internet compared to 44% [21] and 61% [23] reported by others.

Given that more Americans own smartphones than when previous research was conducted, a more equitable comparison of technology use to other studies may be the reported use of text messaging. Nonetheless, a notable difference in the use of this communication technology was evident in this study. McClure and colleagues [21] reported that 79% of their sample had access to and used text messaging. Dahne and colleagues [23] found a discrepancy between the percentage of people who had access to text messaging (96%) and the percentage of those who actually used the text messaging feature (83%) [23]. Compared to the values reported by Dahne and colleagues [23], Milward and colleagues [22] reported a much lower percentage of adults involved in substance use treatment who used text regularly (55%). We found that 97% of caregivers used text messaging regularly and only 3% reported texting limits. These previous findings suggest that for some populations, barriers may exist to implement text message–based interventions; however, our results suggest that text message–based interventions and support may be ideal for caregivers of teenagers.

Accessibility and Reliability of Service

Over half of the sample (65.6%) operated without a yearly mobile service contract. This result is similar to existing research of low-income adults involved in substance use treatment [22,23]. These results suggest that adults who are vulnerable to unmet treatment needs may have demographic characteristics in common, which may limit their use of yearly contracts. Similar to other studies, the majority of our participants were economically disadvantaged. However, our results for the number of times phone numbers were changed in the past year diverged from other reports—a variable considered to be a proxy measure of vulnerability for disruption in service and unreliable access to mobile phone technology. Caregivers in our study reported fewer instances of phone number changes in the past year (64% reported never) compared to other studies [37% [21] and 54% [23] reported never). This difference may also be related to their caregiver status and suggests that caregivers may be more reachable than other populations of adults involved in substance use treatment.

A unique feature of our study was the use of outbound calls made by clinic staff while families were enrolled in treatment, providing an objective indicator of disruption in mobile phone service. Results were consistent with self-report of low interruption of mobile phone service and show that participants were highly reachable despite the majority having pay-as-you-go mobile phone service. One interpretation of these results is that caregivers were reachable because of their involvement in treatment, that is, results may have shown a high rate of accessibility and service connection because caregivers were anticipating clinic calls related to appointments. Although this might have influenced outcomes, the finding that age of a caregiver is associated with reachability suggests that this explanation does not fully account for the current findings. Younger caregivers were more likely to experience disruptions in service and to have pay-as-you-go phone service. Thus, our results suggest that younger caregivers using a mobile phone–based intervention for teens involved in substance use treatment may benefit from assistance in maintaining mobile phone service. For example, mobile phone–based programs implemented by provider organizations could pay the cost of mobile service for some caregivers. The cost of maintaining mobile service during the high-risk period for adolescent relapse (first year posttreatment) costs less than continued in-person support by a mental health counselor. Overall, our results suggest that low-income caregivers of teens involved in substance use are reachable and have consistent access to the mobile phones that they own.

Aftercare Requests

Another unique feature of our study was the inclusion of a measure of caregiver interest in aftercare services. Interestingly, only 72% of caregivers expressed an interest in aftercare support. When asked specifically about aftercare delivered via text messaging, this rate increased to 91%. Few caregivers were able to provide details of what they were interested in receiving as part of aftercare support; however, when parenting skills as a topic were specifically surveyed, more than half of the caregivers endorsed content areas consistent with evidence-based treatments for adolescent substance use, including monitoring, use of consequences, ways to initiate positive activities, strategies for communicating with teen, and encouragement and support. These results suggest a mobile phone–based program for caregivers that includes skills covered in evidence-based curricula may be well received.

Notable Concerns

The prospect of using mobile phones to implement treatment is exciting because mobile phones offer the possibility of reducing health disparities along socioeconomic lines. Although these data suggest that low-income caregivers have access to mobile phones with reliable service, results revealed that demographic factors were related to technology use and reliability of service. First, bilingual caregivers were significantly less likely to use text messaging. This finding suggests that the specific communication technology selected for bilingual caregivers should be considered closely. Speaking two languages was unrelated to using a mobile phone to access the internet. It may be the case that digital interventions for bilingual caregivers are more likely to be adopted if delivered online. To help guide the development of programs wishing to implement a text message–based system for bilingual caregivers, additional research is necessary to better understand the reasons for lower rates of text messaging. For example, a significant proportion of this sample was bilingual but preferred English as their spoken language. A follow-up study could explore whether this finding holds true with a less acculturated sample.

Second, caregiver age may play a role in accessibility and reliability of service. Younger caregivers were significantly less likely to have a yearly contract. As noted above, it may be necessary for providers to offer social service support to caregivers based on age. Additional research is needed to determine the approximate age when caregivers are more likely to have a pay-as-you-go service plan to better target efforts in order to help caregivers maintain service during the

implementation of mobile phone–based interventions. Our results also showed that caregiver age was related to the mobile communication technology most frequently used. Older caregivers were least likely to use their mobile phone to access the internet, but age was unrelated to the use of text messaging. Additional research is needed to determine treatment delivery preference. Internet-based interventions optimized for access via mobile phones may be more successful with older caregivers. Text messaging and internet-based interventions may also be equally acceptable for older caregivers.

Third, caregivers parenting teen boys involved in substance use treatment were less likely to have phone disconnections and use text messaging most often among their mobile phone technologies. Future research is needed to better understand these associations. The association between child gender and caregiver use of mobile phone technology is critical for the design of mobile phone–based interventions for caregivers of teens involved in substance use treatment. These results may be an indication of caregiver efforts to monitor and supervise their teenager, suggesting that the use of mobile technology may differ by gender. These results may also be an indication of differences in what is required to monitor teen girls versus teen boys in this age of mobile technology. Researchers and providers may need to use more engagement efforts with caregivers of teen boys involved in substance use treatment when an intervention is designed for mobile delivery.

Limitations

Several limitations of this study should be noted, as they may impact interpretation and generalizability of the findings. First, since the goal of this study was to inform mHealth development of an aftercare service, the objective assessment of access and reliability of service may have provided different results if these data were collected once caregivers and teens were discharged from treatment. However, in a population of adults engaged in substance use treatment, Dahne and colleagues [25] showed rates of mobile phone ownership and use prior to treatment and expected ownership and use following treatment were similar among low-income adults. Additionally, in a study using ecological momentary assessment to examine medication adherence for diabetes among adolescents, researchers contacted teens through a phone call for a 10-day period following initial treatment [35]. During this time, researchers did not observe a decline in the response rate for these teens [35]. Although this period did not provide an extended assessment of reliability of mobile phone access among teens, it conveys a possible trend of consistent mobile phone access from treatment into aftercare. Importantly, children under the age of 18 years are more likely to have a mobile phone that is covered as part of a single-family plan [36], suggesting that caregivers of the youth in that study may have also experienced consistent and reliable mobile phone access from treatment into aftercare.

Second, our sample of predominately Hispanic caregivers was both a unique aspect of the study and a limitation. Minority caregivers are seldom represented in research, and there is no formative work with Hispanic caregivers of teens involved in substance use treatment. Yet, the generalization of the study results is limited. Result suggests that most caregivers in this sample were acculturated. Further, we were unable to explore associations among variables in the subsample of caregivers who identified as Hispanic.

Third, additional variables important in formative work for designing mHealth interventions for caregivers were not explored. For example, parent skills training usually includes a discussion of problematic child behaviors. We were unable to collect information about whether caregivers would want teens' health information included in an mHealth aftercare support program.

Finally, our analyses of the association between demographic variables and mobile phone characteristics and use, accessibility, and reliability of service should be interpreted with caution, given the exploratory nature of these analyses.

Conclusions

Results of this study suggest that the development of mobile phone–based interventions for caregivers of teens in substance use treatment is promising. The results of the survey demonstrate that mobile phone–based interventions designed for delivery using smartphone technology are feasible. Although caregivers have experience using most of the technology on their phones, usage differs by age and language. Self-report and objective data suggest that caregivers who have reliable access to mobile phones are interested in treatment delivered via mobile phones. Further research is needed to better understand the delivery of these services based on caregiver language, age, and gender of the child, as results suggest that tailoring may be needed.

Acknowledgments

Outcomes reported in this publication were supported by the Delivery Systems Reform Incentive Payment Program (085144601.2.101) and the National Center for Advancing Translations Sciences, National Institutes of Health, through Grant KL2 TR002646. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We would like to extend thanks to the caregivers who participated in this study and the supportive roles performed by valued clinic staff members: Luisana Campos, BA; Sharon Cates, MA; Marissa Mendoza, MA; and Isabella Santos, BA.

Conflicts of Interest

None declared.

References

1. Cornelius JR, Maisto SA, Pollock NK, Martin CS, Salloum IM, Lynch KG, et al. Rapid relapse generally follows treatment for substance use disorders among adolescents. Addictive Behaviors 2003 Mar;28(2):381-386. [doi: 10.1016/S0306-4603(01)00247-7]
2. Sinha R. New findings on biological factors predicting addiction relapse vulnerability. Curr Psychiatry Rep 2011 Oct;13(5):398-405 [FREE Full text] [doi: 10.1007/s11920-011-0224-0] [Medline: 21792580]
3. Waldron HB, Slesnick N, Brody JL, Charles W T, Thomas R P. Treatment outcomes for adolescent substance abuse at 4- and 7-month assessments. Journal of Consulting and Clinical Psychology 2001;69(5):802-813. [doi: 10.1037//0022-006X.69.5.802]
4. Henderson C, Rowe C, Dakof G, Hawes S, Liddle H. Parenting practices as mediators of treatment effects in an early-intervention trial of multidimensional family therapy. Am J Drug Alocohol Ab 2009;35(4):220-226. [doi: 10.1080/00952990903005890}]
5. Schmidt SE, Liddle HA, Dakof GA. Changes in parenting practices and adolescent drug abuse during multidimensional family therapy. Journal of Family Psychology 1996;10(1):12-27. [doi: 10.1037/0893-3200.10.1.12]
6. Black JJ, Chung T. Mechanisms of Change in Adolescent Substance Use Treatment: How Does Treatment Work? Substance Abuse 2014 Dec 11;35(4):344-351. [doi: 10.1080/08897077.2014.925029]
7. Winters KC, Lee S, Botzet A, Fahnhorst T, Nicholson A. One-year outcomes and mediators of a brief intervention for drug abusing adolescents. Psychology of Addictive Behaviors 2014 Jun;28(2):464-474. [doi: 10.1037/a0035041]
8. Stanger C, Budney AJ, Kamon JL, Thostensen J. A randomized trial of contingency management for adolescent marijuana abuse and dependence. Drug and Alcohol Dependence 2009 Dec;105(3):240-247. [doi: 10.1016/j.drugalcdep.2009.07.009]
9. Stanger C, Ryan SR, Scherer EA, Norton GE, Budney AJ. Clinic- and Home-Based Contingency Management Plus Parent Training for Adolescent Cannabis Use Disorders. Journal of the American Academy of Child & Adolescent Psychiatry 2015 Jun;54(6):445-453.e2. [doi: 10.1016/j.jaac.2015.02.009]
10. Godley SH, Garner BR, Passetti LL, Funk RR, Dennis ML, Godley MD. Adolescent outpatient treatment and continuing care: Main findings from a randomized clinical trial☆. Drug and Alcohol Dependence 2010 Jul 01;110(1-2):44-54. [doi: 10.1016/j.drugalcdep.2010.02.003]
11. Reardon T, Harvey K, Baranowska M, O'Brien D, Smith L, Creswell C. What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. Eur Child Adolesc Psychiatry 2017 Jan 4;26(6):623-647. [doi: 10.1007/s00787-016-0930-6]
12. Anderson J, Howarth E, Vainre M, Jones P, Humphrey A. A scoping literature review of service-level barriers for access and engagement with mental health services for children and young people. Children and Youth Services Review 2017 Jun;77:164-176. [doi: 10.1016/j.childyouth.2017.04.017]
13. Ozechowski T, Liddle H. Family-based therapy for adolescent drug abuse: Knowns and unknowns. Clin Child Fam Psychol Rev 2000;3(4):269-298. [doi: 10.1023/A:1026429205294]
14. Cunningham PB, Henggeler SW. Engaging Multiproblem Families in Treatment: Lessons Learned Throughout the Development of Multisystemic Therapy. Family Process 1999 Sep;38(3):265-281. [doi: 10.1111/j.1545-5300.1999.00265.x]
15. Harrison ME, McKay MM, Bannon, Jr WM. Inner-City Child Mental Health Service Use: The Real Question Is Why Youth and Families Do Not Use Services. Community Ment Health J 2004;40(2):119-131. [doi: 10.1023/B:COMH.0000022732.80714.8b]
16. Santiago CD, Kaltman S, Miranda J. Poverty and Mental Health: How Do Low-Income Adults and Children Fare in Psychotherapy? J Clin Psychol 2012 Dec 20;69(2):115-126. [doi: 10.1002/jclp.21951]
17. Humensky JL. Are adolescents with high socioeconomic status more likely to engage in alcohol and illicit drug use in early adulthood? Subst Abuse Treat Prev Policy 2010 Aug 5;5(1):1. [doi: 10.1186/1747-597X-5-19]
18. Wisdom JP, Cavaleri M, Gogel L, Nacht M. Barriers and facilitators to adolescent drug treatment: Youth, family, and staff reports. Addiction Research & Theory 2010 Nov 30;19(2):179-188. [doi: 10.3109/16066359.2010.530711]
19. Mensinger JL, Diamond GS, Kaminer Y, Wintersteen MB. Adolescent and Therapist Perception of Barriers to Outpatient Substance Abuse Treatment. Am J Addict 2006 Jan;15(s1):16-25. [doi: 10.1080/10550490601003631]
20. Pew Research Center Internet & Technology. 2017. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2017-11-04] [WebCite Cache ID 72HdjZi3o]
21. McClure EA, Acquavita SP, Harding E, Stitzer ML. Utilization of communication technology by patients enrolled in substance abuse treatment. Drug Alcohol Depend 2013 Apr 01;129(1-2):145-150 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.10.003] [Medline: 23107600]
22. Milward J, Day E, Wadsworth E, Strang J, Lynskey M. Mobile phone ownership, usage and readiness to use by patients in drug treatment. Drug Alcohol Depend 2015 Jan 1;146:111-115. [doi: 10.1016/j.drugalcdep.2014.11.001] [Medline: 25468818]
23. Dahne J, Lejuez CW. Smartphone and Mobile Application Utilization Prior to and Following Treatment Among Individuals Enrolled in Residential Substance Use Treatment. J Subst Abuse Treat 2015 Nov;58:95-99 [FREE Full text] [doi: 10.1016/j.jsat.2015.06.017] [Medline: 26231698]

24. Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014 Jan 13;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
25. Smart RG, Adlaf EM, Walsh GW. Neighbourhood Socio-Economic Factors in Relation to Student Drug Use and Programs. Journal of Child & Adolescent Substance Abuse 1994 Jul 25;3(1):37-46. [doi: 10.1300/J029v03n01_04]
26. Morales-Campos DY, Parra-Medina D, Esparza LA. Picture this!: using participatory photo mapping with Hispanic girls. Fam Community Health 2015;38(1):44-54 [FREE Full text] [doi: 10.1097/FCH.0000000000000059] [Medline: 25423243]
27. Mojica CM, Parra-Medina D, Yin Z, Akopian D, Esparza LA. Assessing media access and use among Latina adolescents to inform development of a physical activity promotion intervention incorporating text messaging. Health Promot Pract 2014 Jul;15(4):548-555 [FREE Full text] [doi: 10.1177/1524839913514441] [Medline: 24357863]
28. van Boekel LC, Brouwers EPM, van Weeghel J, Garretsen HFL. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: systematic review. Drug Alcohol Depend 2013 Jul 01;131(1-2):23-35. [doi: 10.1016/j.drugalcdep.2013.02.018] [Medline: 23490450]
29. Rieckmann T, Fussell H, Doyle K, Ford J, Riley K, Henderson S. Adolescent substance abuse treatment: Organizational change and quality of care. J Addict Offender Couns 2011 Apr;31(2):80-93 [FREE Full text] [doi: 10.1002/j.2161-1874.2011.tb00069.x] [Medline: 23750096]
30. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. International Journal of Information Management 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
31. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
32. Smith A. Pew Research Center Internet & Technology. Chapter One:A Portrait of Smartphone Ownership URL: http://www.pewinternet.org/2015/04/01/chapter-one-a-portrait-of-smartphone-ownership/ [accessed 2017-12-01] [WebCite Cache ID 72I1UBgeE]
33. Correa T, Straubhaar JD, Chen W, Spence J. Brokering new technologies: The role of children in their parents' usage of the internet. New Media & Society 2013 Oct 15;17(4):483-500. [doi: 10.1177/1461444813506975]
34. Olmstead K. Pew Research Center. A third of Americans live in a household with three or more smartphones URL: http://www.pewresearch.org/fact-tank/2017/05/25/a-third-of-americans-live-in-a-household-with-three-or-more-smartphones/ [accessed 2018-09-09] [WebCite Cache ID 72JGUZZ15]
35. Mulvaney SA, Rothman RL, Dietrich MS, Wallston KA, Grove E, Elasy TA, et al. Using mobile phones to measure adolescent diabetes adherence. Health Psychol 2012 Jan;31(1):43-50 [FREE Full text] [doi: 10.1037/a0025543] [Medline: 21967662]
36. Pew Research Center. Teens and Mobile Phones URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2010/PIP-Teens-and-Mobile-2010-with-topline.pdf%20 [accessed 2017-08-25]

Abbreviations

mHealth: mobile health
ULM: unable to leave voice message

Edited by G Eysenbach; submitted 03.10.18; peer-reviewed by A Gilmore, H Jin, SE Brown, K Davison; comments to author 25.10.18; revised version received 02.04.19; accepted 21.04.19; published 26.09.19.

Please cite as:
Ryan-Pettes SR, Lange LL, Magnuson KI
Mobile Phone Access and Preference for Technology-Assisted Aftercare Among Low-Income Caregivers of Teens Enrolled in Outpatient Substance Use Treatment: Questionnaire Study
JMIR Mhealth Uhealth 2019;7(9):e12407
URL: https://mhealth.jmir.org/2019/9/e12407
doi:10.2196/12407
PMID:31573920

Original Paper

A Tablet-Based App for Carpal Tunnel Syndrome Screening: Diagnostic Case-Control Study

Koji Fujita[1*], MD, PhD; Takuro Watanabe[2*], MEng; Tomoyuki Kuroiwa[1], MD, PhD; Toru Sasaki[1], MD; Akimoto Nimura[3], MD, PhD; Yuta Sugiura[2], PhD

[1]Department of Orthopedic and Spinal Surgery, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan
[2]School of Science for Open and Environmental Systems, Graduate School of Science and Technology, Keio University, Yokohama, Japan
[3]Functional Joint Anatomy, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan
[*]these authors contributed equally

Corresponding Author:
Koji Fujita, MD, PhD
Department of Orthopedic and Spinal Surgery
Graduate School of Medical and Dental Sciences
Tokyo Medical and Dental University
1-5-45 Yushima Bunkyo-ku
Tokyo, 1138519
Japan
Phone: 81 358035279
Email: fujiorth@tmd.ac.jp

Abstract

Background: Carpal tunnel syndrome (CTS), the most common neuropathy, is caused by a compression of the median nerve in the carpal tunnel and is related to aging. The initial symptom is numbness and pain of the median nerve distributed in the hand area, while thenar muscle atrophy occurs in advanced stages. This atrophy causes failure of thumb motion and results in clumsiness; even after surgery, thenar atrophy does not recover for an extended period. Medical examination and electrophysiological testing are useful to diagnose CTS; however, visits to the doctor tend to be delayed because patients neglect the symptom of numbness in the hand. To avoid thenar atrophy-related clumsiness, early detection of CTS is important.

Objective: To establish a CTS screening system without medical examination, we have developed a tablet-based CTS detection system, focusing on movement of the thumb in CTS patients; we examined the accuracy of this screening system.

Methods: A total of 22 female CTS patients, involving 29 hands, and 11 female non-CTS participants were recruited. The diagnosis of CTS was made by hand surgeons based on electrophysiological testing. We developed an iPad-based app that recorded the speed and timing of thumb movements while playing a short game. A support vector machine (SVM) learning algorithm was then used by comparing the thumb movements in each direction among CTS and non-CTS groups with leave-one-out cross-validation; with this, we conducted screening for CTS in real time.

Results: The maximum speed of thumb movements between CTS and non-CTS groups in each direction did not show any statistically significant difference. The CTS group showed significantly slower average thumb movement speed in the 3 and 6 o'clock directions (P=.03 and P=.005, respectively). The CTS group also took a significantly longer time to reach the points in the 2, 3, 4, 5, 6, 8, 9, and 11 o'clock directions (P<.05). Cross-validation revealed that 27 of 29 CTS hands (93%) were classified as having CTS, while 2 of 29 CTS hands (7%) did not have CTS. CTS and non-CTS were classified with 93% sensitivity and 73% specificity.

Conclusions: Our newly developed app could classify disturbance of thumb opposition movement and could be useful as a screening test for CTS patients. Outside of the clinic, this app might be able to detect middle-to-severe-stage CTS and prompt these patients to visit a hand surgery specialist; this may also lead to medical cost-savings.

(JMIR Mhealth Uhealth 2019;7(9):e14172) doi:10.2196/14172

KEYWORDS

carpal tunnel syndrome; screening test; movement; thumb

Introduction

Carpal tunnel syndrome (CTS) is a common condition that causes numbness, tingling, and pain in the hand, and is caused by a compression of the median nerve in the carpal tunnel, a narrow passageway on the palm side of the wrist [1]. CTS is the most common neuropathy and affects 5%-10% of women over the age of 40 years [2]. The initial symptom of CTS is numbness of the hand from the thumb to the ring finger; as the condition progresses, atrophy of the thenar muscle occurs [3]. Thenar muscle atrophy is strongly connected to failure of thumb motion [4], which causes problems in daily life, such as difficulty with picking up small items, fastening buttons, and opening bottles.

Patients tend to delay seeing a doctor until the numbness worsens; thus, in most cases, thenar atrophy has occurred by the first hospital visit [5]. Early-stage CTS, prior to thenar atrophy, can be treated conservatively by using a night splint, anti-inflammatory injection, or surgical intervention [1]; however, for advanced-stage CTS with thenar atrophy, carpal tunnel release surgery is the first choice [3,6,7]. Despite carpal tunnel release surgery for CTS with thenar atrophy, recovery of the atrophy takes longer than a year after surgery in most cases [8]. Therefore, performing surgery before thenar atrophy develops is key to avoid inconvenience in daily life activities.

To diagnose CTS, physical examination techniques are used, such as testing for the Tinel sign or performing the Phalen maneuver and the compression maneuver [1]; however, the sensitivity and specificity of these tests are not high [9,10]. Electrophysiological testing—analyzing the conduction velocity in the median nerve— reflects the condition and compression of the nerve itself accurately [11,12]. However, this test requires not only a skillful technician, but also a dedicated and expensive machine; thus, this test is not widely available [13].

With recent advances in technology, mobile phone or tablet devices can now be used as a diagnostic tool in several diseases [14,15] and can enhance patients' access to medical care in the early stage of disease [16]. Here, we developed a tablet app for CTS screening, focusing on thumb movement, and examined the usefulness of this app as a screening tool for CTS.

Methods

Study Design and Participant Recruitment

This study was approved by the Institutional Review Board of Tokyo Medical and Dental University. Paper-based informed consent was provided by all participants.

We recruited 22 female patients with CTS prior to surgery (CTS group, 29 hands) and 11 healthy female volunteers (non-CTS group, 11 hands) between January 2017 and July 2018. Upon recruitment, we obtained information about patients' chief complaints and histories of hand trauma. For all patients, we performed physical examinations and CTS induction maneuvers and obtained x-ray images of their hands. In the CTS group, patients were included if they had a primary diagnosis of CTS and planned to undergo carpal tunnel release surgery. The criteria for primary CTS diagnosis included numbness of fingers; CTS-specific physical findings, such as positive results on a compression test or the Tinel sign, as well as the Phalen test; and an abnormal value in nerve conduction velocity, measured by Neuropack X1 (Nihon Kohden), based on Bland's classification [17]. The following patients were excluded: patients with a history of hand injury or surgery; recurrence after release surgery of carpal tunnel; positive imaging findings indicative of first carpometacarpal or thumb metacarpophalangeal osteoarthritis, which could affect thumb motion; suspicion of disease on cervical spine; or positive magnetic resonance imaging findings of a space-occupying lesion in carpal tunnel.

As the control (non-CTS) group, female volunteers were included if they had undergone total hip arthroplasty in our hospital. We excluded patients from the non-CTS group if they had a history of wrist, hand, or finger injury or surgery; finger numbness; thumb pain; positive physical findings of CTS; or positive imaging findings of osteoarthritis of the first carpometacarpal or thumb metacarpophalangeal (see Table 1).

Table 1. Characteristics of participants in the CTS[a] and non-CTS groups.

Participant characteristics	Non-CTS group	CTS group
Number of participants, N	11	22
Age in years, median (IQR[b])	67 (58-74)	69 (59-80)
Sex (female), n (%)	11 (100)	22 (100)
Number of hands, n	11	29
Bland's classification		
Grade 1	N/A[c]	1
Grade 2	N/A	3
Grade 3	N/A	4
Grade 4	N/A	9
Grade 5	N/A	7
Grade 6	N/A	5

[a]CTS: carpal tunnel syndrome.

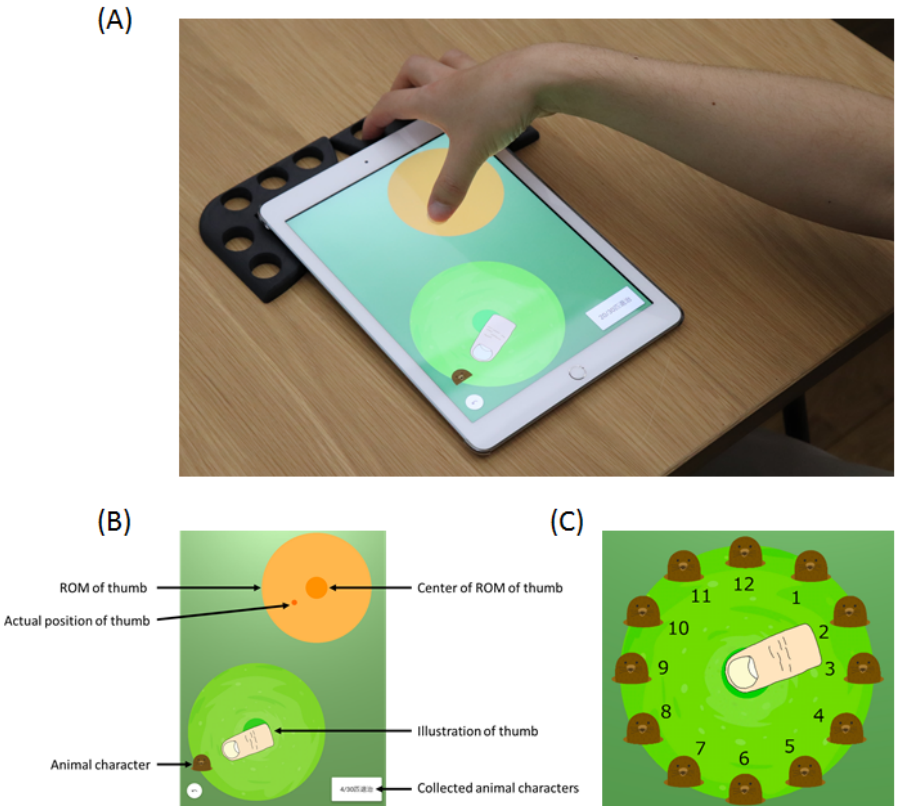
[b]IQR: interquartile range.

[c]N/A: not applicable.

App Design

The app was designed to run on a tablet device, in this case an iPad (Apple Inc); a 3D-printed holder, which fixed the fingers, was attached to the upper part of the tablet device in order to ensure use of only the thumb. A patient needed to slide the thumb along the touch screen to collect animal characters appearing on the screen (see Figure 1, A and B, and Multimedia Appendix 1). The animal characters appeared in 12 clock-like directions, in random order. In this way, it was possible to determine in which direction the patient's thumb movement was restricted (see Figure 1, C). The animal characters completely disappeared after appearing twice from the ground. If the patient missed the animal character in one direction, a new animal character appeared. The animal characters appeared in 12 directions centering on a large green circle; the screening assessment was made based on the movement ability of the thumb centered on this circle. We designed the app such that when the animal character had been collected or completely disappeared, a new animal character appeared at the center of the circle and guided the patient to return the thumb to the center. Thus, the animal characters were set as markers to monitor thumb movement ability; whether the patient could collect the animal characters was not a factor in the screening test.

Figure 1. Participants used this app with the thumb of each hand. (A) The index to small fingers were fixed to the holder. When the patient touched the orange circle with their thumb, the illustration of the thumb appeared. (B) The patient collected animal characters by controlling the illustration. (C) Animal characters appeared in 12 clock-like directions centered on the green circle. ROM: range of motion.

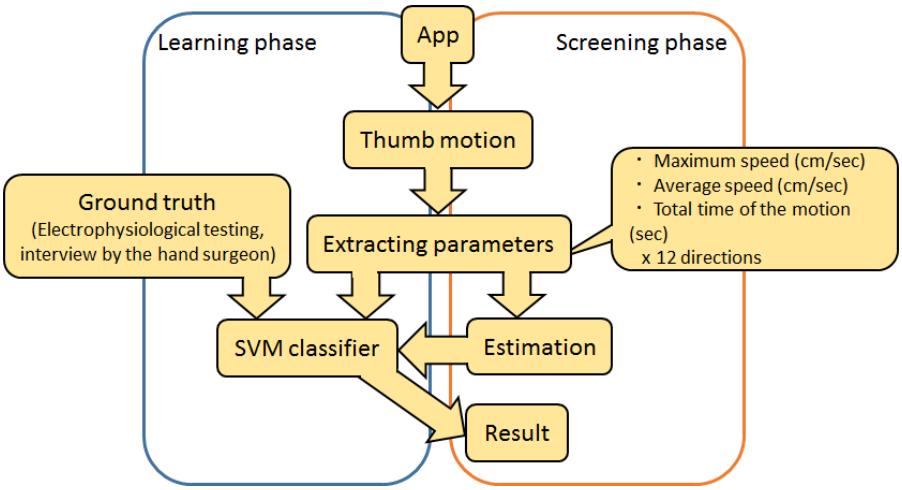


In this study, we used a two-class system to classify CTS or non-CTS using a support vector machine (SVM) [18]. SVM is a supervised learning algorithm for classification and regression analysis. In considering two-class classification, the boundary that distinguished between the two classes was called the decision boundary; the distance of the datum closest to the boundary for each class was called the margin. SVM determined the decision boundary to maximize the margin. The symptoms used as the basis for classification were comprehensively judged by the results of the electrophysiological testing and the medical examination by the hand surgeon. The training data were given by the maximum speed (cm/second), the average speed (cm/second), and the total time (seconds) of the thumb movement for the 12 directions in which the animal characters appeared. In total, 36 parameters per patient were obtained. We calculated the coincidence rate of the classification based on the electrophysiological testing and the app-based screening test.

Statistical Analysis

The raw data were collected in JavaScript Object Notation format and parsed to comma-separated value format. Hyperparameters used in SVM analyses were tuned using a grid search. The classification with the SVM adopted leave-one-out cross-validation. Leave-one-out cross-validation extracted one datum out of the dataset as a testing datum and used the rest as training data. This was repeated until each datum had been used as a testing datum. The SVM employed Python, version 3.7.0 (Python Software Foundation), and scikit-learn, version 0.20.2 (scikit-learn developers), a machine learning library (see Figure 2). The scikit-learn library played a role in the grid search, training, and validation. The Welch *t* test was used to compare changes between non-CTS and CTS individuals in each direction and was performed in R, version 3.5.1 (The R Foundation). A *P* value of less than .05 was considered as statistically significant.

Figure 2. The flow of carpel tunnel syndrome (CTS) screening with the app. The app extracted three parameters from the thumb motion. The app monitored thumb movements in 12 directions centered on a circle; data on three parameters were collected for each of the 12 directions. The parameters were used for support vector machine (SVM) training in the learning phase and to classify CTS or non-CTS in the screening phase.



Results

The maximum speed of thumb movements in the 12 directions was not significantly different (see Table 2). The CTS group showed significantly slower average thumb movement speed in the 3 and 6 o'clock directions (see Table 3) and also took significantly longer to reach the points in the 2, 3, 4, 5, 6, 8, 9, and 11 o'clock directions (see Table 4). Figure 3 shows the total thumb motion time.

By leave-one-out cross-validation, 27 of 29 CTS hands (93%) were classified as having CTS, and 2 of 29 CTS hands (7%) were classified as not having CTS. CTS and non-CTS individuals were classified with 93% sensitivity and 73% specificity (see Figure 4).

Table 2. Radar chart data of the maximum speed of thumb movements in 12 directions.

Direction	Maximum speed of thumb movements (cm/second), median (95% CI)		*P* value
	Control (non-CTS[a])	CTS	
12 o'clock	11.18 (7.79-21.89)	9.97 (8.56-14.52)	.56
1 o'clock	12.42 (9.45-20.22)	9.32 (7.73-12.11)	.11
2 o'clock	10.97 (7.30-17.37)	10.05 (8.37-13.65)	.94
3 o'clock	11.29 (8.35-15.64)	8.68 (7.45-13.19)	.34
4 o'clock	12.47 (9.21-16.00)	10.54 (8.89-14.99)	.48
5 o'clock	14.24 (8.98-22.14)	10.96 (8.71-14.19)	.06
6 o'clock	12.63 (6.82-16.12)	10.59 (7.94-12.73)	.47
7 o'clock	10.75 (6.07-19.30)	8.71 (6.40-12.62)	.40
8 o'clock	8.80 (7.39-16.78)	11.74 (6.90-15.52)	.58
9 o'clock	10.42 (6.47-15.53)	10.82 (6.03-12.85)	>.99
10 o'clock	10.70 (8.01-13.76)	9.50 (6.02-11.90)	.60
11 o'clock	10.81 (6.68-16.67)	9.18 (7.17-12.10)	.44

[a]CTS: carpal tunnel syndrome.

Table 3. Radar chart data of the average speed of thumb movements in 12 directions.

Direction	Average speed of thumb movements (cm/second), median (95% CI)		*P* value
	Control (non-CTS[a])	CTS	
12 o'clock	7.22 (4.96-12.16)	5.20 (4.32-6.60)	.16
1 o'clock	7.75 (4.87-10.80)	5.57 (4.52-8.62)	.33
2 o'clock	6.14 (4.91-9.58)	4.77 (3.70-7.48)	.94
3 o'clock	7.48 (4.33-10.70)	5.59 (3.20-6.59)	.03
4 o'clock	6.98 (4.46-9.53)	6.10 (3.44-7.56)	.21
5 o'clock	13.48 (5.78-15.25)	6.71 (3.41-7.82)	.06
6 o'clock	7.53 (5.80-12.01)	4.90 (2.99-6.38)	.005
7 o'clock	6.82 (5.02-14.53)	5.41 (3.37-6.42)	.06
8 o'clock	6.67 (4.81-13.13)	4.95 (3.48-6.71)	.50
9 o'clock	7.18 (4.83-10.18)	6.06 (3.57-7.63)	.38
10 o'clock	7.64 (5.57-11.16)	4.76 (3.09-7.16)	.16
11 o'clock	8.54 (4.27-12.02)	4.32 (3.13-5.73)	.04

[a]CTS: carpal tunnel syndrome.

Table 4. Radar chart data of the total time of thumb movements in 12 directions.

Direction	Total time of thumb movements (seconds), median (95% CI)		*P* value
	Control (non-CTS[a])	CTS	
12 o'clock	0.53 (0.33-0.90)	0.77 (0.58-0.97)	.11
1 o'clock	0.57 (0.38-0.93)	0.85 (0.63-1.10)	.34
2 o'clock	0.50 (0.40-0.88)	0.80 (0.68-1.10)	.01
3 o'clock	0.47 (0.37-1.00)	0.72 (0.55-1.30)	.01
4 o'clock	0.47 (0.38-0.58)	0.63 (0.57-1.07)	.01
5 o'clock	0.32 (0.23-0.50)	0.75 (0.50-1.18)	.003
6 o'clock	0.48 (0.35-0.58)	0.75 (0.60-1.40)	.01
7 o'clock	0.45 (0.27-0.65)	0.55 (0.45-0.95)	.13
8 o'clock	0.55 (0.25-0.80)	0.58 (0.40-1.08)	.02
9 o'clock	0.63 (0.40-0.73)	0.63 (0.58-0.92)	.02
10 o'clock	0.52 (0.32-0.75)	0.70 (0.43-0.95)	.06
11 o'clock	0.37 (0.30-0.82)	0.65 (0.43-1.12)	.01

[a]CTS: carpal tunnel syndrome.

Figure 3. Representation of the median time to reach a point. The carpal tunnel syndrome (CTS) group took longer to reach most of the 12 points.

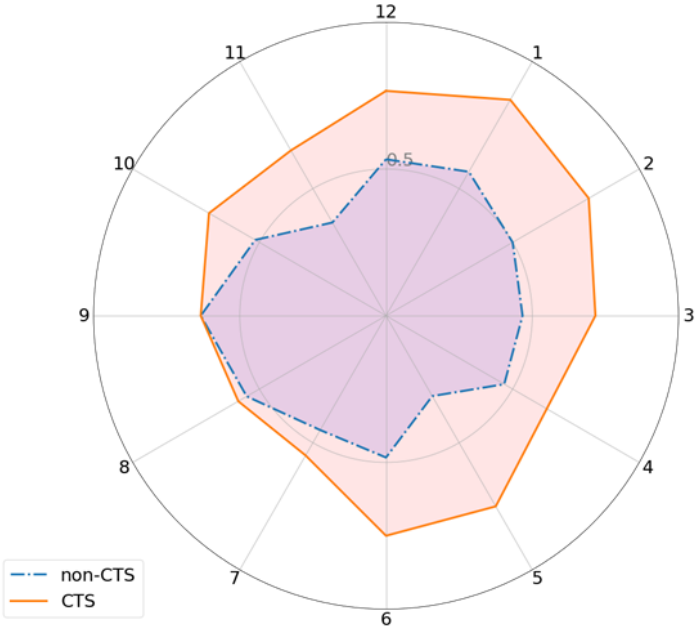
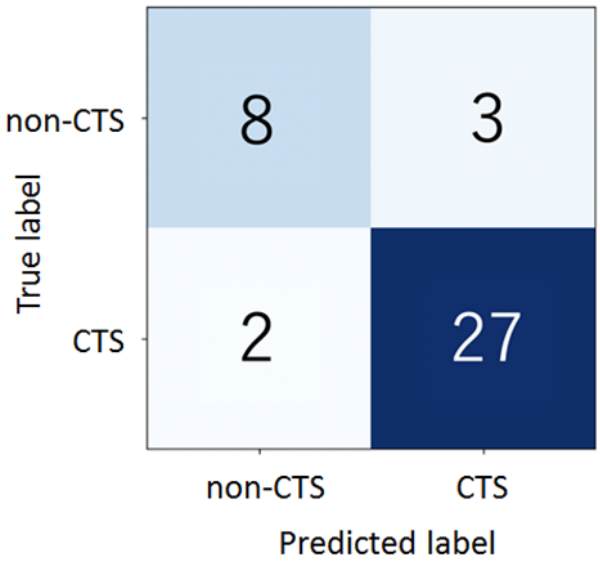


Figure 4. Representation of the median time in seconds to reach a point. The carpal tunnel syndrome (CTS) group took longer to reach most of the 12 points.



Discussion

Principal Findings

In this study, we developed a screening app for tablet devices that could detect thumb movement. With this app, CTS was detected with 93% sensitivity, 73% specificity, and 90% positive predictive value, which is as good as a diagnosis based on maneuver tests by a doctor. Testing for the Tinel sign showed 62% sensitivity, 93% specificity, and 88% positive predictive value; the Phalen maneuver showed 96% sensitivity, 80% specificity, and 79% positive predictive value.

Some diseases show specific finger or hand movements (eg, tremor in Parkinson disease), and doctors typically focus on these movements during the medical examination to diagnose and monitor the disease [19]. However, such visual information is difficult to quantify. Some studies have attempted to quantify these movements by using sensors or cameras. Motion capture analysis of the finger in cervical spondylosis [20] and small three-axis gyroscope analysis of thumb movement in CTS [4] have been reported. These methods have advantages in allowing precise measurement of detailed motion; on the other hand, special and precise devices and techniques are needed to achieve this and they are difficult to apply in clinical practice or daily life.

Tremor and writing disturbance in Parkinson disease have been well studied, and tablet or watch devices have come into clinical use recently [21-23]. Obtaining electrocardiograms or blood pressure by touch sensor can help monitor health status [24,15];

in addition, eye or skin diseases can be diagnosed using mobile phone cameras, and mental status can be examined through speech tone using mobile phone apps [25]. Widely used tablets and mobile phones have touch sensors, gyroscopes, cameras, and global positioning system. Development of apps is also relatively easy; therefore, use of these devices to analyze and assess the status of diseases in the daily living environment has gained attention. Furthermore, compared to medical devices, tablets and mobile phones are much cheaper.

In this study, the correct classification rate of CTS was approximately 90%, which was comparable to that obtained with some diagnostic maneuvers specific to CTS. We focused on disturbance of thumb opposition movements and tracking the movements while playing a game. On the other hand, we misclassified approximately 10% of hands. This may be because thenar atrophy had proceeded for a couple of months or years; acute onset of CTS will not cause thenar atrophy at the time of this test. Also, we included the patients with mild CTS status without thenar atrophy or thumb movement disturbance. To improve the screening accuracy, we plan to include a detection method of sensory disturbance and subjective scoring of CTS into the screening algorithm. Subjective scoring has been well studied in the clinical field; the Boston Carpal Tunnel Syndrome Questionnaire or the Disabilities of the Arm, Shoulder, and Hand questionnaire are routinely used to assess patients' status. These subjective questionnaires can be easily included into our app. With these developments, we plan to assess the severity of CTS using this app.

Limitations

There are some limitations to this study. First, we did not analyze other diseases with finger numbness, such as cervical spondylosis, diabetes neuropathy, or cubital tunnel syndrome. Thus, the specificity of screening in this app has not been well addressed. Second, conditions associated with thumb movement, such as trapeziometacarpal joint arthritis or a history of hand fracture, were not assessed. However, compared to these conditions, CTS is much more common, and thus this app will be useful as a screening tool. Furthermore, we made a plastic holder for the index, middle, ring, and small fingers, but the design of this holder requires further consideration to ensure consistency.

Conclusions

We showed that our newly developed app could classify disturbance of thumb opposition movement and can be useful as a screening test for CTS patients. Outside of the clinic, this app might be able to detect middle-to-severe-stage CTS and could prompt these patients to visit hand surgery specialists; this may also facilitate cost-savings.

Acknowledgments

This work was supported by grants from the Japan Science and Technology Agency (JST) Center for Advanced Intelligence Project—Public/Private R&D Investment Strategic Expansion PrograM (AIP-PRISM) (grant number: JPMJCR18Y2) and the JST Precursory Research for Embryonic Science and Technology (PRESTO) program (grant number: JPMJPR17J4.) We would like to thank Editage for English editing of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The participant used this app with the thumb. The index to small fingers were fixed to the folder.
[MP4 File (MP4 Video)24362 KB - mhealth_v7i9e14172_app1.mp4]

References

1. Wipperman J, Goerl K. Carpal tunnel syndrome: Diagnosis and management. Am Fam Physician 2016 Dec 15;94(12):993-999 [FREE Full text] [Medline: 28075090]
2. Middleton SD, Anakwe RE. Carpal tunnel syndrome. BMJ 2014 Nov 06;349:g6437. [doi: 10.1136/bmj.g6437] [Medline: 25378457]
3. Padua L, Coraci D, Erra C, Pazzaglia C, Paolasso I, Loreti C, et al. Carpal tunnel syndrome: Clinical features, diagnosis, and management. Lancet Neurol 2016 Nov;15(12):1273-1284. [doi: 10.1016/S1474-4422(16)30231-9] [Medline: 27751557]
4. Kuroiwa T, Fujita K, Nimura A, Miyamoto T, Sasaki T, Okawa A. A new method of measuring the thumb pronation and palmar abduction angles during opposition movement using a three-axis gyroscope. J Orthop Surg Res 2018 Nov 16;13(1):288 [FREE Full text] [doi: 10.1186/s13018-018-0999-3] [Medline: 30445972]
5. Fernandes CH, Meirelles LM, Raduan Neto J, Nakachima LR, Dos Santos JB, Faloppa F. Carpal tunnel syndrome with thenar atrophy: Evaluation of the pinch and grip strength in patients undergoing surgical treatment. Hand (N Y) 2013 Mar;8(1):60-63 [FREE Full text] [doi: 10.1007/s11552-012-9471-8] [Medline: 24426894]
6. Carlson H, Colbert A, Frydl J, Arnall E, Elliot M, Carlson N. Current options for nonsurgical management of carpal tunnel syndrome. Int J Clin Rheumtol 2010 Mar;5(1):129-142 [FREE Full text] [doi: 10.2217/IJR.09.63] [Medline: 20490348]
7. Kim P, Lee H, Kim T, Jeon I. Current approaches for carpal tunnel syndrome. Clin Orthop Surg 2014 Sep;6(3):253-257 [FREE Full text] [doi: 10.4055/cios.2014.6.3.253] [Medline: 25177448]

8. Garg B, Manhas V, Vardhan A, Srivastava DN, Das CJ, Vibha D, et al. Thumb opposition recovery following surgery for severe carpal tunnel syndrome: A clinical, radiological, and electrophysiological pilot study. J Hand Surg Am 2019 Mar;44(2):157.e1-157.e5. [doi: 10.1016/j.jhsa.2018.05.004] [Medline: 29934085]
9. Kuhlman KA, Hennessey WJ. Sensitivity and specificity of carpal tunnel syndrome signs. Am J Phys Med Rehabil 1997;76(6):451-457. [Medline: 9431262]
10. Wiesman IM, Novak CB, Mackinnon SE, Winograd JM. Sensitivity and specificity of clinical testing for carpal tunnel syndrome. Can J Plast Surg 2003;11(2):70-72 [FREE Full text] [doi: 10.1177/229255030301100205] [Medline: 24222987]
11. Sonoo M, Menkes DL, Bland JDP, Burke D. Nerve conduction studies and EMG in carpal tunnel syndrome: Do they add value? Clin Neurophysiol Pract 2018;3:78-88 [FREE Full text] [doi: 10.1016/j.cnp.2018.02.005] [Medline: 30215013]
12. Werner RA, Andary M. Electrodiagnostic evaluation of carpal tunnel syndrome. Muscle Nerve 2011 Oct;44(4):597-607. [doi: 10.1002/mus.22208] [Medline: 21922474]
13. Basiri K, Katirji B. Practical approach to electrodiagnosis of the carpal tunnel syndrome: A review. Adv Biomed Res 2015;4:50 [FREE Full text] [doi: 10.4103/2277-9175.151552] [Medline: 25802819]
14. 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]
15. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: Systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
16. Uchino M, Kawashima M, Uchino Y, Suzuki N, Mitamura H, Mizuno M, et al. The evaluation of dry eye mobile apps for screening of dry eye disease and educational tear event in Japan. Ocul Surf 2018 Oct;16(4):430-435. [doi: 10.1016/j.jtos.2018.06.002] [Medline: 29883737]
17. Bland JD. A neurophysiological grading scale for carpal tunnel syndrome. Muscle Nerve 2000 Aug;23(8):1280-1283. [Medline: 10918269]
18. Noble WS. What is a support vector machine? Nat Biotechnol 2006 Dec;24(12):1565-1567. [doi: 10.1038/nbt1206-1565] [Medline: 17160063]
19. Jankovic J. Parkinson's disease: Clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008 Apr;79(4):368-376. [doi: 10.1136/jnnp.2007.131045] [Medline: 18344392]
20. Omori M, Shibuya S, Nakajima T, Endoh T, Suzuki S, Irie S, et al. Hand dexterity impairment in patients with cervical myelopathy: A new quantitative assessment using a natural prehension movement. Behav Neurol 2018;2018:5138234 [FREE Full text] [doi: 10.1155/2018/5138234] [Medline: 30073036]
21. Hubble RP, Naughton GA, Silburn PA, Cole MH. Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: A systematic review. PLoS One 2015;10(4):e0123705 [FREE Full text] [doi: 10.1371/journal.pone.0123705] [Medline: 25894561]
22. Rovini E, Maremmani C, Cavallo F. How wearable sensors can support Parkinson's disease diagnosis and treatment: A systematic review. Front Neurosci 2017;11:555 [FREE Full text] [doi: 10.3389/fnins.2017.00555] [Medline: 29056899]
23. Wile DJ, Ranawaya R, Kiss ZH. Smart watch accelerometry for analysis and diagnosis of tremor. J Neurosci Methods 2014 Jun 15;230:1-4. [doi: 10.1016/j.jneumeth.2014.04.021] [Medline: 24769376]
24. Carpenter A, Frontera A. Smart-watches: A potential challenger to the implantable loop recorder? Europace 2016 Jun;18(6):791-793. [doi: 10.1093/europace/euv427] [Medline: 26847074]
25. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6:e856 [FREE Full text] [doi: 10.1038/tp.2016.123] [Medline: 27434490]

Abbreviations

AIP-PRISM: Center for Advanced Intelligence Project—Public/Private R&D Investment Strategic Expansion PrograM
CTS: carpal tunnel syndrome
IQR: interquartile range
JST: Japan Science and Technology Agency
N/A: not applicable
PRESTO: Precursory Research for Embryonic Science and Technology
SVM: support vector machine

Edited by G Eysenbach; submitted 28.05.19; peer-reviewed by SB Tsai, GE Iyawa, A Notarnicola; comments to author 18.06.19; revised version received 10.07.19; accepted 19.07.19; published 13.09.19.

Please cite as:
Fujita K, Watanabe T, Kuroiwa T, Sasaki T, Nimura A, Sugiura Y
A Tablet-Based App for Carpal Tunnel Syndrome Screening: Diagnostic Case-Control Study
JMIR Mhealth Uhealth 2019;7(9):e14172
URL: http://mhealth.jmir.org/2019/9/e14172/
doi:10.2196/14172
PMID:31586365

Original Paper

Validation of an mHealth App for Depression Screening and Monitoring (Psychologist in a Pocket): Correlational Study and Concurrence Analysis

Roann Munoz Ramos[1,2], PhD; Paula Glenda Ferrer Cheng[3,4], MA; Stephan Michael Jonas[1,5], Dr rer medic

[1]Department of Medical Informatics, RWTH Aachen University Hospital, Aachen, Germany

[2]College of Education, Graduate Studies, De La Salle University-Dasmarinas, Dasmarinas City, Cavite, Philippines

[3]Vivech System Solutions Inc, Manila, Philippines

[4]Department of Psychology, College of Science, University of Santo Tomas, Manila, Philippines

[5]Department of Informatics, Technical University of Münich, Münich, Germany

Corresponding Author:
Roann Munoz Ramos, PhD
Department of Medical Informatics
RWTH Aachen University Hospital
Pauwelstrasse 30
Aachen, 52074
Germany
Phone: 49 2418080352
Email: rramos@mi.rwth-aachen.de

Abstract

Background: Mobile health (mHealth) is a fast-growing professional sector. As of 2016, there were more than 259,000 mHealth apps available internationally. Although mHealth apps are growing in acceptance, relatively little attention and limited efforts have been invested to establish their scientific integrity through statistical validation. This paper presents the external validation of Psychologist in a Pocket (PiaP), an Android-based mental mHealth app which supports traditional approaches in depression screening and monitoring through the analysis of electronic text inputs in communication apps.

Objective: The main objectives of the study were (1) to externally validate the construct of the depression lexicon of PiaP with standardized psychological paper-and-pencil tools and (2) to determine the comparability of PiaP, a new depression measure, with a psychological gold standard in identifying depression.

Methods: College participants downloaded PiaP for a 2-week administration. Afterward, they were asked to complete 4 psychological depression instruments. Furthermore, 1-week and 2-week PiaP total scores (PTS) were correlated with (1) Beck Depression Index (BDI)-II and Center for Epidemiological Studies–Depression (CES-D) Scale for congruent construct validation, (2) Affect Balance Scale (ABS)–Negative Affect for convergent construct validation, and (3) Satisfaction With Life Scale (SWLS) and ABS–Positive Affect for divergent construct validation. In addition, concordance analysis between PiaP and BDI-II was performed.

Results: On the basis of the Pearson product-moment correlation, significant positive correlations exist between (1) 1-week PTS and CES-D Scale, (2) 2-week PTS and BDI-II, and (3) PiaP 2-week PTS and SWLS. Concordance analysis (Bland-Altman plot and analysis) suggested that PiaP's approach to depression screening is comparable with the gold standard (BDI-II).

Conclusions: The evaluation of mental health has historically relied on subjective measurements. With the integration of novel approaches using mobile technology (and, by extension, mHealth apps) in mental health care, the validation process becomes more compelling to ensure their accuracy and credibility. This study suggests that PiaP's approach to depression screening by analyzing electronic data is comparable with traditional and well-established depression instruments and can be used to augment the process of measuring depression symptoms.

(JMIR Mhealth Uhealth 2019;7(9):e12051) doi:10.2196/12051

KEYWORDS

mobile health; depression; validation; Psychologist in a Pocket; PiaP

Introduction

Background

Mobile technology has gained widespread acceptance and is seamlessly integrated in day-to-day activities, expanding especially into the field of health care. Mobile health (mHealth) is considered to be among the fastest growing sectors nowadays with a compound annual growth rate of 32.5% [1] and more than 259,000 apps available from over 59,000 publishers worldwide. Although mHealth apps definitely have their inherent appeal and value, very little attention and effort has been given to establish their scientific integrity or validity [2-4]. This is especially true in apps targeting mental health.

Validity ensures whether a novel approach is comparable with or is in agreement with the existing traditional methodology or instrument. Current scientific status of apps targeting mental health and behavioral disorders lack supporting data and empirical evidence on efficacy and outcome. Overall, studies on app validation and clinical effectiveness have not kept up with the pace of app development [5]. For instance, a scant 2% or 32 out of the 1536 downloadable mHealth apps for depression in 2013 were based on scientific publications [6]. Only 14 of 1065 articles on smartphone apps for bipolar and major depressive disorders reported having conducted scientific studies, mostly pilot or feasibility tests [7]. The United Kingdom's National Health Service has a list of 14 recommended apps in their library, 4 of which provide evidence based on patient reports [8].

In addition to the general lack of science-based development, most existing research on mobile technology and mental health care is methodologically limited with very small sample sizes [9,10] or are supported with feasibility studies only [11,12]. This shows the need for validation of accuracy and reliability of published apps.

The challenge of the validation process is the absence of a universal agreement on mHealth app metrics to identify high quality mobile apps, such as standardized evaluation and rating tools. Setting common evaluation benchmarks for existing health apps can be a challenging task because of their varied features, functions, and suitability. Although rating scales and classification platforms have been developed for mobile apps [4,13], these criteria cannot be implemented to all mHealth apps. Even major professional organizations, such as the American Psychological Association and the American Psychiatric Association, have yet to provide general guidelines as basis for mobile app evaluation [14]. The US Food and Drug Administration does not intend to regulate apps that appear to be of low risk nor transform a smartphone into a medical device [15].

Objective

This paper tackles the issue of mHealth app credibility by applying the psychometric approach of construct validation to a mobile app in mental health. Validation aims to determine whether or not relationships with other variables exist, and, if such relationships exist, to what magnitude. In this work, we focused on the validation of an app in depression detection through ecological momentary assessment (EMA).

EMA allows for a continuous detection of an individual's subtle and incremental mood changes during daily life. Compared with traditional psychological assessments such as self-reports and questionnaires, EMA's feature of real-time assessment avoids or reduces recall bias through recurrent and repeated data recording of daily cognitive and emotional dynamics. Various studies suggest that EMA provides accurate data regarding depression symptoms [16]. Mobile apps can support EMA through unobtrusive monitoring of day-to-day activities and social interactions.

The *Psychologist in a Pocket* (PiaP) [17] is an Android-based mental health app which aims to support and assist mental health professionals and complement traditional assessment approaches in depression detection and monitoring through EMA [18]. As it relies on EMA, PiaP reduces or eliminates the limitations of retrospective measurements (patient interviews and self-report) currently being used in mental health care assessment. Examples of the limitations that PiaP addresses are the reliance on the patient's memory and the overlooking of subtle or underreported symptoms by mental health practitioners.

PiaP's basic assumptions are as follows: (1) Everyday language—its usage, content, and themes—is a reliable indicator of the state of one's mental health; (2) Individuals tend to reveal personal information when using electronic media; and 3) Depressed or depression-prone individuals tend to self-focus and to ruminate on the negative aspects of their lives. PiaP aims at detecting changes in the nature of electronic text inputs through a lexicon of words in English and Tagalog related to depression, which were developed using both top-down and bottom-up processes (see [19] for app details and [18] for technical details). Sources for the lexicon were (1) symptom classification systems of the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 criteria for major depressive disorder and the International Statistical Classification of Diseases and Related Health Problems–10 criteria for depressive disorder, (2) focus group discussions, (3) interviews with mental health professionals, and (4) established psychological tests. As a result of these approaches, PiaP lexicon has a total of 13 symptom categories: mood, interest, appetite and weight, sleep, psychomotor agitation, psychomotor retardation, fatigue, guilt and self-esteem, concentration, suicide, alcohol and substance abuse, anxiety, and histrionic behavior. In addition, PiaP includes the category of first-person pronouns to reflect self-focus tendencies.

In the following sections, the construct validation of the PiaP depression lexicon is described. We hypothesize (Hypothesis 1, H1) that construct validity of the PiaP can be proven based on the measures for (H1.1) congruent, (H1.2) convergent, and (H1.3) divergent construct validations. In addition (Hypothesis 2, H2), statistical agreement of the PiaP with a test measuring the same variable (Beck Depression Index [BDI]-II) is hypothesized.

Methods

Tripartite Model of Test Construction

The development and validation of the PiaP lexicon is based on the tripartite model of test construction [20,21]. PiaP lexicon progressed through 3 stages, which are (1) theoretical-substantive (test items are generated according to theoretical requirements), (2) internal-structural (rational items are subjected to validation to establish internal consistency via construct validation, item analysis, and tests), and (3) external-criterion (entire test is investigated for its measurement of its construct as compared with other established measurement tools). A major advantage of this model is that it combines the strength of each phase in coming up with a reliable and valid measurement tool [22]. Items that are deemed to be inadequate are removed throughout the phases.

As PiaP is designed for depression-screening purposes, it underwent the technical phases of item or keyword construction. As a result, 2 versions (V1 and V2) of the PiaP lexicon were developed for validation. Stage 1 of the tripartite model provided the PiaP V1 keywords. Included are main keywords, derivatives of main keywords, and spelling variations (PiaP V1 total=835,286). During stage 2, PiaP V1 underwent internal validation to determine its internal psychometric properties (content validity, item analysis, and internal consistency). Only internally valid depressive-symptom keywords from PiaP V1 were included in PiaP V2 for use in stage 3 (external validation; PiaP V2 total=781,936).

Research proposal was first subjected to ethical review and approval by the Ethics Review Committee of the Graduate School, University of Santo Tomas (Manila, Philippines). After obtaining ethics approval, several potential universities were considered. Research letters were sent out to 6 universities in Manila and nearby provinces. Of the 6, 3 universities agreed to take part in the 3-stage study.

In this paper, only the results from stage 3 of the tripartite model are presented and discussed (see [19] for stages 1 and 2).

Participants

A total of 510 college students from stage 2 initially agreed to participate for 2 weeks during stage 3 of the research. Using homogenous sampling, they were purposively selected from Metro Manila colleges and universities, based on the following selection criteria: (1) must be enrolled in a tertiary academic institution at the time of data gathering, (2) should be aged between 16 and 25 years, (3) should have a mobile device that functions under Android operating system for PiaP to function, and (4) should have internet access at the time of PiaP download and upload of their encrypted data to the researcher. (Please see Multimedia Appendices 1 and 2 for sample screenshots; the presentation for the app is available in Multimedia Appendix 3).

Of the 510 participants, 332 could not be contacted immediately after inclusion despite follow-ups and reminders; thus, they were considered as *immediate dropouts*. After a 2-week administration of the PiaP V2, the remaining 178 participants were required to complete the following psychological tests to prove the research hypotheses: (1) Beck Depression Inventory (BDI)-II (H1.1 and H2); (2) Center for Epidemiological Studies–Depression (CES-D) Scale (H1.1); (3) Affect Balance Scale (ABS)–Negative Affect (H1.2); (4) Satisfaction With Life Scale (SWLS; H1.3); and (5) ABS–Positive Affect (H1.3).

Only 53 completed both the trial period and data collection. Participants (n=125) were excluded from data analysis for the following reasons:

- Sent empty encrypted psychological test files (n=2)
- Did not send encrypted psychological test files for unknown reasons (n=3)
- Did not send encrypted psychological test files because of internet problems (n=3)
- No data recorded owing to not following PiaP V2 setup instructions (n=4)
- Had changed phones (from Android to iPhone; n=5)
- Had Android version incompatibility with PiaP V2 (n=6)
- Dropped out (n=10)
- Experienced unexpected technical difficulties (n=10)
- Did not accomplish all psychological tests (n=33)
- Discontinued app after using PiaP V2 for a couple of hours/few days (n=49)

Data collection and analysis was based on 53 undergraduate students with a mean age of 17.42 (SD 1.03) years (Table 1). The average BDI-II score is 17.49 (SD 11.15), which is equivalent to a mild level of depressive symptoms.

Ethical Considerations

Voluntary participation was emphasized. Informed consent forms were distributed and filled up during each of the research stages. Moreover, participants were duly informed and reminded of the right to withdraw from the study at any time.

As privacy, data security, and anonymity of respondents were of paramount importance, several points were ensured:

1. Downloading the app needs only 1-time internet access. After downloading, PiaP runs offline. As a result, each of the participant's text inputs were stored locally (ie, in their mobile devices).
2. Only the researchers have sole and exclusive access to participant data (password protection). Participants were instructed to upload encrypted files to a designated cloud-based storage using the PiaP app. After data collection, all data were deleted or removed from the cloud storage.
3. In lieu of names, each participant was assigned and identified via a number code.

In addition, participants who were found to have significant BDI-II depressive symptom scores that warrant attention were individually referred to a clinical psychologist or counselor from their respective universities.

Table 1. Participant statistics (N=53).

Characteristics	Value
Gender (female), n (%)	43 (81)
Age (years), mean (SD)	17 (1)
Number of years at university, mean (SD)	2 (1)
BDI[a]-II score, mean (SD)	18 (12)
BDI-II level, n (%)	
Minimal	21 (40)
Mild	13 (24)
Moderate	7 (13)
Severe	12 (23)

[a]BDI: Beck Depression Inventory.

Construct Validation Process

In psychometrics, one type of validity is construct validity—the extent to which a measure adequately assesses the construct it purports to assess [23]. A *construct* (also known as *psychological construct*) is an attribute measured in a test. As a construct is generally not directly observable, this is validated through evidences of its relationships or correlations with psychometrically sound psychological tests, which either measure the same attribute or a different construct.

To accomplish this, 3 types of construct validity can be analyzed: (1) *Congruent construct validity* refers to a test's congruency or relationship with a known valid and reliable measure of the same construct [24] (eg, 2 measures of depressive symptoms); (2) *Convergent construct validity* correlates scores on a new test with the scores of established tests of related constructs [25] (eg, negative affect and depressive symptoms); and (3) *Divergent construct validity* provides discriminant evidence by proving that a particular test has low correlations with measures of unrelated constructs [26] (eg, life satisfaction and depressive symptoms).

To prove hypotheses H1.1, H1.2, and H1.3, the congruent, convergent, and divergent constructs needed to be selected.

The study's construct is *depressive symptoms*. It is characterized by negatively valenced words (words that describe unpleasant emotions) grouped according to 1 of the PiaP 13 symptoms based on a prior-developed lexicon and the frequency of first-person pronoun usage (see Cheng, et al [19] and Ramos, Cheng et al [27] for the development of the mentioned lexicon).

For congruent validity, the study characterization is compared with standardized tests for the same construct.

For convergent validity, the construct *negative affect* was chosen as previous researches have indicated a relationship between depression and negative affect [28]. Increases in negative affect, in response to everyday life challenges, reflect vulnerability to depression [29].

For divergent validity, the constructs *positive affect* and *life satisfaction* were chosen. As life satisfaction has been shown to be inversely associated with depression [30,31], positive affect and life satisfaction are considered to be a major indicator of subjective well-being [32]. For the convergent construct, *negative affect* was selected. *Positive affect*, similar to negative affect, is the emotional, affective component of subjective well-being. However, unlike negative affect, positive affect is the *pleasurable engagement with the environment* [33] and can be a protective factor against depression [34]. *Life satisfaction* is a distinct attribute as it constitutes the cognitive component of subjective well-being. It is an overall assessment about one's current life situation based on his or her personal criteria [32,35,36]. It is highly unlikely that a person who is satisfied with life can also be depressed at the same time [37].

Next, correlation was calculated to determine construct validity of PiaP (*depressive symptoms*) against the following psychological measures:

- Congruent construct validity (H1.1)
  - (1) BDI-II
  - (2) CES-D Scale
- Convergent construct validity (H1.2)
  - (3) ABS–Negative Affect component
- Divergent construct validity (H1.3)
  - (4) SWLS
  - (5) ABS–Positive Affect component

Note that BDI-II and CES-D Scale measure depressive symptoms before testing. Therefore, the PiaP total scores (PTS) of each respondent spanning 2 weeks and 1 week were correlated with BDI-II and CES-D Scale, respectively.

Statistical Analysis

In determining the construct validity of PiaP against the psychological measures used in the study, Pearson product-moment correlation (PPMC) of scores on all tests were calculated [38]. PPMC was employed to determine the strength of association between PiaP's interval scales scores with each of the psychological tests. In this research, positive correlations are evidences of congruent and convergent validities, whereas negative correlations are expected in divergent construct validation.

Study findings are explained according to Hinkle et al's [39] rule of thumb in interpreting the size of the correlation coefficient (Table 2).

To determine the practical significance of the results, Cohen *d* effect size (ES) was used to interpret the correlation values (Table 3). ES presents the magnitude of reported effects in a standardized manner, regardless of the scale used to measure a variable [40].

Although correlation quantifies the degree of relation, it does not automatically imply good agreement between 2 methods. Thus, to prove H2, further statistical validation to compare 2 different types of measurements (PiaP and BDI-II) of the same variable (depression symptoms) was performed by applying Bland-Altman (B-A) plot and analysis. The researchers selected BDI-II as the established psychological test with which PiaP was compared, as this test is considered the gold standard of self-rating scales designed to measure the current severity of depressive symptoms [41].

Psychologist in a Pocket Normative Structure

PiaP's set of norms was based on data collected from 924 days of PiaP usage of 510 randomly selected college student participants from the study's stage 2. Participants' average number of days of PiaP usage is 10.62. The overall tally of text inputs per day of all relevant words (regardless of symptom category) detected by the depression lexicon is referred to as the PiaP total score (PTS). Specifically, the PTS is increased by 1 score point for each typed word present in the PiaP lexicon. During the 2-week period, a total of 31,336 text inputs from all the participants was obtained, with an average of 11.40 (SD 17.77) text inputs per daily evaluation, with a score range of 0 (no depression-related keyword detected in text inputs) to 164 (maximum number of text inputs detected as matching the keywords in the depression lexicon).

For the interpretation of the PTS, quartiles were calculated to determine the levels of depressive symptoms from normal to critical (Table 4). The normal level represents scores from individuals who do not experience depression yet had typed words representative of depression and its symptoms (eg, for research purposes). Score ranges from above normal to critical levels signify that the text inputs suggest varying degrees of depression as detected by the lexicon.

It is important to note that gender-specific norms were not created as studies with adolescents conclude that gender does not influence depressive symptomatology [42,43].

Psychological Tests

Beck Depression Inventory–II

BDI–II [44,45] is a 21-item self-report measuring the intensity of current depressive symptoms (sadness, pessimism, loss of pleasure, etc) based on the DSM, particularly for ages 13 to 80 years. Respondents report each symptom on a 4-point Likert scale retrospectively for the 2 weeks prior the test. The highest possible score is 63 with minimal (0-13), mild (14-19), moderate (20-28), and severe (29-63) ranges.

Center for Epidemiological Studies–Depression Scale

The CES-D Scale, initially developed for epidemiological research, is a 20-item screening tool to detect current depressive symptoms during the week before taking the test, with an emphasis on depressed mood [46,47]. It covers 4 factors: depressive affect, somatic symptoms, positive affect, and interpersonal relations. Respondents choose on a 4-point Likert scale. Scores of 16 and above indicate significant symptoms, with 60 as the highest possible score.

Affect Balance Scale

ABS [48] targets objective well-being through the assessment of positive and negative affect. The 10-item scale focuses on feelings experienced by respondents over the past few weeks, with 5 items each to describe positive and negative affect. Respondents choose on a binary scale *Yes* (score of 1) or *No* (score of 0). Total affect balance score is computed by subtracting the negative affect score from the positive affect score and then adding a constant of 5 to avoid values below 0. A score of 0 means low affect balance, whereas 10 reflects high affect balance.

Table 2. Interpreting correlation values.

Absolute size of correlation	Interpretation
0.90 to 1.00	Very high positive (negative) correlation
0.70 to 0.90	High positive (negative) correlation
0.50 to 0.70	Moderate positive (negative) correlation
0.30 to 0.50	Low positive (negative) correlation
0.00 to 0.30	Negligible correlation

Table 3. Interpretation of Cohen *d* (effect size).

Effect size	Interpretation
0.50	Large
0.30	Medium
0.10	Small

Table 4. Psychologist in a Pocket total score interpretation.

Level	Brief description	Psychologist in a Pocket total score range (text input)
Normal	Typical or average number of depression-related keywords typed by an individual without depression	0-19
Above normal	Higher than average amount of depression-related keywords typed by an individual with some (mild) signs of depression	20-38
High	Considerable amount of depression-related text inputs by an individual with possible moderate signs of depression	39-65
Critical	Elevated amount of depression-related text inputs by an individual with a possible clinical or serious case of depression	66-164

Satisfaction With Life Scale

The SWLS is designed to measure life satisfaction as a whole and does not tap positive or negative affect, happiness, or satisfaction related to various life domains [49]. Participants indicate how much they agree or disagree with each of the 5 items measuring global satisfaction using a 7-point scale. Participants within the higher score range of 30 to 35 consider life as enjoyable and that major domains of life are well. Scores between 5 to 9 reflect extreme dissatisfaction in multiple areas of life.

Results

Descriptive Statistics

In Table 5, we present an overview of the measures used in this study. The number of observations for PiaP reflect the 1-week and 2-week tallies of depression-related keywords (relevant inputted keywords) of the 53 participants as identified by the PiaP depression lexicon. As CES-D Scale is covering only 1 week, it was correlated with the 1-week period, whereas data from the 2-week period was used to correlate with BDI-II scores. There was a notable decrease of depression-related keywords in the second week of PiaP administration.

Depression levels of the participants range from mild to moderate, as indicated by their mean scores in the 2 depression measures used, BDI-II and CES-D Scale. Score in ABS, which comprises ABS–Positive Affect and ABS–Negative Affect, reflect an average level of happiness (ABS total score=5.66). However, for the purposes of this research, we looked at these 2 scale components separately. Participants reported having mild negative affect while experiencing moderate positive affect. Finally, participants are slightly satisfied with their lives, as inferred from the SWLS mean score.

Hypothesis 1: Construct Validity Correlations

Table 6 presents the correlation coefficient results for the 3 construct validation approaches of 1-week and 2-week PTS with each of the psychological instruments used.

The exact *P* values have been provided below.

Table 5. Descriptive statistics (Psychologist in a Pocket and psychological tests).

Measure (score range)	Number of observations	Mean (SD)	Interpretation
PiaP[a] 1-week (0-3154)	3154 keywords	59.64 (78.238)	High
PiaP 2-weeks (0-5214)	5214 keywords	101.06 (93.140)	Critical
BDI[b]-II (0-63)	53 participants	17.49 (11.154)	Mild
CES-D Scale[c] (0-60)	53 participants	19.81 (10.958)	Moderate
ABS[d]–Negative Affect (0-5)	53 participants	2.49 (1.589)	Mild
ABS–Positive Affect (0-5)	53 participants	3.15 (1.199)	Moderate
SWLS[e] (5-35)	53 participants	20.58 (5.716)	Average

[a]PiaP: Psychologist in a Pocket.

[b]BDI: Beck Depression Index.

[c]CES-D Scale: Center for Epidemiological Studies–Depression Scale.

[d]ABS: Affect Balance Scale.

[e]SWLS: Satisfaction With Life Scale.

Table 6. Construct validation results (correlation coefficient) and hypothesis (N=53 for all analyses).

Psychological tests	Psychologist in a Pocket, correlation coefficient		Effect size	Hypothesis	Hypothesis support
	1-week	2-week			
BDI[a]-II	—[b]	0.50[c]	Large	Hypothesis 1.1	Yes
CES-D Scale[d]	0.42[c]	—	Medium	Hypothesis 1.1	Yes
ABS[e]–Negative Affect	0.25	0.19	N/A[f]	Hypothesis 1.2	No
ABS–Positive Affect	−0.29[g]	−0.20	Medium	Hypothesis 1.3	Yes
SWLS[h]	−0.29[g]	−0.32[g]	Medium	Hypothesis 1.3	Yes

[a]BDI: Beck Depression Index.

[b]Not applicable.

[c]Significant finding *P*=.01.

[d]CES-D Scale: Center for Epidemiological Studies–Depression Scale.

[e]ABS: Affect Balance Scale.

[f]No effect size due to no significant correlation between PTS and ABS-Negative Affect.

[g]Significant finding *P*=.05.

[h]SWLS: Satisfaction With Life Scale.

Congruent Construct Validity (Hypothesis 1.1): Correlations Between Psychologist in a Pocket and Depression Tests

PiaP's construct, *depression symptoms*, was validated with 2 psychological tests of depression. Using PPMC, congruent construct validity was determined by correlating the participants' (1) 1-week PTS with CES-D Scale scores and (2) 2-week PTS with BDI-II scores. These PiaP timeframes were considered as CES-D Scale instructs the respondents to recall depressive symptoms occurring for the week before testing, whereas BDI-II evaluates depressive symptoms for the previous 2 weeks before test administration. At 0.01 level of significance (2-tailed), results show significant low to moderate positive correlations between (1) PiaP and CES-D Scale (r=0.42, n=53, P=.002) and (2) PiaP and BDI-II (r=0.50, n=53, P<.001), respectively. Furthermore, Cohen *d* 's ES values for 1-week PTS and CES-D Scale (d=0.42) and 2-week PTS and BDI-II (d=0.50) suggest a moderate to high practical significance, respectively.

Convergent Construct Validity (Hypothesis 1.2): Correlations Between Psychologist in a Pocket and Affect Balance Scale–Negative Affect

Although the correlations are positive, they are not significant. There is no significant correlation between the 2-week PTS and ABS–Negative Affect scores (r=0.19, n=53, P=.17). In addition, there is no significant correlation between the 1-week PTS and ABS–Negative Affect scores (r=0.25, n=53, P=.07). In addition, Cohen *d* 's ES indices for both ABS–Negative Affect and (1) 1-week PTS (d=0.25) and (2) 2-week PTS (d=0.19) indicate low practical significance.

Divergent Construct Validity (Hypothesis 1.3): Correlations Between Psychologist in a Pocket with Affect Balance Scale–Positive Affect and Satisfaction With Life Scale

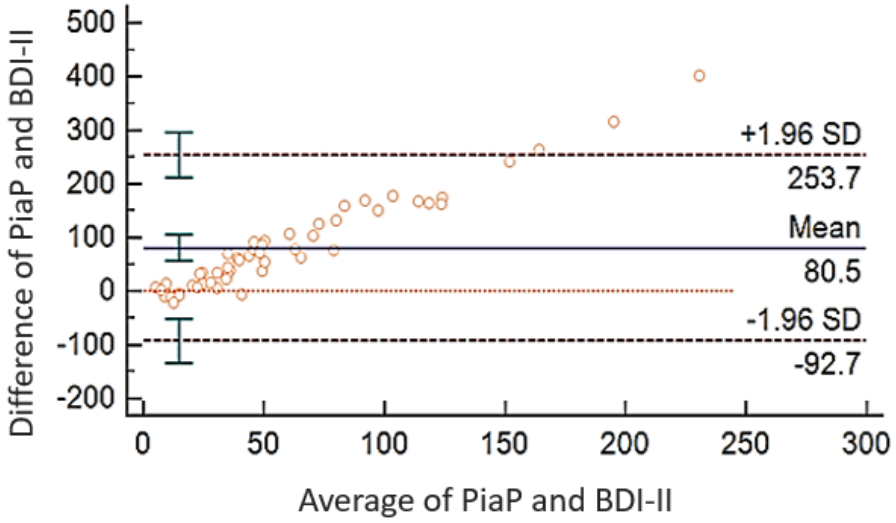
At 0.05 level of significance (2-tailed), a significant but negligible correlation exists between 1-week PTS and ABS–Positive Affect (r=−0.29, n=53, P=.04). A negative but nonsignificant relationship exists between 2-week PTS and ABS–Positive Affect (r=−0.20, n=53, P=.15). Cohen *d* 's ES for both ABS–Negative Affect and (1) 1-week PTS (d=−0.29) and (2) 2-week PTS (d=−0.20) results are in the low practical significance range.

A significant but negligible correlation at 0.05 level of significance (2-tailed) was also obtained between SWLS and 1-week PTS (r=−0.29, n=53, P=.04), whereas there is a low positive significant correlation at 0.05 level of significance between SWLS and 2-week PTS (r=−0.32, n=53, P=.02). Cohen *d* 's ES for SWLS and (1) 1-week PTS (d=−0.29) and (2) 2-week PTS (d=−0.32) scores are in the low to moderate practical significance range, respectively.

Hypothesis 2: Concordance Analysis

MedCalc statistical software [50] was used to compute and to create the B-A plot. The concordance between the difference of PiaP and BDI-II scores and the average of PiaP and BDI-II scores is analyzed (Figure 1). Mean difference of raw scores is 80.50, which is within the CI of 56.1289 to 104.8522. Limits of agreement values are from −92.7 to 253.7. Upper confidence limit of 253.7 falls within the upper 95% CI limit (CIL; 211.8261 to 295.6209), whereas the lower confidence limit of −92.7 is within the range of the lower 95% CIL (−50.8449 to −134.6397). Out of 53 participants, only 3 were outliers.

Figure 1. Bland-Altman plot analysis of Psychologist in a Pocket (PiaP) and Beck Depression Index-II (BDI-II).



Discussion

Primary Contribution

Together with our prior work on lexicon development and content validation [19], this work concludes the tripartite model of test construction on the PiaP. To the best of our knowledge, this is the first time a mobile mental health app has been validated according to the tripartite model of test construction.

Construct validity correlations show correlation with congruent construct, and the concordance analysis further indicates that the PiaP's lexicon is able to reproduce standard test findings. In addition, PiaP is EMA-based and, therefore, does not rely on memory. Symptoms that are easily overlooked by psychological tests can be detected in a more timely manner. In addition, mobile phone–captured data might be more sensitive than paper-and-pencil–collected data [51]. Thus, PiaP can be an addition to the classical pen-and-paper tests and give a more detailed picture on mood changes.

Although the congruent correlation values of PiaP with the BDI-II and the CES-D Scale reflect that they measure the same construct, ES values quantify (1) the differences between PiaP with the 2 paper-and-pencil tests and (2) PiaP's effectiveness to screen for depression symptoms via text analysis. Furthermore, this shows that mobile phones offer a platform where language can be studied and used to identify people with depression through their free texts and novel ways of communication. For PiaP users, this could mean a more feasible and comfortable way of reporting their symptoms, while providing a reliable, immediate, and more encompassing screening (and monitoring) of depression symptoms.

Although correlation for convergent and divergent constructs seem low, this is expected as high correlation should mostly occur for the congruent construct. Simply put, convergent and divergent constructs behave similar (or similar inverted) to the intended measure but not identical. Thus, no perfect correlation should be reached.

General Remarks

More than 5000 observations or text inputs of depression-related words were made by PiaP during the 2-week test period. The resulting high SD values of PiaP scores indicate great variability in the number of responses between the participants. This variability is likely because of the nature of text inputs. Logging of text messages and text evaluations are based on free text inputs during daily usage without any specific prompts. This PiaP approach to depression detection is unlike structured psychological (depression) tests, wherein replies to target questions or stimuli require a specific kind of response. In addition, PiaP texts are captured in real time or close in time to experience, allowing for a steady and unlimited detection of numerous and varying mood changes.

The decrease in the number of depression text inputs from the participants (from 3154 inputs in week 1 to 2060 inputs in week 2) may be attributed to academic-related factors. In week 2 of data gathering, there was presumably lesser stress in the preparation of class requirements and exams before the Christmas break, whereas higher academic pressure in week 1 may have led to depression and anxiety [52] or perceived lack of achievement [53].

Low to moderate correlations between PiaP and the psychological tests utilized may be because of the restriction in the range of scores included in the sample. Restricted range occurs when the scores of 1 or both variables in a sample have a range of values that is less than the range of scores in the population [26], thus reducing the correlation found in a sample relative to the correlation that exists in the population. As only 53 participants successfully complied with the required 2-week PiaP run and the completion of psychological tests, this limited the range of scores available for analysis.

The large quantity of items or keywords in the PiaP lexicon may have contributed to the low or insignificant correlation results. This is not surprising as the psychometrics of word usage is in contrast with the typical test development such that compiled words in lexica are not normally distributed, have low base rates, and do not adhere to the traditional psychometric

laws. Thus, standard reliability measures are not always appropriate in such a scenario [54].

Hypothesis 1: Construct Validity Correlations

Congruent Construct Validation (Hypothesis 1.1)

The congruent construct validation attempts to determine whether the construct or attribute of the psychological approach in question correlates with a gold standard. Significant positive correlations with BDI-II and CES-D Scale imply that PiaP's measure is compatible with the depressive symptoms measured in BDI-II and CES-D Scale. In addition, ES provides additional meaning to the results by providing more concrete and meaningful interpretations. In this study, ES ranged from medium to high, implying that depression signs are observable in their text inputs.

Convergent Construct Validation (Hypothesis 1.2)

Contrary to the study's hypothesis, there is no significant correlation between depression and negative affect. This finding might be because of the fact that depression is a phenomenon with complex and varied features. In addition, the experience of depression might not be manifested through negative affect alone nor its absence demonstrated through positive affect or positive emotion. As Beck suggested in the cognitive theory of depression, negative thought processes and rumination, which are common and debilitating aspects of depression, should be the main focus of evaluation, as depression displays itself in negative thinking before it creates negative affect or mood [55].

Divergent Construct Validation (Hypothesis 1.3)

Divergent constructs of *positive affect* and *life satisfaction* were hypothesized to be inconsistent with the experience of depression.

Positive affect has a weak to negligible correlation. This suggests that, although positive affect has been shown to be low or absent in an individual experiencing depression, it is independent from negative affect, regardless of the intensity of affective experience [56]. Positive affect and negative affect are 2 broad mood factors which are salient in self-reported mood [33]. Having low levels of positive affect may not immediately point to negative affectivity but may be manifested as lethargy or fatigue. Among the participants, low levels of positive affect were consistently related only to depressive symptoms such as loss of pleasurable engagement.

Life satisfaction appears to be the stronger contrary attribute to depressive symptoms, as evidenced by the more stable and consistent negative correlation between PiaP and SWLS. Life satisfaction is a (negative) predictor of depression [57], second only to negative thoughts. Sample text inputs of research participants who obtained low scores in SWLS fall under the following PiaP categories: depressed mood, suicide, loss of interest, and fatigue.

Hypothesis 2: Concordance Analysis (Bland-Altman plot and analysis)

Concordance analysis reveals that PiaP's evaluation of depression symptoms via text or lexical analysis is comparable with the use of BDI-II, implying that PiaP is able to identify the presence of depressive symptoms similar to commonly used structured depression tests. It indicates that PiaP's lexica are valid depression indicators as reflected in BDI-II. It likewise suggests that PiaP's text analysis approach is able to reveal current psychological states, making it comparable with BDI-II's appraisal of current symptoms of depression.

In addition, PiaP's degree of agreement with BDI-II implies that it can support continued mental health appraisal, such as in an ongoing depression monitoring and screening of patients in between their appointments with doctors and/or therapy sessions.

Limitations

One limitation of this work is the high dropout attrition rate. Despite having agreed to take part in both stages 2 and 3 of this study, a sizeable proportion of participants did not respond to follow-ups for stage 3. Although high attrition rates are avoided in traditional clinical trials, such a phenomenon is a naturally occurring and distinct feature of remote electronic health trials [58,59]. In addition, adherence to mental health care apps tend to be poor among individuals with mild to severe depression [60]. As a result of the high attrition rate, the final research group consisted only of 53 participants. This lower-than-expected sample size may undermine the study's significant findings. However, the researchers applied the 3 approaches to external validation and, to strengthen the positive correlation results, added the B-A analysis particularly for the congruent construct validation. In addition, the medium-to-high ES values imply that the effectiveness of PiaP's approach in identifying depression symptoms, as compared with paper-and-pencil tests, is consistent and obvious.

A second limitation of PiaP is the limitation to text input. Behavioral symptoms [61] or weight change and appetite disturbance [61] could be important in detecting a person with depression. The individual's behavioral or motoric expressions of affect may not have been clearly detected as they are more difficult to verbalize. Hence, it is suggested that PiaP be validated with behavioral markers of depression such as movement and sleep patterns.

Finally, several results have either significant yet low correlation or no correlation. As previously mentioned, depression is a complex condition with cognitive, affective, and behavioral manifestations. As PiaP scoring relies on language usage, which tends to reflect the cognitive and affective elements of depression, the app is unable to screen for behavioral signs of depression, which cannot be expressed via text.

Comparison With Prior Work

We compare our work with studies on mobile apps for depression in terms of (1) application of EMA, (2) lexicon development, and (3) construct validation.

First, PiaP, as it employs EMA, does its evaluation with a time stamp upon the exact occurrence of the symptoms using text analysis. Chung et al [62] designed a mobile app that recorded daily self-reported ratings for the Korean version of the Center for Epidemiologic Studies Depression Scale–Revised (K-CESD-R). Although the K-CESD-R Mobile app was

completed by their 20 participants every day for 2 weeks to avoid recall bias, it still did not employ EMA real-time measurement unlike PiaP.

Second, PiaP considered the cultural expression of depression in text analysis in the creation of its English-Tagalog lexicon. This includes the mixed usage of Tagalog and English (Taglish), textolog (shortening of words), emoticons, and emojis, thus allowing for the recognition of "possible cultural variations in the expression of depressive symptoms via electronic data" [63,64] and providing a more nuanced screening. Compared with BinDhim et al [65], although they proved the feasibility of using a mobile app for depression screening by utilizing an app that was an electronic version of the Patient Health Questionnaire (PHQ)-9, they did not use text analysis.

Third, PiaP applied congruent construct validation to determine whether its construct of *depressive symptoms* corresponds to the depression construct of established psychological measures for depression. In Chung et al [62] and BinDhim et al [65] studies, each used only 1 test—K-CESD-R and PHQ-9, respectively as a basis for the electronic (mobile app) version. In the case of PiaP, aside from using CES-D Scale to determine construct validation of the PiaP lexicon, the researchers also used BDI-II, considered to be the gold standard in depression identification [66].

Conclusions

A major point to consider from this study is that the language used in contemporary avenues (such as social media communication and mobile technology) serves as a channel for expressing depression-associated emotions while avoiding stigmatization, thereby making lexical data analysis an added dimension to depression-screening. Language—the use or choice of words—can express most depression symptoms that are better expressed in verbal behavior, specifically those that are more cognitive in nature. With social media and other forms of communication being incorporated in mobile phones, it becomes easier to express oneself for individuals who may be experiencing depression, as they prefer to spend more time online rather than have face-to-face interactions.

The study also alludes to the value of combining current technology with mental assessment. Mobile technology and, consequently, EMA should be maximized for a timely identification, screening, monitoring, and follow-up of individuals with depression and other mental health issues.

As an mHealth app for depression screening, PiaP provides several advantages. First, PiaP has proven both its internal [19] and external validities, thus satisfying the increasing need for the scientific testing of mHealth apps. With its reliance on EMA, PiaP provides prompt information regarding the user's psychological state and eliminates or reduces errors and biases associated with interviews and self-reports of traditional mental health screening approaches, specifically in depression. Finally, PiaP's lexical analysis of electronic data yields a layer of refinement to depression identification. With this leverage, PiaP can be used as an accessible and novel supplement and technological support to traditional approaches in depression screening and monitoring.

Acknowledgments

The authors would like to thank the following: (from Germany) Dr Jó Ágila Bitsch (exceet Secure Solutions) for codeveloping PiaP; Mr Tim Ix, Mr Paul Smith, and Dr Sarah Winter for their work on the earlier PiaP prototypes; Mr Eugen Seljutin and Mr Marko Jovanović for the additional technical support; and (from the Philippines) Dr Portia Lynn Quetulio-See (University of Santo Tomas) for research consultation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Psychologist in a Pocket (PiaP) research version opening message.
[PDF File (Adobe PDF File)142 KB - mhealth_v7i9e12051_app1.pdf]

Multimedia Appendix 2
Psychologist in a Pocket (PiaP) set-up screen.
[PDF File (Adobe PDF File)114 KB - mhealth_v7i9e12051_app2.pdf]

Multimedia Appendix 3
Presentation during the PAP 55th Annual Convention (20-22 Sept 2018, Manila, Phil).
[PDF File (Adobe PDF File)58 KB - mhealth_v7i9e12051_app3.pdf]

References

1. Research2Guidance. 2016. mHealth Economics 2016 – Current Status and Trends of the mHealth App Market URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/ [accessed 2018-08-17] [WebCite Cache ID 71kiZf3qX]

2. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
3. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mhealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
4. Yasini M, Marchand G. Toward a use case based classification of mobile health applications. Stud Health Technol Inform 2015;210:175-179. [doi: 10.3233/978-1-61499-512-8-175] [Medline: 25991125]
5. Torous JB, Chan SR, Yellowlees PM, Boland R. To use or not? Evaluating ASPECTS of smartphone apps and mobile technology for clinical care in psychiatry. J Clin Psychiatry 2016 Jun;77(6):e734-e738. [doi: 10.4088/JCP.15com10619] [Medline: 27136691]
6. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
7. Torous JB, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interv 2015 May;2(2):169-173 [FREE Full text] [doi: 10.1016/j.invent.2015.03.002]
8. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015 Nov;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
9. 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]
10. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010 Apr 30;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
11. 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]
12. Eonta AM, Christon LM, Hourigan SE, Ravindran N, Vrana SR, Southam-Gerow MA. Using everyday technology to enhance evidence-based treatments. Prof Psychol: Res Pract 2011 Dec;42(6):513-520 [FREE Full text] [doi: 10.1037/a0025825]
13. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
14. 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]
15. Food and Drug Administration. 2016. Examples of Mobile Apps For Which the FDA Will Exercise Enforcement Discretion URL: https://www.fda.gov/medicaldevices/digitalhealth/mobilemedicalapplications/ucm368744.htm [accessed 2018-08-18] [WebCite Cache ID 71lahMJWi]
16. Armey MF, Schatten HT, Haradhvala N, Miller IW. Ecological momentary assessment (EMA) of depression-related phenomena. Curr Opin Psychol 2015 Aug 1;4:21-25 [FREE Full text] [doi: 10.1016/j.copsyc.2015.01.002] [Medline: 25664334]
17. PiaP: Psychologist in a Pocket. URL: https://piap.mobi/ [accessed 2019-08-22]
18. Bitsch J, Ramos R, Ix T, Ferrer-Cheng PG, Wehrle K. Psychologist in a pocket: towards depression screening on mobile phones. Stud Health Technol Inform 2015;211:153-159. [doi: 10.3233/978-1-61499-516-6-153] [Medline: 25980862]
19. Cheng PG, Ramos RM, Bitsch JA, Jonas SM, Ix T, See PL, et al. Psychologist in a pocket: lexicon development and content validation of a mobile-based app for depression screening. JMIR Mhealth Uhealth 2016 Jul 20;4(3):e88 [FREE Full text] [doi: 10.2196/mhealth.5284] [Medline: 27439444]
20. Boyle GJ, Matthews G, Saklofske DH. The Sage Handbook of Personality Theory and Assessment: Personality Measurement and Testing. Volume 2. Thousand Oaks, CA: Sage Publicationa; 2008.
21. Millon T, Millon C, Davis R, Grossman S. MCMI-III Manual: Millon Clinical Multiaxial Inventory-III. Fourth Edition. Minneapolis, MN: Pearson Education; 2009.
22. Meagher SE, Grossman SD, Millon T. Treatment planning and outcome assessment in adults: the millon clinical multiaxial inventory (MCMI-III). In: Maruish MW, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. Third Edition. Volume 3. New York: Erlbaum Publishing; 2004:479-508.
23. Westen D, Rosenthal R. Quantifying construct validity: two simple measures. J Pers Soc Psychol 2003 Mar;84(3):608-618. [doi: 10.1037/0022-3514.84.3.608] [Medline: 12635920]
24. Colman AM. A Dictionary of Psychology. Third Edition. Oxford: Oxford University Press; 2014.
25. Cohen RJ, Swerdlik ME. Psychological Testing and Assessment: An Introduction to Tests and Measurement. Seventh Edition. New York: McGraw-Hill; 2009.
26. Kaplan RM, Saccuzzo DP. Psychological Testing: Principles, Applications, and Issues. Eighth Edition. Belmont, CA: Wadsworth Publishing; 2012.

27. Ramos R, Cheng P, de Castro F. Attitudes toward mhealth: a look at general attitudinal indices among selected Filipino undergraduates. In: Mohan B, editor. Construction of Social Psychology: Advances in Psychology and Psychological Trends. Lisboa, Portugal: InScience Press; 2015:186-202.
28. Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, McCormick RA. Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. J Abnorm Psychol 1995 Feb;104(1):3-14. [doi: 10.1037//0021-843X.104.1.3] [Medline: 7897050]
29. Wichers M, Simons CJ, Kramer IM, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011 Oct;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
30. Moksnes UK, Løhre A, Lillefjell M, Byrne DG, Haugan G. The association between school stress, life satisfaction and depressive symptoms in adolescents: life satisfaction as a potential mediator. Soc Indic Res 2014;125(1):339-357 [FREE Full text] [doi: 10.1007/s11205-014-0842-0]
31. Koivumaa-Honkanen H, Kaprio J, Honkanen R, Viinamäki H, Koskenvuo M. Life satisfaction and depression in a 15-year follow-up of healthy adults. Soc Psychiatry Psychiatr Epidemiol 2004 Dec;39(12):994-999. [doi: 10.1007/s00127-004-0833-6] [Medline: 15583908]
32. Pavot W, Diener E. The satisfaction with life scale and the emerging construct of life satisfaction. J Posit Psychol 2008 Apr;3(2):137-152. [doi: 10.1080/17439760701756946]
33. Watson D, Clark LA, Carey G. Positive and negative affectivity and their relation to anxiety and depressive disorders. J Abnorm Psychol 1988 Aug;97(3):346-353. [doi: 10.1037/0021-843X.97.3.346] [Medline: 3192830]
34. Geschwind N, Nicolson NA, Peeters F, van Os J, Barge-Schaapveld D, Wichers M. Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. Eur Neuropsychopharmacol 2011 Mar;21(3):241-247 [FREE Full text] [doi: 10.1016/j.euroneuro.2010.11.004] [Medline: 21146375]
35. Diener E. Subjective well-being. Psychol Bull 1984 May;95(3):542-575. [doi: 10.1037/0033-2909.95.3.542] [Medline: 6399758]
36. Shin DC, Johnson DM. Avowed happiness as an overall assessment of the quality of life. Soc Indic Res 1978;5(1-4):475-492. [doi: 10.1007/BF00352944]
37. Headey B, Kelley J, Wearing A. Dimensions of mental health: life satisfaction, positive affect, anxiety and depression. Soc Indic Res 1993;29(1):63-82. [doi: 10.1007/BF01136197]
38. Mukaka MM. A guide to appropriate use of Correlation coefficient in medical research. Malawi Med J Sep; ? 2012;24(3):71. [Medline: 23638278]
39. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. Fifth Edition. Boston: Houghton Mifflin; 2003.
40. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
41. Cusin C, Yang H, Yeung A, Fava M. Rating scales for depression. In: Baer L, Blais MA, editors. Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health: Current Clinical Psychiatry. New York: Humana Press; 2009:7-35.
42. Peyton M, Critchley CR. The development of the experiences of low mood and depression questionnaire. N Am J Psychol 2005;7(1):35-42 [FREE Full text]
43. Lee RB, Maria MS, Estanislao S, Rodriguez C. Factors associated with depressive symptoms among Filipino university students. PLoS One 2013;8(11):e79825 [FREE Full text] [doi: 10.1371/journal.pone.0079825] [Medline: 24223198]
44. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: Manual. San Antonio, TX: Psychological Corporation; 1996.
45. Shean G, Baldwin G. Sensitivity and specificity of depression questionnaires in a college-age sample. J Genet Psychol 2008 Sep;169(3):281-288. [doi: 10.3200/GNTP.169.3.281-292] [Medline: 18788328]
46. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psych Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
47. Carleton RN, Thibodeau MA, Teale MJ, Welch PG, Abrams MP, Robinson T, et al. The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. PLoS One 2013;8(3):e58067 [FREE Full text] [doi: 10.1371/journal.pone.0058067] [Medline: 23469262]
48. Bradburn NM. The Structure of Psychological Well-Being. Chicago: Aldine Publishing; 1969.
49. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
50. MedCalc Statistical Software. 2016. URL: https://www.medcalc.org; [accessed 2019-02-04]
51. van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
52. Eremsoy CE, Çelimli S, Gençöz T. Students under academic stress in a Turkish university: variables associated with symptoms of depression and anxiety. Curr Psychol 2005;24(2):123-133 [FREE Full text] [doi: 10.1007/s12144-005-1011-z]

53. Liu Y, Lu Z. Chinese high school students' academic stress and depressive symptoms: gender and school climate as moderators. Stress Health 2012 Oct;28(4):340-346. [doi: 10.1002/smi.2418] [Medline: 22190389]
54. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 8;29(1):24-54. [doi: 10.1177/0261927X09351676]
55. Beck JS. Cognitive Behavior Therapy: Basics and Beyond. Second Edition. New York, NY: The Guilford Press; 2011.
56. Diener E, Larsen RJ, Levine S, Emmons RA. Intensity and frequency: dimensions underlying positive and negative affect. J Pers Soc Psychol 1985 May;48(5):1253-1265. [doi: 10.1037/0022-3514.48.5.1253] [Medline: 3998989]
57. Yavuzer Y, Karatas Z. Investigating the relationship between depression, negative automatic thoughts, life satisfaction and symptom interpretation in Turkish young adults. In: Breznoscakova D, editor. Depression. UK: IntechOpen; 2017:71-89.
58. 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]
59. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
60. DiMatteo M, Haskard KB, Williams SL. Health beliefs, disease severity, and patient adherence: a meta-analysis. Med Care 2007 Jun;45(6):521-528. [doi: 10.1097/MLR.0b013e318032937e] [Medline: 17515779]
61. Kanter JW, Busch AM, Weeks CE, Landes SJ. The nature of clinical depression: symptoms, syndromes, and behavior analysis. Behav Anal 2008;31(1):1-21 [FREE Full text] [doi: 10.1007/bf03392158] [Medline: 22478499]
62. Chung K, Jeon MJ, Park J, Lee S, Kim CO, Park JY. Development and evaluation of a mobile-optimized daily self-rating depression screening app: a preliminary study. PLoS One 2018;13(6):e0199118 [FREE Full text] [doi: 10.1371/journal.pone.0199118] [Medline: 29944663]
63. Chang MX, Jetten J, Cruwys T, Haslam C. Cultural identity and the expression of depression: a social identity perspective. J Commun Appl Soc Psychol 2016 Oct 20;27(1):16-34. [doi: 10.1002/casp.2291]
64. Lovey K, Torrez J, Fine A, Moriarty G, Coppersmith G. Cross-Cultural Differences in Language Markers of Depression Online. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology. 2018 Presented at: CLPsych'18; June 5, 2018; New Orleans, LA p. 78-87 URL: http://www.aclweb.org/anthology/W18-0608 [doi: 10.18653/v1/W18-0608]
65. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
66. Kneipp SM, Kairalla JA, Stacciarini J, Pereira D. The Beck Depression Inventory II factor structure among low-income women. Nurs Res 2009;58(6):400-409. [doi: 10.1097/NNR.0b013e3181bee5aa] [Medline: 19918151]

Abbreviations

ABS: Affect Balance Scale
B-A: Bland-Altman
BDI-II: Beck Depression Index-II
CES-D Scale: Center for Epidemiological Studies–Depression Scale
CIL: confidence interval limit
DSM-5: Diagnostic and Statistical Manual of Mental Disorders–5
EMA: ecological momentary assessment
ES: effect size
K-CESD-R: Korean version of the Center for Epidemiologic Studies Depression Scale–Revised
mHealth: mobile health
PHQ: Patient Health Questionnaire
PiaP: Psychologist in a Pocket
PPMC: Pearson product-moment correlation
PTS: PiaP total score
SWLS: Satisfaction With Life Scale

Edited by G Eysenbach; submitted 30.08.18; peer-reviewed by N Shen, LA Lee; comments to author 31.01.19; revised version received 21.03.19; accepted 28.07.19; published 16.09.19.

Please cite as:
Ramos RM, Cheng PGF, Jonas SM
Validation of an mHealth App for Depression Screening and Monitoring (Psychologist in a Pocket): Correlational Study and Concurrence Analysis
JMIR Mhealth Uhealth 2019;7(9):e12051
URL: https://mhealth.jmir.org/2019/9/e12051/
doi:10.2196/12051
PMID:31538946

Original Paper

Using Video Feedback Through Smartphone Instant Messaging in Fundamental Nursing Skills Teaching: Observational Study

Xiaoxian Yang[1], MA; Ri-Hua Xie[2,3*], PhD; Si Chen[1], BA; Wei Yu[1], MA; Yan Liao[4], BEng; Daniel Krewski[5,6], PhD; Shi Wu Wen[6,7,8*], PhD

[1]School of Nursing, Wuxi Taihu University, Wuxi, China

[2]Department of Nursing, Nanhai Hospital, Southern Medical University, Foshan, China

[3]General Practice Center, Nanhai Hospital, Southern Medical University, Foshan, China

[4]Ottawa Hospital Research Institute, Ottawa, ON, Canada

[5]McLaughlin Centre for Population Health Risk Assessment, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[6]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[7]Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada

[8]Department of Obstetrics & Gynecology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Ri-Hua Xie, PhD
Department of Nursing
Nanhai Hospital
Southern Medical University
28 Liguan Rd, Lishui
Foshan, 528244
China
Phone: 86 189 2869 7126
Email: xierihua928@hotmail.com

Abstract

Background: Video feedback has been shown to be an effective teaching tool that can improve student learning when having them view their own performance. However, the literature on the effect of integrating smartphones with video feedback in fundamental nursing skills teaching is sparse.

Objective: This study aimed to explore the potential effects of video feedback through smartphone-based instant messaging on teaching undergraduate nursing students fundamental nursing skills.

Methods: We conducted a study on teaching fundamental nursing skills to 6 classes of second-year undergraduate nursing students. In 2 classes (the intervention group), the instructor elected to use smartphone-based video feedback to facilitate teaching; instructors in the other 4 classes (the control group) elected to use routine methods of teaching without video feedback. Scores from the final examination, in-class assignments, and the General Self-Efficacy Scale questionnaire were collected and compared between the two groups. Multiple linear regression analysis was performed to estimate the independent effect of video feedback after adjusting for gender, age, and prior experience in the use of WeChat/QQ in learning applications. An ad hoc questionnaire was used for student evaluation of the novel smartphone-based video feedback teaching method.

Results: A total of 195 nursing students (65 in the video feedback group and 130 in the control group) completed the study and were included in the final analysis. Mean and standard deviation of scores on the final examination, bed making, aseptic procedure, vital signs measurement, and oxygen therapy were 91.29 (SD 2.36), 90.52 (SD 3.18), 93.23 (SD 3.16), 91.65 (SD 4.21), and 92.06 (SD 3.58), respectively, in the video feedback group and 89.99 (SD 3.12), 81.71 (SD 8.63), 87.12 (SD 5.50), 87.45 (SD 8.00), and 90.37 (SD 6.36), respectively, in the control group (differences were statistically significant). The mean and standard deviation of scores for assignments in catheterization and enema and General Self-Efficacy Scale were 89.69 (SD 3.22), 91.14 (SD 3.15), and 24.52 (SD 5.35), respectively, in the video feedback group and 88.82 (SD 7.48), 90.79 (SD 6.08), and 24.50 (SD 6.16), respectively, in the control group (differences were not statistically significant). The majority (over 98%) of nursing students were satisfied with this smartphone-based video feedback teaching method.

Conclusions: Video feedback through smartphone-based instant messaging may be an effective way to improve nursing students' academic performance and professional skills.

(JMIR Mhealth Uhealth 2019;7(9):e15386) doi:10.2196/15386

KEYWORDS

video feedback; smartphone; mobile phone; student nurses; fundamental nursing skill; teaching

Introduction

Video feedback has been demonstrated to be an effective teaching tool that can improve student skills by having them view their own performance on video [1]. Students can identify what they did well, what they did not do so well, and what they could improve through video feedback in nursing skills training [2]. A meta-analysis showed that video feedback has a positive effect on learning [1]. However, video feedback with standard equipment for large classroom teaching is not convenient and lacks flexibility. It is therefore important to find a way to improve the process for video feedback in medical education.

Smartphones are very popular with the youth and could be exploited to improve learning as a consequence of personal behavior changes [3,4]. In recent years, China has developed popular apps such as WeChat and QQ, which support instant text messaging and voice and video calling via smartphones. WeChat and QQ are the most popular personal communication tools used by university and college students in China, making them attractive options for implementing enhanced teaching methods in medicine and nursing [5,6]. The penetration rate of both smartphone and WeChat/QQ use is almost 100% among university and college students, providing a convenient basis for integrating the video feedback in nursing skills training [7,8].

Attempts have been made to use smartphones to achieve positive results for a range of medical and nursing education issues, including coordination supporting among groups [7], theory and practice integrating [9], student participation/engagement [10], and communication skill enhancement [7]. Nursing skills practice could be recorded in video format by students with their smartphones and then sent to instructors via instant messaging. The instructors, in turn, could provide comments and suggestions on student performance, providing rapid content-related feedback. As the literature on the potential benefits of video feedback on the teaching of fundamental nursing skills is sparse, our study was conducted to evaluate the effects of video feedback through smartphone-based instant messaging on teaching fundamental nursing skills.

Methods

Study Participants

This study was carried out in Wuxi Taihu University School of Nursing between October 1 and November 30, 2018. There were 6 classes of second-year undergraduate nursing students who were taking the fundamental nursing skills course during that semester. There were 3 instructors, each of whom taught 2 classes with no overlap/exchanges/substitutions throughout the entire semester. The instructors were qualified nursing educators with the same seniority. At the beginning of the semester, an instructor for 2 of the 6 classes elected to use smartphone-based video feedback to facilitate her teaching. One author of the research team approached the instructors with a request to undertake a study to evaluate the effect of this new method of teaching. After the instructors agreed to participate in the study, an approval from the institutional review board of Taihu University was obtained. Students from the 2 classes whose instructor planned to use the smartphone-based video feedback comprised the intervention group (video feedback group), while students from other 4 classes whose instructors planned to use routine teaching methods (no video feedback) formed the control group. All students provided written informed consent to participate in the study. Fundamental nursing skills taught in this study were performed on manikin simulators and included bed making, aseptic procedures, vital signs measurement, oxygen therapy, catheterization, and enema administration. Students from both groups were asked to practice and complete one assignment for each skill after class. In the video feedback group, the instructor first explained the basic material and demonstrated the basic procedures in the classroom. Afterward, the nursing students worked in groups of 3 or 4 to record videos of each other on their smartphones. Each video lasted 5 to 15 minutes and took more than 10 minutes to upload and download. The students then sent the videos to the instructor via instant messaging for evaluation and feedback on their nursing practices (the instructor spent 5 to 15 minutes going through each video).

Outcome Measures

Scores on the final examination and on the 6 nursing skills assignments (bed making, aseptic procedure, vital signs measurement, oxygen therapy, catheterization, and enema) were the main outcomes of interest. Teaching and evaluation at Wuxi Taihu University School of Nursing were performed by different faculty members (faculty members involved in teaching cannot evaluate their own students).

The General Self-Efficacy (GSE) Scale [11], which is designed to assess optimistic self-beliefs related to coping with a variety of demands in life, was a secondary outcome. The GSE Scale comprises 10 questions, each of which is scored on a 4-point Likert scale (1=not at all true, 2=hardly true, 3=moderately true, 4=exactly true). Total score ranges between 10 and 40. This scale was originally developed by Schwarzer and Jerusalem [11] and has been used and evaluated in several populations and cultures. The Chinese version of the GSE Scale has been used in college and university students in China [12]. Its Cronbach alpha coefficient is 0.87, the retest reliability is 0.83, the half-fold reliability is 0.90, and the validity ranges from 0.60 to 0.77 [13].

Finally, an ad hoc questionnaire was developed to measure the students' own evaluations of the video feedback teaching

method in the video feedback group. The questionnaire has 6 items with 5 response categories expressing the degree of agreement with each item.

Data Analysis

We first compared the distribution of baseline characteristics between the intervention and control groups. We then compared the scores on the final examination, the 6 nursing skills assignments, and GSE scores between the two groups. Multiple linear regression analysis was conducted to estimate the independent effect of video feedback on student performance after adjusting for gender, age, and prior experience in the use of WeChat/QQ in learning applications. Two-sided tests were used in all comparisons between the two study groups. Finally, student evaluations of the video feedback teaching method were presented. All analyses were performed using SAS v9.4 (SAS Institute Inc).

Results

Baseline Characteristics of Study Participants

A total of 195 nursing students (65 in the video feedback group and 130 in the control group) completed the study and were included in the final analysis. Table 1 shows the distribution of baseline characteristics of the two groups. There were no differences in gender, age, or prior experience in the use of WeChat/QQ in learning applications between the two groups.

Comparison of Scores in the Six Nursing Skills Between the Intervention and Control Groups

Table 2 presents the mean and SD of the final examination scores and aggregate and individual scores for the 6 indicators of nursing skills. Scores on the final examination, bed making, aseptic procedure, vital signs measurement, and oxygen therapy were significantly higher in the video feedback group than in the control group. No significant differences in scores on catheterization and enema between the two study groups were observed.

Comparison of Scores on the General Self-Efficacy Scale Between the Intervention and Control Groups

Table 3 compares the mean and standard deviation of the total GSE score and the 10 individual items comprising the GSE Scale between the two study groups. A total of 148 students (48 students in video feedback group and 100 students in control group) provided answers to the questionnaire. Mean and SD of the total scores for GSE Scale were 24.52 (SD 5.35) in the video feedback group and 24.50 (SD 6.16) in control group (differences in the total score and individual item scores were not statistically significant).

Table 1. Comparison of baseline characteristics of intervention and control groups.

Characteristics	Video feedback group (n=65)	Control group (n=130)	*P* value
Gender, n (%)			.33
Male	5 (7.7)	16 (12.3)	
Female	60 (92.3)	114 (87.7)	
Age (years), mean (SD)	19.65 (0.82)	19.54 (0.75)	.34
Prior experience in the use of WeChat/QQ in learning applications, n (%)			.18
Yes	22 (33.9)	57 (43.9)	
No	43 (66.1)	73 (56.1)	

Table 2. Comparison of scores in the 6 nursing skills between the intervention and control group.

Test scores	Video feedback group (n=65), mean (SD)	Control group (n=130), mean (SD)	Crude mean difference (95% CI)	Adjusted mean difference[a] (95% CI)
Skill 1: bedmaking	90.52 (3.18)	81.71 (8.63)	8.82 (6.63, 11.00)	8.82 (6.67, 10.97)
Skill 2: aseptic procedure	93.23 (3.16)	87.12 (5.50)	6.12 (4.66, 7.57)	6.12 (4.66, 7.57)
Skill 3: vital signs measurement	91.65 (4.21)	87.45 (8.00)	4.19 (2.10, 6.28)	4.19 (2.12, 6.27)
Skill 4: oxygen therapy	92.06 (3.58)	90.37 (6.36)	1.69 (0.02, 3.37)	1.69 (0.01, 3.38)
Skill 5: catherization	89.69 (3.22)	88.82 (7.48)	0.87 (–1.04, 2.78)	0.87 (–1.02, 2.76)
Skill 6: enema	91.14 (3.15)	90.79 (6.08)	0.35 (–1.24, 1.93)	0.35 (–1.21, 1.90)
Average: 6 skills	91.38 (2.10)	87.71 (4.43)	3.67 (2.53, 4.82)	3.67 (2.54, 4.81)
Average: final examination	91.29 (2.36)	89.99 (3.12)	1.30 (0.43, 2.17)	1.30 (0.44, 2.16)

[a]Adjusted for gender, age, and prior experience in the use of WeChat/QQ in learning applications.

Table 3. Comparison of scores of the General Self-Efficacy (GSE) Scale between the intervention and control groups.

GSE Scale scores	Video feedback group (n=48), mean (SD)	Control group (n=100), mean (SD)	Crude mean difference (95% CI)	Adjusted mean difference[a] (95% CI)
Item 1: I can always manage to solve difficult problems if I try hard enough.	2.78 (0.72)	2.82 (0.74)	–0.05 (–0.30, 0.21)	–0.05 (–0.31, 0.21)
Item 2: If someone opposes me, I can find the means and ways to get what I want.	2.50 (0.80)	2.47 (0.67)	0.03 (–0.22, 0.28)	0.03 (–0.22, 0.28)
Item 3: It is easy for me to stick to my aims and accomplish my goals.	2.00 (0.80)	2.15 (0.74)	–0.15 (–0.41, 0.11)	–0.15 (–0.41, 0.11)
Item 4: I am confident that I could deal efficiently with unexpected events.	2.21 (0.74)	2.29 (0.71)	–0.08 (–0.33, 0.17)	–0.08 (–0.34, 0.17)
Item 5: Thanks to my resourcefulness, I know how to handle unforeseen situations.	2.23 (0.78)	2.23 (0.71)	–0.00 (–0.25, 0.25)	–0.00 (–0.26, 0.25)
Item 6: I can solve most problems if I invest the necessary effort.	2.79 (0.71)	2.71 (0.69)	0.08 (–0.32, 0.16)	0.08(–0.32, 0.16)
Item 7: I can remain calm when facing difficulties because I can rely on my coping abilities.	2.63 (0.76)	2.65 (0.70)	–0.02 (–0.28, 0.23)	–0.02 (–0.27, 0.22)
Item 8: When I am confronted with a problem, I can usually find several solutions.	2.44 (0.74)	2.43 (0.69)	0.01 (–0.24, 0.25)	0.01 (–0.24, 0.25)
Item 9: If I am in trouble, I can usually think of a solution.	2.69 (0.72)	2.55 (0.63)	0.14 (–0.09, 0.37)	0.14 (–0.09, 0.36)
Item 10: I can usually handle whatever comes my way.	2.25 (0.79)	2.22 (0.79)	0.03 (–0.24, 0.30)	0.03 (–0.24, 0.30)
Total score	24.52 (5.35)	24.50 (6.16)	0.02 (–1.97, 1.93)	0.02 (–1.98, 1.94)

[a]Adjusted for gender, age, and prior experience in using WeChat/QQ in learning applications.

Student Evaluations of Video Feedback

Table 4 presents student evaluations of the smartphone-based video feedback teaching method. Of the students who provided answers to the questionnaire, 98% (54/55) of nursing students were satisfied with the smartphone-based video feedback. Many of the nursing students strongly agreed that the video feedback teaching method can improve skill proficiency (32/55, 58%), practice passion (28/55, 51%), learning interest (27/55, 49%), learning effectiveness (31/55, 56%) and autonomous learning ability (32/55, 58%).

Table 4. Student evaluations of video feedback in the video group (n=55).

Item	Strongly agree, n (%)	Agree, n (%)	Uncertain, n (%)	Disagree, n (%)	Strongly disagree, n (%)
Improved skill proficiency	32 (58)	21 (38)	2 (4)	0 (0)	0 (0)
Improved practice passion	28 (51)	25 (45)	2 (4)	0 (0)	0 (0)
Improved learning interest	27 (49)	24 (44)	4 (7)	0 (0)	0 (0)
Improved learning effectiveness	31 (56)	24 (44)	0 (0)	0 (0)	0 (0)
Improved autonomous learning ability	32 (58)	20 (36)	3 (5)	0 (0)	0 (0)
Satisfied with the smartphone-based video feedback method	35 (63)	19 (34)	1 (1)	0 (0)	0 (0)

Discussion

Principal Findings

Our study found that video feedback through smartphone-based instant messaging may have the potential to improve the performance of nursing students in fundamental nursing skills, especially with respect to skills related to bed making, aseptic procedure, vital signs measurement, and oxygen therapy. Overall, most nursing students were satisfied with the smartphone-based video feedback teaching method. Although the study failed to demonstrate an improvement in overall self-efficacy, nursing students perceived that their interests and autonomous learning abilities had been improved.

Strengths

To our knowledge, this is the first study incorporating smartphone instant messaging with video feedback in fundamental nursing skills teaching. Although the effect of video feedback in medical education has been well established, the need for standard equipment in regular video feedback makes this teaching method not convenient to some extent. Through smartphone-based video messaging, students were able to record the videos at a time and place that was convenient to them and send their videos to the instructor for timely and precise feedback. In Wuxi Taihu University School of Nursing, faculty members not involved in the teaching of the particular course acted as the evaluators of student performances, thereby

avoiding bias in the assessment. In contrast to Western countries, Chinese universities admit students as high school graduates according to their performance on the National University/College Entrance Examination. Top-ranked universities have priority to admit students with higher scores on the examination and superior academic performance in high school. Leading national universities recruit top students from all provinces across the country, while local universities like Wuxi Taihu University mainly recruit students locally. Once admitted, students are assigned to different classes by the university administration in a somewhat random fashion. In our study, students from the intervention and control groups were similar in gender, age, and previous life experience, so any differences between the groups should be attributable to the intervention and not to inherent differences between the two groups. The most important previous life experience relevant to this study is the prior experience using WeChat/QQ messaging in learning, which was not different between the two groups (Table 1). We used multiple regression analysis in the comparison of outcomes between the 2 study groups to adjust for age, gender, and previous life experience, ensuring no residual confounding in the comparison. The skills were performed on mannikins, so there were no ethical concerns related to videography.

Limitations

We acknowledge limitations of this study. First, whether to use the smartphone-based video feedback was a choice by the course instructor. Although all 3 instructors are qualified nursing educators with the same seniority, the instructor who elected to use smartphone-based video feedback to facilitate her teaching may be more motivated and this may have resulted in better quality in her teaching. Because this was not a randomized controlled trial, we cannot be sure about this source of bias. Second, the answers to some of the GSE Scale questions may be somewhat inaccurate and imprecise because some of the participants completed the questionnaire several weeks after the course was over. Thus, recall bias may exist. Third, the questionnaire for student evaluation of the smartphone-based video feedback was developed on an ad hoc basis without formal validation or reliability assessment. Fourth, although faculty members who evaluated student performance did not participate in teaching the course, they were from the same school and knew who the course instructor was. As a result, the skills evaluation could not be considered entirely blind. Fifth, there may be a chance that some students in the intervention group may not have actually done the video or the instructor may not have actually sent feedback to some students. Either way, actual effect of video feedback may have been diluted. Unfortunately, we did not collect these data and could not assess the impact of the quality of video feedback on the observed effect. Sixth, the time needed for video uploading and downloading depended on the network speed and could have been frustrating for students and instructors alike if the network was slow.

Implications

This study explored a method of improving fundamental nursing skills teaching for undergraduate nursing students. It integrated smartphone technology, a mobile app, and video feedback in facilitating teaching. Smartphone-based feedback could offer a novel, flexible study method, and the feedback could allow participants to know whether they are performing well or not [14]. Students could identify problems or errors in their performance while reviewing their video and then repeat the procedures in the correct manner. This learning experience is conducive to deepening student understanding of clinical skills practice, empowering students to standardize their own skills practice and explore the limits of their own skills and abilities. Problems or errors encountered in video practice could be sent to the instructor promptly for rapid feedback and correction. In the future, application of these skills could be further improved and the tasks and procedures standardized to optimize operational performance in clinical work. This method could also increase the frequency of practice after class, which is a key point of improving nursing skills. Thus, smartphone-based instant messaging video feedback could improve fundamental nursing skills for nursing students, consistent with a previous study in Korea [15]. Of the 6 skills assessed, smartphone-based video feedback had a stronger effect on skills in bed making, aseptic procedure, vital signs measurement, and oxygen therapy. The reasons for the lack of significant improvement in catheterization and enema administration by video feedback are unclear. Catheterization and enema administration were more complicated than the other procedures, and they were the last two skills to be learned in the semester. We speculate that students may have limited time to prepare video and practice in the end of semester.

Self-efficacy is related to one's beliefs as to whether or not they are capable of completing a certain task [16]. Studies have demonstrated that general self-efficacy is positively correlated with self-learning ability [17], indicating that improving self-efficacy could encourage nursing students to learn by themselves [18]. Using a smartphone [19] or a personal digital assistant [20] in nursing education could improve self-efficacy because both have the ability to meet their unique needs and improve confidence while learning. However, our study failed to find an improvement in self-efficacy. Some of the participants completed the GSE Scale questionnaire several weeks after the course was over; therefore, they may have not answered the questions accurately and precisely. It is also possible that one curriculum may not be sufficient to improve self-efficacy. As shown in Table 4, the majority of students perceived that the smartphone-based video feedback teaching method could improve their interest, effectiveness, and capacity of autonomous learning, which may lead to improvement in their performance on the final examination and assignments of skills, despite the lack of a significant improvement in self-efficacy.

Most students were satisfied with the smartphone-based video feedback teaching method, a finding consistent with a previous study [21]. A systematic review of the use of mobile technology in undergraduate education also showed that nursing faculty members have become more interested in incorporating such technologies into their teaching strategies [7]. In our study, the majority of students strongly agreed that video feedback could improve the proficiency of nursing skills and autonomous learning. The nursing students may become more motivated to learn when these technologies are incorporated in education

[14]. The attitudes of nursing students toward an instant message–based video feedback teaching paradigm shows that this teaching method is both feasible and acceptable. It would be informative to undertake a formal qualitative evaluation of both instructors and students to further explore attitudes toward and acceptance of this new method of teaching. It should be pointed out that although it is easy for instructors to go through the videos and video feedback may improve quality of teaching, its use may increase instructor workload with respect to reviewing videos and sending feedback. On the other hand, it may reduce the workload associated with other aspects of teaching by reducing the need for face-to-face consultations. These issues need to be further explored and considered by teachers and school administrators before smartphone-based video feedback is widely implemented in nursing teaching curricula.

Conclusions

Our study suggests that the use of video feedback through smartphone-based instant messaging may be an effective way to improve nursing students' overall performance and skills. This novel teaching modality makes use of relatively inexpensive smartphone technology, which is now almost universally available and familiar to the millennials who will become tomorrow's health professionals. Extending the use of smartphone-based video feedback teaching techniques more broadly across multi-year academic curricula and other areas of health sciences could lead to even better results than those observed in this limited study, including not only increased performance but increased self-efficacy.

Acknowledgments

The authors thank the participating students and teachers in Wuxi Taihu University School of Nursing for their contributions to this study. DK is the Natural Sciences and Engineering Research Council of Canada Chair in Risk Science at the University of Ottawa. This work has been partly supported by the Canadian Institute of Health Research (FDN-148438), Southern Medical University (LC2019ZD019), Nanhai Hospital of the Southern Medical University (YNKT201802), and Wuxi Taihu University (17WUNS002).

Authors' Contributions

RHX and XY conceptualized and designed the study. XY, SC, and WY made major contributions to the acquisition of the data, and XY and YL made major contributions to the analysis and interpretation of the data. XY drafted the paper; RHX, DK, and SWW critically reviewed and revised the paper; and all authors approved the final version of the manuscript. Each author certified that they had participated sufficiently in the work to believe in its overall validity and take public responsibility for appropriate portions of its content.

Conflicts of Interest

None declared.

References

1. Fukkink RG, Trienekens N, Kramer LJC. Video feedback in education and training: putting learning in the picture. Educ Psychol Rev 2010 Oct 16;23(1):45-63. [doi: 10.1007/s10648-010-9144-5]
2. Bowden T, Rowlands A, Buckwell M, Abbott S. Web-based video and feedback in the teaching of cardiopulmonary resuscitation. Nurse Educ Today 2012 May;32(4):443-447. [doi: 10.1016/j.nedt.2011.04.003] [Medline: 21546137]
3. Hides L, Dingle G, Quinn C, Stoyanov SR, Zelenko O, Tjondronegoro D, et al. Efficacy and outcomes of a music-based emotion regulation mobile app in distressed young people: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11482 [FREE Full text] [doi: 10.2196/11482] [Medline: 30664457]
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. Shaban S, Elzubeir M, Al Houqani M. Online assessment standard setting for multiple choice questions. Int J Med Educ 2016 May 05;7:142-143 [FREE Full text] [doi: 10.5116/ijme.5708.e5c4] [Medline: 27149419]
6. Cui H, Li R, Fang Y, Horn B, Welsch RE. Heterogeneous characters modeling of instant message services users' online behavior. PLoS One 2018;13(5):e0195518 [FREE Full text] [doi: 10.1371/journal.pone.0195518] [Medline: 29734327]
7. Lee H, Min H, Oh S, Shim K. Mobile technology in undergraduate nursing education: a systematic review. Healthc Inform Res 2018 Apr;24(2):97-108 [FREE Full text] [doi: 10.4258/hir.2018.24.2.97] [Medline: 29770243]
8. Clavier T, Ramen J, Dureuil B, Veber B, Hanouz J, Dupont H, et al. Use of the smartphone app Whatsapp as an e-learning method for medical residents: multicenter controlled randomized trial. JMIR Mhealth Uhealth 2019 Apr 09;7(4):e12825 [FREE Full text] [doi: 10.2196/12825] [Medline: 30964435]
9. Willemse JJ. Undergraduate nurses reflections on Whatsapp use in improving primary health care education. Curationis 2015 Aug 13;38(2):1512 [FREE Full text] [doi: 10.4102/curationis.v38i2.1512] [Medline: 26304053]

10. Rambe P, Bere A. Using mobile instant messaging to leverage learner participation and transform pedagogy at a South African University of Technology. Br J Educ Technol 2013 Jun 04;44(4):544-561. [doi: 10.1111/bjet.12057]
11. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio. Windsor: NFER-Nelson; 1995:35-37.
12. Cheung S, Sun SYK. Assessment of optimistic self-beliefs: further validation of the Chinese version of the General Self-Efficacy Scale. Psychol Rep 2016 Sep;85(3_suppl):1221-1224. [doi: 10.2466/pr0.1999.85.3f.1221]
13. Zhang JXS. Measuring optimistic self-beliefs: a Chinese adaptation of the General Self-Efficacy Scale. Psychologia 1995;38(3):174-181. [doi: 10.1037/t00393-000]
14. Lee MK. Effects of mobile phone-based app learning compared to computer-based web learning on nursing students: pilot randomized controlled trial. Healthc Inform Res 2015;21(2):125. [doi: 10.4258/hir.2015.21.2.125]
15. Kang K. Effectiveness of video-record method on fundamental nursing skill education: focused on enema. J Korean Acad Fundam Nurs 1996 Dec;3(2):273-283.
16. Bandura A. Self-efficacy. In: Ramachandran VS, editor. Encyclopedia of Human Behavior Vol 4. New York: Academic Press; 1994:71-81.
17. Wolters CA. Understanding procrastination from a self-regulated learning perspective. J Educ Psychol 2003;95(1):179-187. [doi: 10.1037//0022-0663.95.1.179]
18. Yuan QL, Gao JL. The relationship between academic self-efficacy, achievement motivation and self-learning ability of undergraduate nursing students. J Nursing 2008;23(3).
19. Kim MS, Park JH, Park K. [Development and effectiveness of a drug dosage calculation training program using cognitive loading theory based on smartphone application]. J Korean Acad Nurs 2012 Oct;42(5):689-698. [doi: 10.4040/jkan.2012.42.5.689] [Medline: 23221658]
20. Goldsworthy S, Lawrence N, Goodman W. The use of personal digital assistants at the point of care in an undergraduate nursing program. Comput Inform Nurs 2006;24(3):138-143. [Medline: 16707943]
21. Williams MG, Dittmer A. Textbooks on tap: using electronic books housed in handheld devices in nursing clinical courses. Nurs Educ Perspect 2009;30(4):220-225. [Medline: 19753854]

Abbreviations

GSE Scale: General Self-Efficacy Scale

Edited by G Eysenbach; submitted 06.07.19; peer-reviewed by R Chan, A Sulley, N Ozcakar, F Kaliyadan, J Sanchez Viamonte; comments to author 27.07.19; revised version received 19.08.19; accepted 21.08.19; published 05.09.19.

Please cite as:
Yang X, Xie RH, Chen S, Yu W, Liao Y, Krewski D, Wen SW
Using Video Feedback Through Smartphone Instant Messaging in Fundamental Nursing Skills Teaching: Observational Study
JMIR Mhealth Uhealth 2019;7(9):e15386
URL: https://mhealth.jmir.org/2019/9/e15386
doi:10.2196/15386
PMID:31489839

Original Paper

Research and Implementation Lessons Learned From a Youth-Targeted Digital Health Randomized Controlled Trial (the ARMADILLO Study)

Lianne Gonsalves[1,2,3], MSPH; Winnie Wangari Njeri[4], BA; Megan Schroeder[5], MPH; Jefferson Mwaisaka[4,6], MA; Peter Gichangi[4,7,8], MBChB, MMED, PhD

[1]Department of Reproductive Health and Research, World Health Organization, Geneva, Switzerland
[2]Swiss Tropical and Public Health Institute, Basel, Switzerland
[3]University of Basel, Basel, Switzerland
[4]International Centre for Reproductive Health-Kenya, Mombasa, Kenya
[5]Ona, Nairobi, Kenya
[6]Department of Population, Family & Reproductive Health, University of Ghana, Accra, Ghana
[7]Department of Human Anatomy, University of Nairobi, Nairobi, Kenya
[8]University of Ghent, Ghent, Belgium

Corresponding Author:
Lianne Gonsalves, MSPH
Department of Reproductive Health and Research
World Health Organization
Avenue Appia 20
Geneva,
Switzerland
Phone: 41 795006554
Email: gonsalvesl@who.int

Abstract

Background: Evidence is lacking on the efficacy of sexual and reproductive health (SRH) communication interventions for youth (aged 15-24 years), especially from low- and middle-income countries. Therefore, the World Health Organization initiated the Adolescent/Youth Reproductive Mobile Access and Delivery Initiative for Love and Life Outcomes (ARMADILLO) program, a free, menu-based, on-demand text message (SMS, short message service) platform providing validated SRH content developed in collaboration with young people. A randomized controlled trial (RCT) assessing the effect of the ARMADILLO intervention on SRH-related outcomes was implemented in Kwale County, Kenya.

Objective: This paper describes the implementation challenges related to the RCT, observed during enrollment and the intervention period, and their implications for digital health researchers and program implementers.

Methods: This was an open, three-armed RCT. Following completion of a baseline survey, participants were randomized into the ARMADILLO intervention (arm 1), a once-a-week contact SMS text message (arm 2), or usual care (arm 3, no intervention). The intervention period lasted seven weeks, after which participants completed an endline survey.

Results: Two study team decisions had significant implications for the success of the trial's enrollment and intervention implementation: a hands-off participant recruitment process and a design flaw in an initial language selection menu. As a result, three weeks after recruitment began, 660 participants had been randomized; however, 107 (53%) participants in arm 1 and 136 (62%) in arm 2 were "stuck" at the language menu. The research team called 231 of these nonengaging participants and successfully reached 136 to learn reasons for nonengagement. Thirty-two phone numbers were found to be either not linked to our participants (a wrong number) or not in their primary possession (a shared phone). Among eligible participants, 30 participants indicated that they had assumed the introductory message was a scam or spam. Twenty-seven participants were confused by some aspect of the system. Eleven were apathetic about engaging. Twenty-four nonengagers experienced some sort of technical issue. All participants eventually started their seven-week study period.

Conclusions: The ARMADILLO study's implementation challenges provide several lessons related to both researching and implementing client-side digital health interventions, including (1) have meticulous phone data collection protocols to reduce

wrong numbers, (2) train participants on the digital intervention in efficacy assessments, and (3) recognize that client-side digital health interventions have analog discontinuation challenges. Implementation lessons were (1) determine whether an intervention requires phone ownership or phone access, (2) digital health campaigns need to establish a credible presence in a busy digital space, and (3) interest in a service can be sporadic or fleeting.

Clinical Trial: International Standard Randomized Controlled Trial Number (ISRCTN): 85156148; http://www.isrctn.com/ISRCTN85156148

(JMIR Mhealth Uhealth 2019;7(8):e13005) doi:10.2196/13005

KEYWORDS

adolescent health; sexual and reproductive health; health communication; mHealth; Kenya; intervention research

Introduction

In the past 10 years, the digital health field—or the use of digital, wireless, and mobile technologies for health [1]—has exploded in size and scope of interventions. Digital health solutions have been promoted with enthusiasm due to the technology's rapid and widespread proliferation, and the potential of digital health to improve access to health information and services, especially in resource-poor settings [2]. Today, the World Health Organization's Classification of Digital Health Interventions describes a spectrum of solutions for supporting health care providers, health system and resource managers, health data services, and clients of the health system [3].

Young people (individuals between 15 and 24 years) are an especially promising population to reach with digital health interventions. They often face special vulnerabilities, especially related to sexual and reproductive health (SRH). In developing regions, an estimated 33 million women aged 15 to 24 years have an unmet need for contraception [4]; 16 million girls aged 15 to 19 years give birth each year and 3.9 million girls aged 15 to 19 years undergo unsafe abortions [5]. In Kenya, young people between the ages of 15 and 24 years constitute one-fifth of the total population [6]. The most recent Kenya Demographic and Health Survey, found that 37% of young women aged 15 to 19 years and 49% aged 20 to 24 years who are currently married, and 49% of young women aged 15 to 19 years and 64% aged 20 to 24 years who are sexually active but not married, are currently using any form of modern contraception [7]. Even among currently married women, those aged 15 to 24 years still have an unmet need for family planning that is higher than the national estimate of unmet need among all women of reproductive age (15-49 years) [7].

Despite demonstrated SRH needs in Kenya, and around the world, young people have traditionally faced a wide variety of financial, cultural, social, and legal obstacles to obtaining SRH services [8,9]. They are also, however, voracious adopters and innovators when it comes to mobile phone technology [10]. Recent years have seen an explosive proliferation of mobile phone ownership, thus closing ownership gaps across education and wealth levels [11-13]. Kenya leads East Africa in mobile phone infrastructure and innovation (for example, with higher-than-average coverage in rural areas, and long-time use of mobile money programs driving increases in mobile phone access) [13]. As such, youth-targeted digital health solutions appear to be a logical intervention for privately disseminating needed information to a population with special SRH-related vulnerabilities.

Unfortunately, evidence on the efficacy of client-targeted SRH communication interventions for young people, especially in low- and middle-income countries, is severely lacking [14]. Therefore, in 2014, the World Health Organization's Department of Reproductive Health and Research initiated the Adolescent/Youth Reproductive Mobile Access and Delivery Initiative for Love and Life Outcomes (ARMADILLO) study, joined by the International Centre for Reproductive Health–Kenya, and Kenya-based technology partner Ona. An additional research partner also implemented the ARMADILLO study in Peru.

The ARMADILLO study was envisioned as a proof-of-concept intervention study. The intervention was designed as a free, automated, menu-based and on-demand text messaging (short message service; SMS) platform that would provide validated information across a variety of youth-identified SRH domains. The study itself was implemented in two stages. A formative stage 1 identified relevant SRH domains, and then developed and tested the SMS text messaging content and intervention appeal among youth aged 15 to 24 years with qualitative methods [15]. For the stage 2 efficacy assessment [16], we opted to conduct a randomized controlled trial (RCT), long considered the gold standard in health research study design, so that the ARMADILLO study might address repeated global calls for rigorous evidence that digital interventions can (either directly or as secondary outcomes) positively impact health outcomes [14,17-19].

It is established that conducting RCTs on digital health interventions can be challenging due to the study's rigid design, as well as the cost and time often required [20]. However, the purpose of this paper is to describe some additional implementation challenges that arose during the ARMADILLO RCT in Kenya. These issues, which arose during the period of enrollment and early during the intervention period, have implications for both digital health researchers and programmers attempting similar, client-side health communication interventions, especially with young people.

Methods

Overview

The full procedures for the ARMADILLO trial (registration number: ISRCTN85156148) are described in full elsewhere

[16]; briefly, this was an open, three-armed RCT conducted in a peri-urban area in Kwale County, Kenya. The RCT sought to determine whether the provision of on-demand SRH information via text message (arm 1) would result in significant improvement over several SRH knowledge, attitudinal, and behavioral outcomes as compared with periodic messages encouraging self-learning (arm 2) or usual care (no intervention, arm 3). The primary outcome measured change in an index of myths and misconceptions related to contraception. Secondary outcomes measured change in knowledge, attitudes, and behavior for key SRH outcomes (eg, knowledge of HIV/AIDS and its transmission, attitudes around violence against women, engagement in sexual activity).

Following recruitment and consent, participants completed a baseline survey capturing sociodemographic information and primary and secondary outcome measures. Participants were then randomized into a 1:1:1 ratio using a computer-based randomization tool (developed using Node.js and docker). The intervention period lasted seven weeks, at which point data collectors visited participants to administer an endline survey of SRH outcomes. After completing an additional eight-week period, during which no participants received any intervention, participants completed a final, follow-up assessment of SRH outcomes.

The last of the ARMADILLO study participants finished their study period and follow-up period by August 2018. After minor modifications, the full ARMADILLO architecture, consisting of all domains and their subdomain messages, was linked via an overarching domain-selection menu message and made available to participants from all arms for 2 months. The system was taken offline in December 2018 and remains offline while primary and secondary analyses from the trial are being conducted.

The ARMADILLO study obtained ethics review and approval from the World Health Organization's Research Ethics Review Committee (A65892b) and the University of Nairobi/Kenyatta National Hospital (P274/05/2017).

Recruitment

Participants were identified via a household enumeration of eligible youth, which took place October 2017. In this enumeration, the research team used an official record of households (developed in preparation of Kenya's 2019 national census) to map all households in the study area. Trained data collectors recruited from the study area then visited every household (a total of 2132) to identify eligible youth. Household members were deemed eligible if they were between the ages of 18 and 24 years, literate, had their own mobile phone (meaning it was primarily in their possession and they controlled when and with whom they shared access) and reported regular use, had a mobile phone with them at the time of recruitment, and reported current use of text messages.

When RCT recruitment began in February 2018, one eligible youth per household was preselected randomly for recruitment; if they opted not to participate, no one else in the household was eligible. Enrollment of participants was rolling and took place in three waves over seven weeks.

Data Collection

All surveys were completed on a mobile phone via digital form (ODK Collect); surveys were primarily administered by data collectors, although participants filled in the digital forms themselves for certain sensitive questions. Twenty-one individuals from the study site community were hired to serve as data collectors for the RCT. An almost-equal number of male and female data collectors were selected to ensure that all participants would be recruited, consented, and enrolled by someone of the same sex. Most data collectors had completed at least some secondary education.

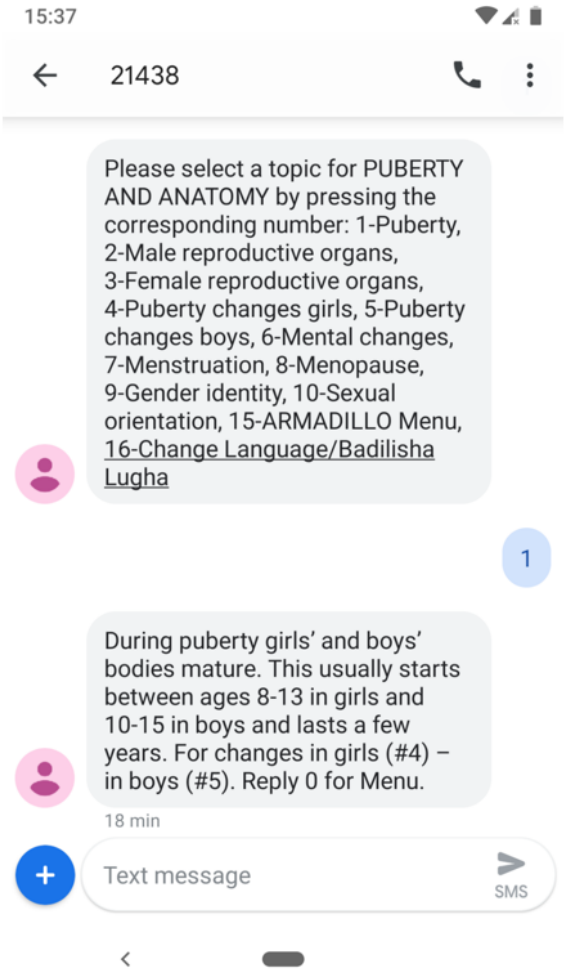
Before participant recruitment, data collectors underwent a three-day training that covered an overview of the study and its purpose, the process for ethically recruiting and consenting individuals, and how to collect data (for participant surveys) via a digital form on mobile phones. Given the taboo nature of an SRH-related study conducted in a conservative community, the training included a special focus on making sure that young participants would feel comfortable speaking with data collectors.

Study Arms Description

After being randomized, participants were intended to automatically enter into one of the three arms the following day: they would receive either their first domain menu (arm 1), domain contact (arm 2) message, or no message (arm 3), marking the start of their intervention period.

Arm 1 provided the ARMADILLO intervention: SMS text messaging content around seven youth-identified SRH domains: puberty/anatomy, pregnancy, relationships, sex, contraception, HIV, and gender-based violence. Arm 1 participants received one SMS text message pushed to their phones every week providing them with a new, unlocked domain menu; at their convenience, they could request further information on any of 5 to 12 numbered subdomains, which then provided them with two to three SMS text messages of validated health information, developed by youth. An SMS quiz was pushed to arm 1 participants' phones at the end of the week to maintain engagement. Any participant who responded received a phone credit equivalent to US $0.50. An example interaction with a domain message, user reply, and subdomain message can be seen in Figure 1.

Figure 1. Example interaction with the ARMADILLO (Adolescent/Youth Reproductive Mobile Access and Delivery Initiative for Love and Life Outcomes) puberty and anatomy domain (seen by arm 1 participants).



Arm 2 matched the system-initiated contacts with participants of arm 1 without providing them access to the ARMADILLO content itself. The intended purpose of arm 2 was to assess whether changes in SRH outcomes were due to exposure to messaging content or the "contact" of the digital intervention itself; that is, would a young person, encouraged by SMS text message to go learn about a given SRH topic, be inspired to seek out information on their own (eg, by talking to friends and family, or looking up information online)? Practically, this meant that arm 2 participants received two pushed messages per week: one alerting them to an SRH domain on which to seek information and an SMS text message quiz at the end of the week. Any response to the SMS quiz received a phone credit equivalent to US $0.50. For both arms 1 and 2, in a reflection of the demographics of the study area, messages were available in both English and colloquial or "street" Coastal Swahili.

Finally, arm 3 was a control arm. Arm 3 participants emulated standard access to SRH information and thus received no messages from this study. The ARMADILLO system was stored on RapidPro (an open-source communication platform), hosted by the technology partner, Ona.

Results

Implementation Challenges

Two seemingly minor methodological decisions had significant implications for the overall success of the trial's enrollment and intervention implementation. We describe these below.

Challenge 1: A Hands-Off Recruitment Process

During the enrollment period, data collectors were instructed to identify the randomly selected youth from each household based on this young person's age, sex, and education level (intentionally, no further identifying information had been collected from the young people during the enumeration process). Data collectors confirmed the youth's identity by collecting these demographic details again. They determined eligibility by asking if the youth owned a phone, asking to see the phone, and collecting the phone number. There was no additional check to verify phone ownership or confirm the phone number provided was in service.

If the young person was confirmed to be eligible and expressed an interest in participating, data collectors consented the youth and began the baseline survey. As part of the consent form (read aloud by the data collector to the youth), all three arms of the ARMADILLO study were described. These were the only

instructions on any of the three arms that participants received, an intentionally hands-off approach to emulate as much of a real-world environment for the intervention as possible. As a result, any participant unfamiliar with the intervention's format (or unclear that they could expect to *receive* any intervention) struggled.

Challenge 2: Design Flaw in Language Selection Menu

The ARMADILLO system was built for two languages, Swahili and English, but there was no clear preference in the study area for which language could serve as a default. As such, the day after enrollment we decided that if a participant was randomized into either arm 1 or 2, they would receive a single SMS text message asking them to indicate in which language they wished to receive messages. A response to this initial language selection SMS text message triggered their first domain menu or domain contact and the start of their seven-week study period. However, a critical by-product of this decision, and a design flaw, was that if a participant did not respond to this message, they were left in a study timeline "stasis." Their seven-week intervention period would not trigger until they responded to the initial SMS text message; as such, they would not time out of the study (and therefore be able to participate in endline data collection) because they had never timed in.

As a result, three weeks after the first study participants were enrolled, passive monitoring of participants' progression through the system revealed a number of arm 1 and arm 2 participants who were trapped at language selection because they had not responded to the initial SMS text message from the study. At this point (as seen in Figure 2), 203 participants had been randomized to arm 1, 221 to arm 2, and 236 to arm 3. Among the 424 participants in arms 1 and 2, only 181 (42.6%) had successfully selected a language and initiated their seven-week intervention period; an estimated 243 still had not proceeded past the entry language menu. Arm 1 had fewer participants stopped at this language menu than arm 2 (107 arm 1 participants versus 136 arm 2 participants); however, over half the participants in each arm were "stuck."

To resolve this, we took a series of successive steps to nudge participants into the system, before eventually integrating a nonresponse mechanism (which should have been done initially). With this mechanism, anyone at the initial language menu now automatically flowed to a Swahili-language domain message (and therefore the seven-week intervention period) after one day of inactivity. Figure 2 describes how the RCT was planned (the green pathway) versus the additional steps the study team had to take to move nonengaging participants into their selected arm.

Reasons for Nonengagement

As seen in Figure 2, nudges started with a few low-interference reminder SMS text messages, which were successful in prompting several participants in each arm to select a language and begin their intervention period, implying that these participants had just required a reminder. Additionally, approximately five weeks after the first participants had been enrolled, the study team called a cross-section of nonengaging participants to encourage them to respond to the language message. During these calls, the study team also learned reasons, unrelated to the language menu option, that participants had not yet responded to the initial message.

When the research team called nonengaging participants, 99 participants in arm 1 remained stuck at the language menu, along with 132 participants in arm 2. The research team called all 231 of these nonengaging participants over two days. The team was successful in reaching 136 participants (59 in arm 1; 77 in arm 2), and we were able to learn their reasons for nonengagement. Among this selected subset of reachable nonengagers, Table 1 lists key reason for nonengagement.

Figure 2. Progression of participants: how entry into the randomized controlled trial was planned (expected engagement) versus additional steps taken (nonengaging route) for each arm.

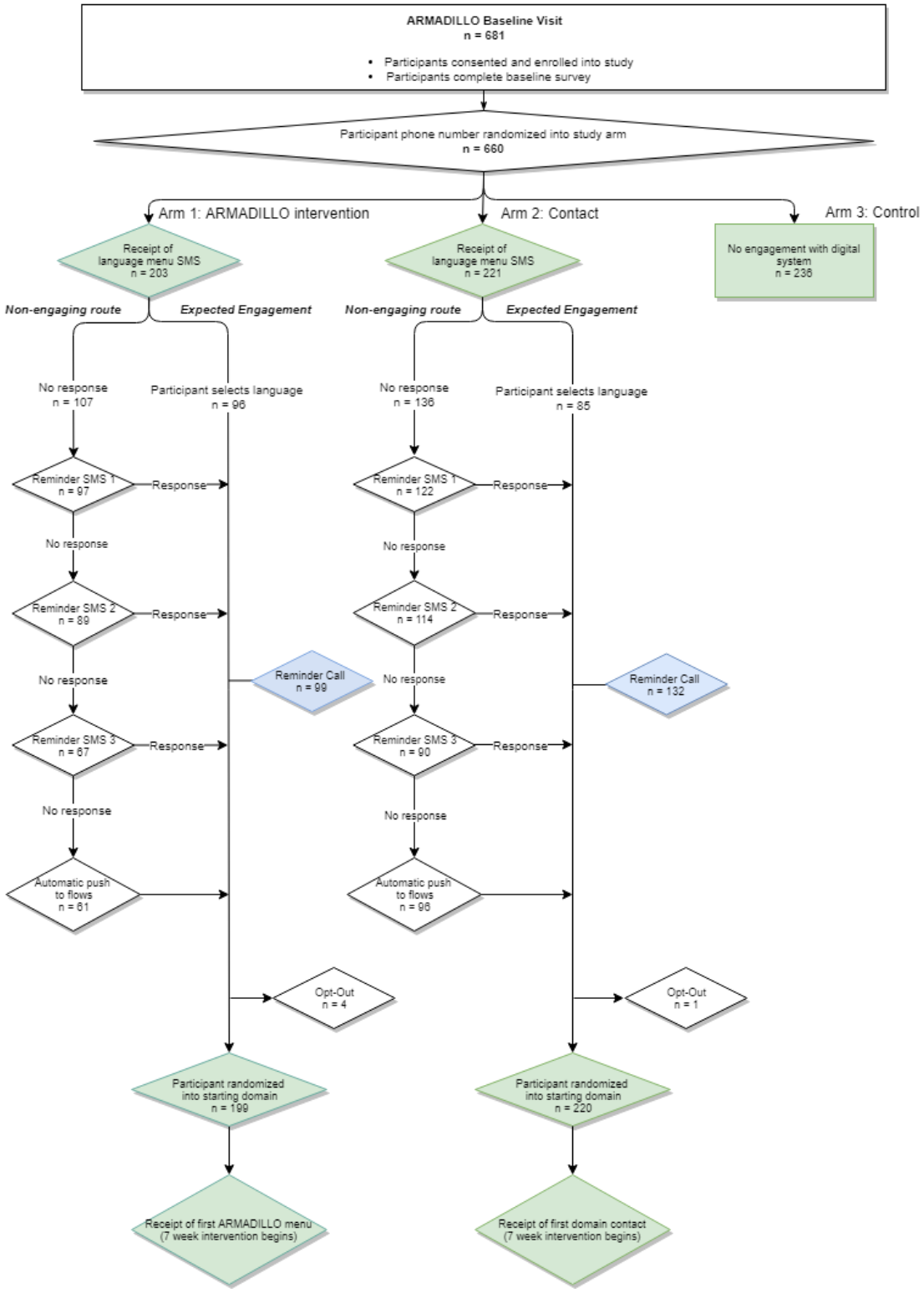


Table 1. Reported reasons for not responding to the ARMADILLO language menu, according to the nonengaging participants in arms 1 and 2 who responded to phone calls from the ARMADILLO team.

Reported reason for nonengagement	Arm 1 (n=59), n (%)	Arm 2 (n=77), n (%)	Total (n=136), n (%)
Eligibility challenges	15 (25)	17 (22)	32 (24)
Person who answered phone was not participant and did not share phone with participant (participant gave someone else's phone number)	7 (12)	12 (16)	19 (14)
Person who answered phone was not participant but did share phone with participant	8 (14)	5 (7)	13 (10)
Did not recognize the system (assumed scam/spam)	8 (14)	22 (29)	30 (22)
Confusion over how to engage with the system	17 (29)	10 (13)	27 (20)
General confusion (nonspecified)	8 (14)	3 (4)	11 (8)
Did not know how to progress/was not sure it was free	4 (7)	3 (4)	7 (5)
Thought the system was supposed to call them	1 (2)	4 (5)	5 (4)
Other (thought messages were time-sensitive; did not know to expect message, thought system was "pushed")	4 (7)	1 (1)	5 (4)
Apathetic about engaging	6 (10)	5 (6)	11 (8)
Technical challenges	9 (15)	15 (19)	24 (18)
Reported not having received messages	3 (5)	6 (8)	9 (7)
Hardware issues: phone was lost or broken	1 (2)	4 (5)	5 (4)
Telco issues: line no longer in service, incomplete number, noneligible mobile network operator	3 (5)	3 (4)	6 (4)
Other (had submitted invalid responses, had multiple phones)	2 (3)	2 (3)	4 (3)
Other reasons	4 (7)	8 (10)	12 (9)

Several nonengagers were found to have violated eligibility criteria, specifically phone ownership. First, 19 phone numbers did not belong to the person who had been recruited to participate (confirmed by the phone owner being outside the 18-24 age range or from outside the study area, and having no recollection of being interviewed). Discussions with the true phone owners indicated that a young study participant may have—out of fear, distrust, or mischievous spirit—opted not to give their own phone number but rather that of a friend, relative, or acquaintance. In other cases, participants may have provided a wrong number, or a data collector may have entered a wrong number. In either case, the phone owners did not recognize the messages from ARMADILLO and did not respond. Phone owners were able to opt out if they wished. However, the individual participants who had provided the numbers were not unenrolled and completed an endline assessment.

An additional 13 calls reached persons who shared a phone with the study participant; for example, one recruited participant was the full-time owner and operator of a phone, but only when his brother was away at university. When the brother returned during the study period, the participant forfeited the phone (purchased by the brother). The brother (and the other nonparticipants in this group) had not recognized the messages arriving to the phone and had not responded. ARMADILLO eligibility criteria had specified phone ownership; therefore, the fluid phone-sharing arrangements meant that these participants violated eligibility criteria as well.

Among eligible participants, the single largest reason for nonengagement was that participants had not recognized the introductory SMS text message as being from the study. Thirty participants indicated they had assumed the introductory message was the start of a scam or that they were being spammed by a third party. An additional 27 of the nonengagers reached indicated being confused by the system. Specific reasons included not being sure how to progress through ARMADILLO, uncertainty that the system was free, thinking that the system was supposed to call them, an assumption that messages had to be responded to within a certain period, and believing ARMADILLO was a push system. An additional 11 expressed some level of apathy with the system, telling the research team that they had been too lazy, too busy, or not interested enough to reply.

Finally, the research team found 24 of the nonengagers had experienced some sort of technical issue. These ranged from numbers being out of service or not on a participating network, participants losing their phones or having other phone issues, and participants reporting either not having received the message or having their responses rejected by the system.

Some participants eventually moved into the system following the call. The remainder, including those who could not be reached by phone, automatically flowed into their first domain shortly thereafter, following the system modification.

Discussion

Principal Findings

This study describes some of the pragmatic implementation challenges that can arise while implementing a rigorous,

multiarm RCT assessing the efficacy of a digital health intervention. Although the quantity and general quality of evidence appear to be increasing in recent years [21,22], RCTs in particular can pose challenges for the digital health field; for example, blinding participants to the intervention they receive is extremely difficult [20]. In addition to being costly, RCTs traditionally also have lengthy recruitment, enrollment, and study periods [23], and the interventions they test—predefined in the trial protocol—remain static for the duration of the study [20]. This can be problematic in a field where innovation and invention advance the field quickly in a short period of time. Finally, as the success of digital interventions can depend as much on contextual factors as on the intervention itself, the appropriateness of RCTs alone to contribute evidence has been debated, with calls evaluations to include robust qualitative components [20,23].

We attempted to account for several of these challenges in the design of the ARMADILLO study. Our RCT had relatively short intervention and follow-up periods, a nod to our transient young population. The development of the study was recognized to be a multiyear process; therefore, we selected SMS text messaging as a delivery channel that, although not at the vanguard of digital health innovation, was and would remain a reliable and universal channel of communication for anyone with a mobile phone. Finally, before developing the RCT, we conducted a robust qualitative phase, which not only vetted the ARMADILLO content but also sought to understand the sociocultural and technological context in which the RCT would be implemented.

However, even while accounting for common challenges to digital health RCTs, two decisions had consequences that threatened the rigor of the planned RCT. One led to data collectors being as hands-off as possible in confirming participant identity and describing the intervention. A second decision introduced an improvised language selection menu to ensure that participants could access messages in their language of choice.

Confronted with dozens of participants stalling at the language menu, we faced a question: why nudge and then push participants into the system at all? Nonengagement with the language menu could be factored into the analysis, for example, by comparing findings using intent-to-treat analysis with those of a per-protocol analysis. The reality was not so simple: the language menu was a last-minute add-on for a study site with two equally used languages; it was separate from the intervention being evaluated (seven weeks of SRH content delivered via text message). As such, nudging or moving participants past the language menu and into their study arm was deemed appropriate. Once participants had flowed into their timed intervention period, we could monitor participants' levels of engagement (or lack thereof); these will be factored into upcoming analyses of the trial results.

Although unplanned and time-intensive, a serendipitous result of calling participants was that the study team was able to communicate directly with nonengaging participants to find out why they had not yet responded. Most reasons had nothing to do with the language menu but rather the intervention or study. However, if the nonresponse mechanism had been built into the language menu initially (as it should have been) so that nonresponding participants flowed directly from language menu to their first domain, we would not have captured the views of these nonengagers.

Broadly, the ARMADILLO study's implementation challenges, which arose during the study's enrollment and initial data collection period, provide several lessons relevant for both research and implementation of client-side digital health interventions.

Research Challenges and Lessons Learned

Lessons learned from the first implementation misstep—the hands-off manner in which participants were recruited and enrolled into the study—can assist future researchers to more carefully design their studies and recruitment procedures, as described subsequently.

Develop Meticulous Phone Data Collection Protocols to Reduce the Possibility of Wrong Numbers

In our calls to nonengagers, we connected with several people who were from the study area but who were not our recruited participants. We have no real way of knowing why we ended up with these wrong numbers. Perhaps it was overenthusiasm: participants who did not quite meet our eligibility criteria but still wanted to participate may have borrowed the phone or phone number of a friend or family member. Conversely, it might have been underenthusiasm; we may have overestimated young people's comfort with participating in a study that would involve SRH messages arriving to their phones.

Data collector trainings establish and drill procedures for recruiting, consenting, and enrolling participants. For digital health interventions, these procedures must also include multiple steps for cross-checking that phone-related eligibility criteria are met, and that the correct phone number is collected. Using ARMADILLO's phone-related eligibility criteria as an example, simple measures can reduce phone-related recruitment error:

1. asking to see the participant's phone;
2. repeating the phone number back to the participant;
3. calling the participant's phone and checking the phone to confirm receipt; and
4. probing to ascertain whether the participant meets ownership criteria as defined by the study (eg, Who buys air time? Who purchased the subscriber identification module (SIM) or phone? Who else can use the phone, and who decides this?)

Train Participants in Efficacy Assessments on the Digital Intervention (Even If It Is Modeled Off Similar Services)

ARMADILLO's on-demand querying of information (arm 1) uses a number-based menu that makes it virtually identical in format to M-PESA, a mobile money service used by more than 18 million people in Kenya [24]. The assumption was that the ping-pong format of user-system interaction would feel familiar to users, and minimal explanation would be necessary. We were also concerned about overly training participants in arm 1 only for them to be confused, disappointed, or less willing to engage if they were to be randomized into one of the other two arms.

Unfortunately, for a minority of participants in arms 1 and 2, the lack of detailed explanations resulted in confusion as to how to engage with the system. Additionally, participants in both arms did not make the connection between the study and the study's SMS text message, even after receiving this SMS text message the day after they were interviewed.

It is tempting, during an evaluation of client-side digital health interventions, to adopt a hands-off approach with users; a user's ability or inability to successfully engage with a system is data in and of itself. This approach can be appropriate for evaluations of service rollout (coverage assessments) or studies of usability or acceptability [1]. However, this randomized trial was an efficacy assessment meant to assess whether the digital intervention affected health outcomes under an ideal research setting. Therefore, with a focus on health rather than usability outcomes, participants must be fully trained to be able to use the system as intended. Data collectors should be instructed to be explicit about the service at the point of recruitment and walk all participants through the following, in detail:

1. How the system works: describing all arms in detail, what cost (if any) is incurred for participants, when and how often they can engage.
2. How to recognize the system: what number or short code does the system use, when can they expect to receive messages (for any pushed interactions).
3. How to use the system: showing example messages on a phone, letting the participant try querying the system on the data collector's phone.

Client-Side Digital Health Interventions Have Analog Discontinuation Challenges

For a study focusing on a young and geographically mobile population, a participant's phone was not only an essential part of the intervention but also an important tool for locating the participants and scheduling endline and follow-up interviews. That said, the young participant's phone itself can become a source of discontinuation. Youth participants lost possession of the phone subscribed to the ARMADILLO system because it was lost or stolen, they upgraded to a new phone, the phone broke (temporarily or permanently), they switched SIMs or providers, or they loaned the phone to a friend or relative for short or long periods of time. Common phone-related discontinuation challenges should be considered with other sources of discontinuation when calculating sample size to ensure that an otherwise robust study does not become underpowered because several participants lose their phone.

Service Rollout Challenges and Lessons Learned

Although ARMADILLO was an RCT and not a full-scale digital health campaign, lessons learned from the calls to nonengagers (made as a result of the second implementation misstep) can contribute to the successful development and rollout of both categories of digital health communication services.

Determine Whether an Intervention Requires Phone Ownership or Phone Access

When developing targeted client communication digital health interventions (for example, SMS text messages to expectant mothers throughout their pregnancies; alerting clients about health tests results) [3], especially interventions around sensitive issues, such as SRH including HIV, it is critical to understand what comprises phone ownership in a given setting.

Outreach to ARMADILLO study users found that phone ownership was a fluid concept; a phone might belong to a user for a certain period during the day, during a certain time of year, or until someone gets an upgrade and passes down their old phone. There is a general need for data on the demographics and practices of phone ownership and phone sharing.

In Kenya, data from a 2009 nationally representative survey of over 30,000 individuals aged 16 years and older showed that although 85% of individuals indicated that they had used a mobile phone, only 44% owned their mobile phone. Phone sharers were predominantly female (65%), and lower levels of phone ownership were observed among the youngest respondents [25]. More recent regional data suggests that women in sub-Saharan Africa are 14% less likely than men to own a mobile phone (defined as having sole or main use of a SIM card or mobile phone which does not require a SIM), and women are 34% less likely than men to use mobile internet [26].

Client-side digital health interventions provide an important mechanism for conveying health information to hard-to-reach populations. However, digital health implementers should take care to consider whether their intervention requires mobile phone ownership (and what that means) or mobile phone access for effective, acceptable, equitable, and safe engagement with users. Digital health interventions will reach their intended populations only when implementers understand the realities of (1) how age, gender, income, or urban-rural status influence likelihood of phone ownership and (2) how phones are shared within households.

Digital Health Campaigns Must Establish a Credible Presence in a Competitive and Busy Digital Space

Similar to preparation for a health communication campaign, the ARMADILLO study team conducted extensive outreach with county-level Ministry of Health officials as well as community leaders to ensure that communities were sensitized to the coming research. However, given the design of this efficacy assessment involved a control group receiving no intervention, special care was taken to avoid contamination across groups of participants in the study area by not advertising the ARMADILLO system within the community itself—not a strategy to be recommended outside a research setting.

A downside to staying quiet about the service was that ARMADILLO did not automatically have the trust or recognition of its participants. ARMADILLO's formative stage found that young people (and, importantly, their parents or caregivers) were enthusiastic about a phone-based health campaign, so long as they knew it was coming from a credible, trustworthy source [27]. That the single largest reason for nonengagement in this study was not recognizing or trusting the sender reinforces those findings.

An additional reason for distrust was likely how the ARMADILLO registered on participants' phones, a weekly SMS text message from a numeric short code. Other large-scale

pushed-SMS text message campaigns use customized names for easy recognition (eg, SMS text message coming from "ARMADILLO"). However, this was not possible given ARMADILLO's ping-pong format, in which participants were expected to interact with the system. At the same time, public awareness and news coverage of mobile-based financial scams is increasing [28], with Kenyans being advised to be on guard against social engineering by scammers in an attempt to gain personal and financial information [29]. Therefore, incoming messages from a numeric short code may have been viewed with added skepticism.

In implementing client-side digital health communication campaigns, the importance establishing its trustworthiness within the community (both intended users and the community at large) cannot be overemphasized.

Interest in a Service Can Be Sporadic and Fleeting

Finally, 11 of the nonengaging participants (8% of the total number of nonengagers reached by phone) saw the messages but then were either too lazy, too busy, or forgot entirely to respond. These participants provide an important reminder that—however exciting a digital system is—intended users may not wait by their phones for messages or opportunities to engage.

Purely on-demand interventions rely on user initiative and therefore user interest for accessing information. However, just as all mobile phone users may forget or get too busy to engage in personal messaging, even pushed message campaigns, whether providing targeted or untargeted client communication, would do well to remember that users' interest and bandwidth to engage will wax and wane over the course of a campaign.

Conclusions

Digital health interventions are lauded for their potential to overcome health client, provider, and system challenges that hinder the coverage or effect of existing health interventions. However, the digital health field is still in its adolescence—and enthusiasm often outpaces evidence. Most recently, a Lancet editorial cautioned against "digital exceptionalism" and highlighted the risk to patients and the health system if we fail to robustly evaluate digital health interventions [30].

The ARMADILLO study was developed with a sole focus on robust evaluation and despite the challenges previously described, preliminary data review has indicated that enough participants received necessary parts of the intervention to be able to power the planned primary and secondary analyses. However, even the process of implementing this multiarm RCT has eliminated certain illusions of digital exceptionalism. Research on digital health interventions faces the same implementation challenges as other research on nondigital health interventions: difficulty reaching the target population, trouble following-up with participants, and overcoming reluctance to engage. If these challenges are not adequately prepared for in future research, there will be adverse implications on the availability and quality of evidence in a field where evidence is sorely needed.

Acknowledgments

We appreciate technical advice from Erick Ngwiri. The manuscript represents the view of the named authors only. The ARMADILLO study was funded by the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction.

Authors' Contributions

LG conceived of the ARMADILLO study, co-led study implementation with JM and WW, and led this manuscript. WW pulled the data for this publication and cross-checked participants' progression through the ARMADILLO system with MS. MS accessed ARMADILLO's backend infrastructure to check participants' progression and developed this publication's flowchart. JM contributed to the organizing of the paper and provided Kenya-specific contextual details. PG is the principal investigator of the ARMADILLO study's Kenya site. He and the other authors provided substantial review of and edits on the manuscript.

Conflicts of Interest

MS is employed by Ona, the social enterprise contracted to develop the ARMADILLO intervention and other study infrastructure. The other authors have no conflicts of interest to declare.

References

1. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252183/9789241511766-eng.pdf?sequence=1 [accessed 2019-06-04]
2. Vital Wave Consulting. mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. Washington, DC: UN Foundation-Vodafone Foundation Partnership; 2009.
3. World Health Organization. Classification of Digital Health Interventions v1. Geneva, Switzerland: World Health Organization; 2018.
4. MacQuarrie KL. Unmet Need for Family Planning Among Young Women: Levels and Trends. Rockville, MD: ICF International; 2014.

5. World Health Organization. 2018 Feb 23. Adolescent pregnancy: key facts URL: http://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy [accessed 2019-06-04]
6. Population and Housing Census 2009. Nairobi, Kenya: Kenya National Bureau of Statistics; 2009.
7. Kenya National Bureau of Statistics. Kenya Demographic and Health Survey 2014. Rockville, MD: The DHS Program, ICF International; 2015.
8. Chandra-Mouli V, McCarraher DR, Phillips SJ, Williamson NE, Hainsworth G. Contraception for adolescents in low and middle income countries: needs, barriers, and access. Reprod Health 2014 Jan 02;11(1):1 [FREE Full text] [doi: 10.1186/1742-4755-11-1] [Medline: 24383405]
9. Adolescent-friendly contraceptive services: mainstreaming adolescent-friendly elements into existing contraceptive services. In: High-Impact Practices in Family Planning (HIPs). Washington DC, USA: USAID; 2015.
10. Interagency Youth Working Group. Mobile technology for health. YouthLens No. 38 2013.
11. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile. JMIR Mhealth Uhealth 2015 Feb 25;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
12. Communications Authority of Kenya. Fourth Quarter Statistics Report for the Financial Year 2015/2016 (April-June 2016). Nairobi, Kenya: Communications Authority of Kenya; 2016.
13. The Mobile Economy: Africa 2016. London: GSMA; 2016. URL: https://www.gsma.com/mobileeconomy/africa/ [accessed 2019-06-04]
14. L'Engle KL, Mangone E, Parcesepe A, Agarwal S, Ippoliti N. Mobile phone interventions for adolescent sexual and reproductive health: a systematic review. Pediatrics 2016 Sep;138(3):A [FREE Full text] [doi: 10.1542/peds.2016-0884] [Medline: 27553221]
15. Gonsalves L, L'Engle KL, Tamrat T, Plourde KF, Mangone ER, Agarwal S, et al. Adolescent/Youth Reproductive Mobile Access and Delivery Initiative for Love and Life Outcomes (ARMADILLO) Study: formative protocol for mHealth platform development and piloting. Reprod Health 2015 Aug 07;12:67 [FREE Full text] [doi: 10.1186/s12978-015-0059-y] [Medline: 26248769]
16. Gonsalves L, Hindin MJ, Bayer A, Carcamo CP, Gichangi P, Habib N, et al. Protocol of an open, three-arm, individually randomized trial assessing the effect of delivering sexual and reproductive health information to young people (aged 13-24) in Kenya and Peru via mobile phones: adolescent/youth reproductive mobile access and delivery initiative for love and life outcomes (ARMADILLO) study stage 2. Reprod Health 2018 Jul 11;15(1):126 [FREE Full text] [doi: 10.1186/s12978-018-0568-6] [Medline: 29996854]
17. Al-Shorbaji N, Geissbuhler A. Establishing an evidence base for e-health: the proof is in the pudding. Bull World Health Organ 2012 May 01;90(5):322-322A [FREE Full text] [doi: 10.2471/BLT.12.106146] [Medline: 22589560]
18. Kumar S, Nilsen W, 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]
19. Sondaal S, Browne J, Amoakoh-Coleman M, Borgstein A, Miltenburg A, 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 May 04;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
20. Dick S, O'Connor Y, Heavin C. Randomized controlled trials as a method of evaluating mobile health interventions. In: Proceedings of the 25th European Conference on Information Systems. 2017 Presented at: 25th European Conference on Information Systems (ECIS); June 5-10, 2017; Guimarães, Portugal p. 2721-2731 URL: https://pdfs.semanticscholar.org/417b/1c641dcd0fb331a9c844ca9006ccd86e30ab.pdf
21. Labrique A, Vasudevan L, Chang LW, Mehl G. H_pe for mHealth: more "y" or "o" on the horizon? Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]
22. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing standards to evaluate the impact of integrating digital health into health systems. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S5-S17 [FREE Full text] [doi: 10.9745/GHSP-D-18-00230] [Medline: 30305335]
23. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
24. Monks K. CNN. 2017 Feb 04. M-Pesa: Kenya's mobile money success story turns 10 URL: https://edition.cnn.com/2017/02/21/africa/mpesa-10th-anniversary/index.html [accessed 2019-06-04]
25. Wesolowski A, Eagle N, Noor A, Snow R, Buckee C. Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS One 2012;7(4):e35319 [FREE Full text] [doi: 10.1371/journal.pone.0035319] [Medline: 22558140]
26. Rowntree O. The Mobile Gender Gap Report 2018. London: GSMA; 2018. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/04/GSMA_The_Mobile_Gender_Gap_Report_2018_32pp_WEBv7.pdf [accessed 2019-06-05]
27. Mwaisaka J, Gonsalves L, Say L, Gichangi P. "What's that on your phone?" The aftermath of parents finding sexual and reproductive health messages on their children's phone in coastal Kenya. J Health Soc Sci 2018;3(2):147-156 [FREE Full text] [doi: 10.19204/2018/whts5]

28. Muchiri D. Mobile money, loan, ATM fraud & crime. Nairobi, Kenya; 2016 Aug 11 Presented at: ICPAK Annual Forensic Audit Conference; Aug 11, 2016; Nairobi, Kenya URL: https://www.icpak.com/wp-content/uploads/2016/08/DAY2-SESSION2-Mobile-Money-Dennis-Muchiri.pdf
29. BBC News. 2018 Jul 20. Phone scam: how Kenyans are losing money URL: https://www.bbc.com/news/world-africa-44899854 [accessed 2019-06-08]
30. The Lancet. Is digital medicine different? Lancet 2018 Jul 14;392(10142):95. [doi: 10.1016/S0140-6736(18)31562-9] [Medline: 30017135]

Abbreviations

ARMADILLO: Adolescent/Youth Reproductive Mobile Access and Delivery Initiative for Love and Life Outcomes
RCT: randomized controlled trial
SIM: subscriber identification module
SMS: short message service
SRH: sexual and reproductive health

Edited by G Eysenbach; submitted 03.12.18; peer-reviewed by K Kalke, W Xie, N Levitz; comments to author 17.04.19; revised version received 03.06.19; accepted 21.07.19; published 27.09.19.

Please cite as:
Gonsalves L, Njeri WW, Schroeder M, Mwaisaka J, Gichangi P
Research and Implementation Lessons Learned From a Youth-Targeted Digital Health Randomized Controlled Trial (the ARMADILLO Study)
JMIR Mhealth Uhealth 2019;7(8):e13005
URL: http://mhealth.jmir.org/2019/8/e13005/
doi:10.2196/13005
PMID:31573939

Original Paper

A Biofeedback App to Instruct Abdominal Breathing (Breathing-Mentor): Pilot Experiment

Corinna Anna Faust-Christmann[1], PhD; Bertram Taetz[1], PhD; Gregor Zolynski[1], PhD; Tobias Zimmermann[1], MSc; Gabriele Bleser[1], PhD

wearHEALTH, Department of Computer Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Corresponding Author:
Corinna Anna Faust-Christmann, PhD
wearHEALTH
Department of Computer Science
Technische Universität Kaiserslautern
Gottlieb-Daimler-Str Gebäude 48
Kaiserslautern, 67663
Germany
Phone: 49 631 205 3456
Email: christmann@cs.uni-kl.de

Abstract

Background: Deep and slow abdominal breathing is an important skill for the management of stress and pain. However, despite multiple proofs on the effectiveness of biofeedback, most breathing apps remain limited to pacing specific breathing patterns, without sensor feedback on the actual breathing behavior.

Objective: To fill this gap, an app named *Breathing-Mentor* was developed. This app combines effective visualization of the instruction with biofeedback on deep abdominal breathing, based on the mobile phone's accelerometers. The aim of this pilot study was to investigate users' feedback and breathing behavior during initial contact with the app.

Methods: To reveal the possible effects of biofeedback, two versions of the mobile app were developed. Both contained the same visual instruction, but only the full version included additional biofeedback. In total, 40 untrained participants were randomly assigned to one of the two versions of the app. They had to follow the app's instructions as closely as possible for 5 min.

Results: The group with additional biofeedback showed an increased signal-to-noise ratio for instructed breathing frequency (0.1 Hz) compared with those using visual instruction without biofeedback ($F_{1,37}$=4.18; P<.048). During this initial contact with the full version, self-reported relaxation effectivity was, however, lower than the group using visual instruction without biofeedback (t_{37}=−2.36; P=.02), probably owing to increased cognitive workload to follow the instruction.

Conclusions: This study supports the feasibility and usefulness of incorporating biofeedback in the Breathing-Mentor app to train abdominal breathing. Immediate effects on relaxation levels should, however, not be expected for untrained users.

(JMIR Mhealth Uhealth 2019;7(9):e13703) doi:10.2196/13703

KEYWORDS

mobile health; relaxation; pain management; biofeedback; respiration; breathing exercises; feasibility study

Introduction

Mobile Stress Management

Chronic stress has been identified as a critical factor that influences people's physical and mental well-being [1,2]. However, the effect of a stressor on an individual's well-being also depends on his or her coping mechanism [3]. In addition to stress management group interventions and self-help literature, the use of stress management apps makes it possible nowadays to learn a broad range of problem-focused and emotion-focused coping methods [4-6]. Moreover, relaxation methods are also commonly integrated in apps for the management of chronic pain [7,8], chronic diseases [9], and anxiety [10]. Through interactive design and gamification, such apps can potentially increase the users' motivation [11,12]. This could reduce the economic burden for the health care system [13]. There are first indications that some stress management apps are indeed effective [14,15], underpinning the usefulness of this prevention approach.

Breathing Apps

Deep and slow diaphragmatic breathing can lead to a state of relaxation. Therefore, it is frequently taught as a basic strategy for the management of stress, anxiety, posttraumatic stress disorder [16], and pain [17]. Traditionally, breathing trainings are guided by health professionals, but the increasing importance for technology-driven approaches such as health apps can be attributed to financial reasons [13]. A broad range of apps specially designed for breathing trainings are available, but breathing exercises are also regularly incorporated in stress and anxiety management apps [5,10,18]. Most of these apps simply pace a distinctive breathing pattern, using audio or visual instructions. Regarding the effectiveness of these instructions, it has been shown that a wave-based visualization of the desired breathing pattern can be more easily followed compared with a circle-based visualization or a traditional audio instruction [19].

Besides pacing, providing biofeedback is another approach for breathing trainings (eg, [20-24]). With biofeedback, information from 1 or multiple sensors is used to gain greater awareness of physiological functions. Besides breathing rate, current stress management apps also target skin conductance [25] and heart rate [26,27]. Most mobile biofeedback solutions, however, require additional costly devices with integrated sensors (eg, a belt [28], wearable textile sensors [21,29], or clothing-adhered biosensors [30]).

In this study, biofeedback refers to feedback about the movement of the abdomen during a breathing task. One example for this kind of biofeedback is the BellyBio Interactive Breathing app for iOS devices by RelaxLine. It uses the mobile phone's built-in accelerometers to capture the abdominal breathing movements. For deep and slow breathing, the sound of the ocean is transformed to relaxing music. However, so far, no study has analyzed the effectiveness of such abdominal breathing feedback. Moreover, the app is recommended only for people who are already familiar with breathing exercises. Direct instructions should be used for novices instead [19].

The *Breathing-Mentor* app is a biofeedback breathing app that was developed to provide such direct instructions. It combines the effective wave-base visualization of the desired breathing pattern [19] with biofeedback on the actual breathing behavior, using the mobile phone's accelerometers. This approach allows comparing the desired breathing pattern with the actual breathing behavior in real time.

To investigate the feasibility and usefulness of the additional biofeedback, a control version without biofeedback, that is, with visual instruction only, was implemented as well. For this purpose, we conducted a user study to reveal how people who are unfamiliar with breathing exercises deal with both versions of *Breathing-Mentor*. The focus of this study was to determine the users' ability to follow the breathing instructions and their subjective usage experience.

Methods

The Interface of Breathing-Mentor

The biofeedback signal is drawn over the sine wave (dark line, not present in the control version). It is obtained from the mobile phone's accelerometers, given that the mobile phone is correctly positioned on the user. The latter is supported through an interactive calibration procedure. During the study, the mobile phone was fixed in a custom (three-dimensional [3D]-printed) frame, and the latter was fixed with an elastic band around the upper abdomen. Figure 1 shows the setup and the training user interface.

Figure 1. The Breathing-Mentor training user interface combines graphical (moving sine wave) and text instructions (inhale/exhale, counting from 1 to 4) for deep, slow abdominal breathing with biofeedback (dark line, not present in the control version).

Signal Processing Approach of Breathing-Mentor

The overall signal processing approach for transforming the accelerometer measurements into the breathing signal as visualized on the screen and used for data analysis is detailed below.

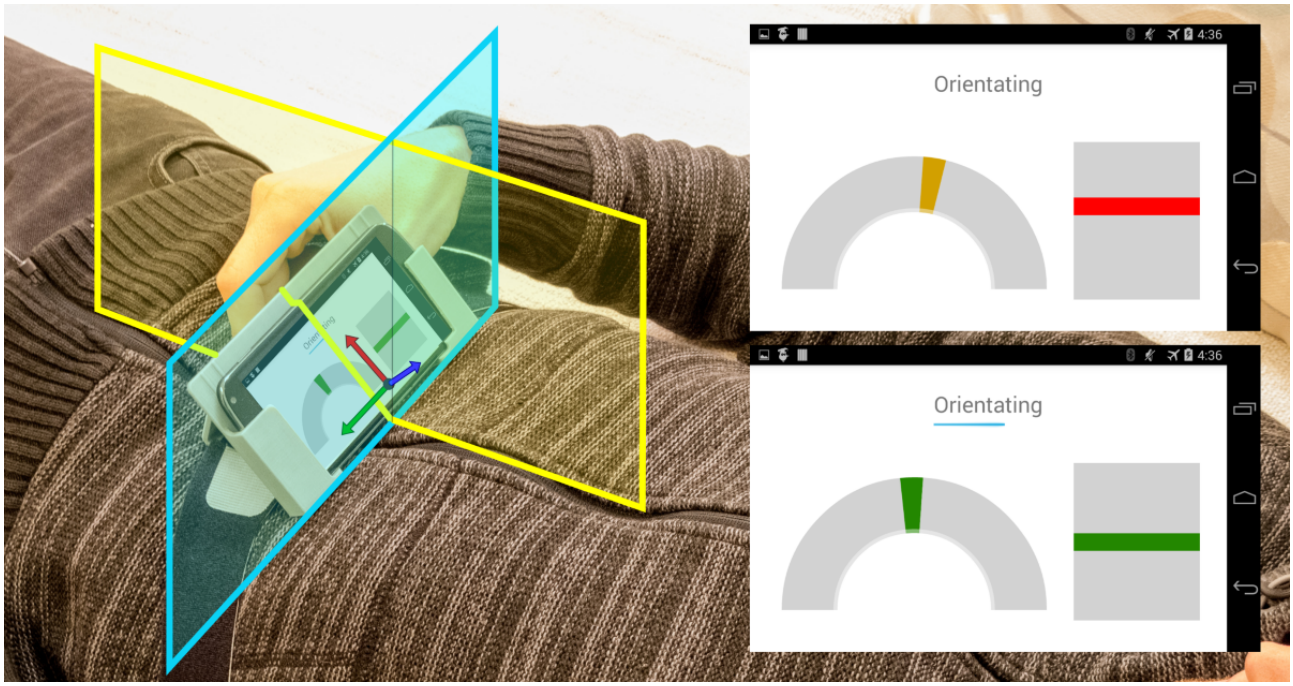
Accelerometers measure 3D linear acceleration, a combination of body acceleration and acceleration resulting from gravity, in the local sensor coordinate system. As the participants are stationary during breathing training, the acceleration resulting from gravity constitutes the major portion of the measurement. Moreover, in the training target pose, this component provides information about lateral and anterior and posterior tilt of the mobile phone with regard to the sagittal and transversal body plane, respectively (see Figure 2). The basic idea is that—with the frame including the mobile phone being placed on the upper abdomen—abdominal breathing results not only in small up and down movements of the mobile phone but also in a change of the mobile phone's orientation, where the tilt change relative to the transversal body plane is dominant. This again results in acceleration measurements with major dispersion direction approximately in the sagittal plane.

These assumptions were confirmed in pretests (1 min, 3 trials) with 5 persons already trained in abdominal breathing. For these trials, the frame including the mobile phone was positioned by the investigator (instructed by the algorithm developer) with its base above the center of the upper abdomen, so that the mobile phone's long edge was approximately leveled, and the display was facing the person. The recorded accelerometer data from these trials were then used to obtain the major dispersion direction as the first principal component. A reference range (representing deep abdominal breathing) was extracted by projecting all recorded accelerometer vectors onto this principal component and calculating the average minimum and maximum values over the test persons.

Calibration Procedure

The average accelerometer vector was also used as reference vector for aiding a repeatable positioning of the custom frame on the participants of the study and thus improving the validity and reliability of the extracted breathing signal. For this, the app provided an interactive calibration procedure with traffic light feedback on the angle deviations between the currently measured accelerometer vector and the reference vector in the xy-plane and in the xz-plane (green: <5°, orange: <15°, red: otherwise; see Figure 2). The angle deviations were calculated using scalar products between the respective vectors. This is based on the assumption that the mobile phone is kept rather stationary during the procedure, and therefore, the accelerometer measures mainly acceleration resulting from gravity, as mentioned above. The angle deviations in the xy-plane and in the xz-plane can be controlled by slowly moving the custom frame including the mobile phone on the upper abdomen laterally or forward and backward, respectively. For a successful alignment, the deviation was required to be in the green area (below 5°) in both planes for 5 seconds. The breathing signal was then obtained from the live accelerometer measurements by applying an infinite impulse response filter (resistor-capacitor low-pass filter) with cut-off frequency of 0.5 Hz, projecting the filtered measurements on the major dispersion direction again using the scalar product and scaling the result so that the reference range mapped to (−1, 1) according to the target sine wave.

Figure 2. Positioning of the smartphone on the upper abdomen and interactive calibration procedure with traffic light feedback, aiding a repeatable positioning during the study. The yellow and cyan rectangles indicate the sagittal and transversal body planes, respectively. The coordinate system denotes the sensor coordinate frame, in which the accelerometer measurements are given. In the user interface, the half circle refers to the alignment in the smartphone's xy-plane and the rectangle refers to the alignment in the xz-plane. For a successful alignment (through manually adjusting the position of the custom frame on the upper abdomen), both marks were required to be in the green area for five seconds.



Pilot Study

To investigate the feasibility and usefulness of the additional biofeedback from the user perspective, we conducted a user study to reveal how people who are unfamiliar with breathing exercises deal with *Breathing-Mentor* compared with those using the control version of the app without additional biofeedback.

Study Protocol

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the local Ethics Committee of the Department for Social Sciences. All participants (see Participants section) gave their informed consent for inclusion in the beginning. They were randomly assigned to the experimental group (EG) with biofeedback or the control group (CG) without biofeedback. In the beginning, previous experience with breathing exercises for relaxation was screened as described by Chittaro and Sioni [19]; see the Results section for details. Then, the investigator fixed the custom frame including the mobile phone (LG Nexus 4, sensor: InvenSense MPU-6050) over the participant's clothes on the upper abdomen using the elastic band while ensuring that the clothes did not fall in folds. The calibration procedure was then performed to ensure correct positioning of the mobile phone for all participants for this study (see Calibration Procedure section for details). Participants lay horizontally during the whole procedure with their head placed comfortably on a pillow. This allowed a direct view on the mobile phone's display.

The 3 measurement blocks are described in Table 1. Each block lasted 5 min, with breathing performance being recorded with the mobile phone's accelerometers (see Signal Processing Approach of Breathing-Mentor section). The baseline block with no mobile phone-guided instruction was included to check if there were systematic differences of abdominal breathing patterns between the 2 groups. In the training block, participants were asked to follow the instructions given by the app as closely as possible, while breathing with the abdomen. The objective was to find out whether the required abdominal breathing pattern (6 cycles/min, 0.1 Hz) could be followed more easily with additional biofeedback on the breathing performance. Although deep abdominal breathing normally increases relaxation sensation in experienced users of breathing exercises, this is not necessarily the case for people who are unfamiliar with breathing tasks. Therefore, a questionnaire regarding the app's effectiveness to support the breathing exercise and its effectiveness to evoke relaxation were assessed directly after the training block (questionnaire provided by Chittaro and Sioni [19]; see the Results section for details). The postmeasurement block was identical to the baseline block without mobile phone-guided instructions. It was included to test whether a single 5-min training session is already enough to evoke changes in the abdominal breathing patterns without further training.

Table 1. Description of the 3 measurement blocks.

Block sequence	Verbal instruction	Screen content	Control questions directly after the block
Baseline	Please breathe as slowly and deeply as possible with the abdomen.	Blank screen. The word *start* appears for 5 seconds. The word *stop* appears after 5 min.	The instruction was easy to follow (1=totally disagree—5=totally agree)
Training	Please follow the instructions on the screen as closely as possible while breathing with the abdomen.	Interface of Breathing-Mentor (see Figure 1), the dark line for biofeedback was not included for the control group.	Questionnaire on the app's effectiveness (1=totally disagree—5=totally agree)
Post	Please breathe as slowly and deeply as possible with the abdomen.	Blank screen. The word *start* appears for 5 seconds. The word *stop* appears after 5 min.	The instruction was easy to follow (1=totally disagree—5=totally agree)

Participants

A total of 40 participants took part in the pilot study. One person from the CG was excluded owing to a chronic respiratory disease, resulting in a final sample size of 39. The mean age was 26.51 years (range 20-42 years, SD 4.41 years). Groups did not differ significantly with regard to age (t_{37}=−0.93, P=.36) or sex ratio (males/females=9/10 in the CG, 10/10 in the EG, χ^2_1=0.03, P=.87).

Statistical Analyses for the Baseline

As no special breathing frequency was instructed in the baseline block, the power of all slow breathing–related frequency bands (0.055-0.195 Hz, width 0.01 Hz each) was compared between both groups in a variance analysis with repeated measurements to check for systematic differences between the groups. Neither systematic group effects nor interaction of group with frequency bands was expected for the baseline block.

Statistical Analyses for the Training Block

For the training block, the following 2 measures of the objective breathing behavior comparable with the study of Chittaro and Sioni [19] were calculated for each minute of the 5-min interval:

The first measure, the spectral power in the recommended frequency band (0.09-0.11 Hz), indicates how intensely the respiratory act is performed for the recommended frequency band. The second measure, respiratory signal-to-noise ratio (SNR), describes the ratio between the power of the recommended breathing frequency band (0.09-0.11 Hz) and the power in the entire breathing spectrum (excluding the band of the recommended frequency, the 0-0.05 Hz band to remove low-frequency fluctuations, and the direct current offset; see [19] for details). It reflects the ability of participants to correctly follow the instructions provided by the app.

Both, the spectral power in the recommended frequency band as well as the respiratory SNR are expected to increase in both groups for the training block owing to the visual instruction for

the 0.1-Hz breathing rate, compared with the baseline condition. If the additional biofeedback actually enhances performance during the breathing exercise, there should be a main effect of group for both dependent measures. The additional within-subject factor *time* (5 steps, 1 min each) allows investigating changes in performance over time. Both groups are expected to increase performance over time for both dependent measures.

Statistical Analyses for the Postmeasurement Block

To test whether a single 5-min training session is already enough to cause changes in the abdominal breathing patterns toward the requested breathing frequency (0.1 Hz), the spectral power in the recommended frequency and the SNR of the postmeasurement block were compared in both groups with the baseline in 2 variance analyses with repeated measures.

Statistical Methods

For single comparisons, *t* tests for independent samples (group comparisons) and *t* tests for dependent samples (comparisons between blocks and minutes) are described. Please note that the given sample size only allows to reveal large effect sizes (0.8). *F* and *P* values are described in the context of variance analyses and *t* and *P* values for *t* tests.

Results

Comparability of Groups Before Training

The screening for previous experience with breathing exercises for relaxation revealed no systematic differences between groups (see Table 2 for details). Although most participants were aware that there is a difference between abdominal and thoracic breathing and that the former can be used for relaxing purposes, only few participants actually practiced breathing and meditation exercises in their daily lives.

For the baseline block, there was a main effect of frequency bands with more power for frequency bands near the normal breathing rate (0.2 Hz; see Figure 3 for details). There were no main effect of group and no interaction between group and frequency bands (see Table 3 for details). Summarizing, the baseline measurements did not reveal any systematic group differences for objective slow abdominal breathing behavior. The subjective ratings on how easy the instruction was to follow did not reveal group differences either (t_{37}=0.56, *P*=.58; EG: mean 4.55, SD 0.69; CG: mean 4.68, SD 0.82).

Table 2. Screening of previous experience with breathing exercises for the control group and experimental group. Absolute frequency of yes and no answers, chi-square values, and *P* values are described for each item.

Item	Control group (yes/no)	Experimental group (yes/no)	Chi-square (*df*)	*P* value
Do you know the difference between abdominal and thoracic breathing?	15/4	16/4	0.01 (1)	.94
Do you know that abdominal breathing is used in the context of relaxing exercises?	13/6	13/7	0.05 (1)	.82
Do you use breathing exercises for relaxation?	4/15	5/15	0.08 (1)	.77
Do you meditate regularly? (at least once a month)	2/17	4/16	0.67 (1)	.41

Figure 3. Mean powers of frequency bands in the baseline block for both groups. Error bars indicate standard error of the mean.

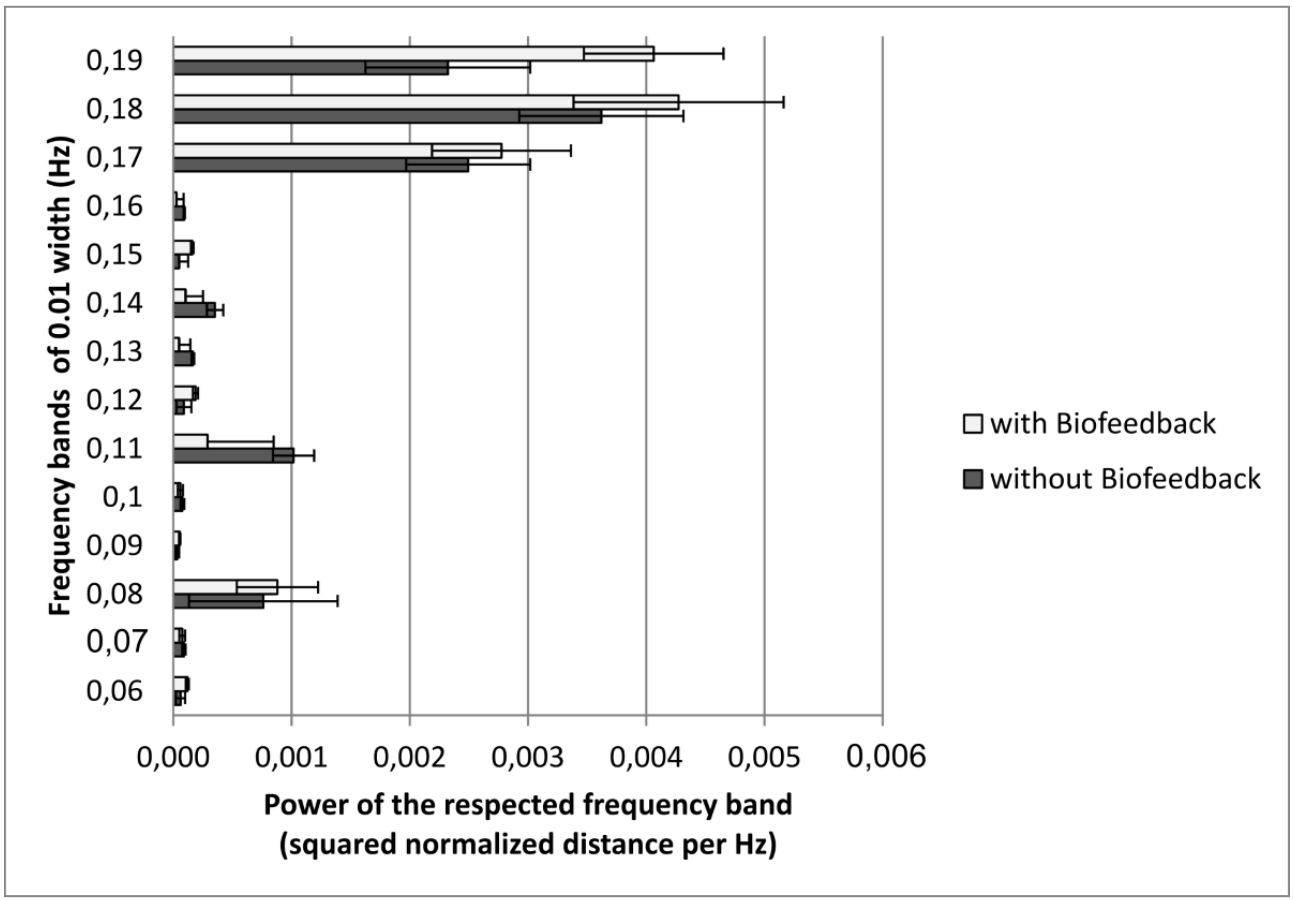


Table 3. Results of the analysis of variance for the power of frequency bands in the baseline block.

Factor	*F* (*df*)	*P* value
Frequency bands	32.21 (13,481)	<.001
Group	0.36 (1,37)	.56
Frequency bands × group	1.36 (13,481)	.18

Effects of Additional Biofeedback

In both groups, there was an increase of the spectral power in the recommended frequency (CG: t_{18}=3.47, *P*=.003; EG: t_{19}=6.12, *P*<.001) as well as in the SNR (CG: t_{18}=5.88, *P*<.001; EG: t_{19}=4.16, *P*=.001) in the training block compared with the baseline block (see Table 4 for means and standard deviations).

To reveal changes over time, 5 time blocks of 1 min each were included as repeated measurements variable to investigate group differences in breathing performance.

For the spectral power in the recommended frequency, the analysis of variance revealed neither significant main effects nor an interaction between group and time (see Table 5 for details).

For the SNR, there was a main effect of time with decreased SNR during the first minute compared with the second minute (CG: t_{18}=−2.48, *P*=.02; EG: t_{19}=−2.36, *P*=.03). There was a main effect of group but no interaction between group and time (see Table 6 and Figure 4 for details).

Comparisons for the subjective ratings of the 2 versions of the app are provided in Table 7. There was an overall trend in favor of the app without biofeedback.

Table 4. Means and SDs of power of the requested frequency band (0.09-0.11 Hz) and the signal-to-noise ratio in both groups for the 3 measurement blocks.

Statistical value	Power: control group	Power: experimental group	Signal-to-noise ratio: control group	Signal-to-noise ratio: experimental group
Baseline				
Mean	0.0000041	0.0000036	0.31	0.64
SD	0.0000067	0.0000063	0.57	1.93
Training				
Mean	0.0000234	0.0000277	9.17	12.18
SD	0.0000243	0.0000171	6.50	12.20
Post				
Mean	0.0000065	0.0000100	0.73	1.28
SD	0.0000112	0.0000159	1.29	2.59

Table 5. Results of the analysis of variance for the power of the recommended frequency in the training block.

Factor	*F* (*df*)	*P* value
Group	0.27 (1,37)	.61
Time	1.30 (4,148)	.27
Group×time	1.01 (4,148)	.41

Table 6. Results of the analysis of variance for the signal-to-noise ratio for the recommended frequency in the training block.

Factor	*F* (*df*)	*P* value
Group	4.18 (1,37)	.048
Time	3.75 (4,148)	.006
Group × time	0.78 (4,148)	.54

Figure 4. Group comparisons of signal to noise ratio (SNR) in the training block over time. SNR increases after the first minute in both groups. The analysis of variance reveals significant group differences but no interaction with time. Error bars indicate standard errors of the mean.

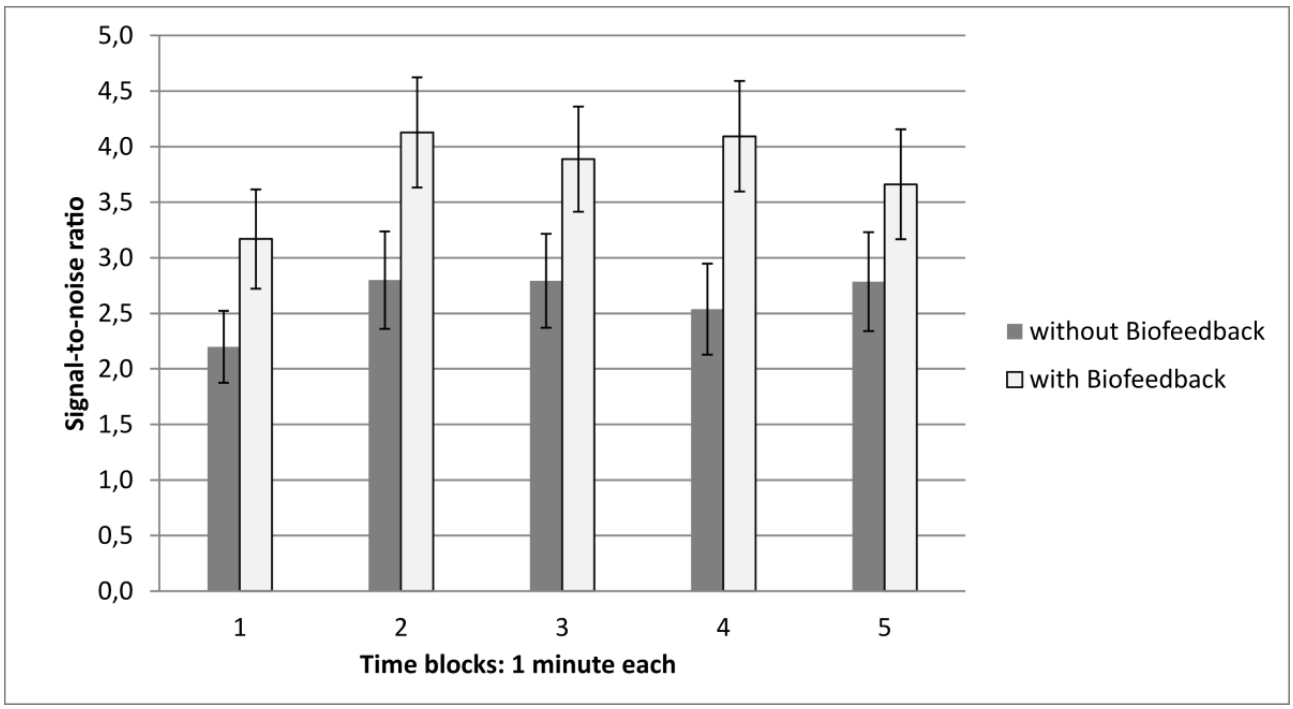


Table 7. Group comparisons of the subjective app ratings [19] with *t* tests for independent samples. Mean, SD, *t* test values, and *P* values are described for each item.

Item (1=strongly disagree, 5=strongly agree)	Mean (SD) CG[a]	Mean (SD) EG[b]	*t* (*df*)	*P* value
The app facilitates relaxation.	3.58 (1.12)	2.90 (1.17)	1.85 (37)	.07
The app is pleasant to use.	4.00 (0.94)	3.70 (0.98)	0.97 (37)	.34
It is easy to follow the app instructions.	4.74 (0.56)	4.15 (1.14)	2.06 (37)	.049
The app effectively teaches how to breathe.	4.37 (1.07)	4.25 (0.91)	0.37 (37)	.71
The app is effective in reducing stress.	3.74 (1.10)	2.90 (1.12)	2.36 (37)	.02
The app is effective in increasing attention to breathing.	4.63 (0.76)	4.55 (0.69)	0.35 (37)	.73

[a]CG: control group.

[b]EG: experimental group.

Comparison of the Post Measurement Block With the Baseline

To test whether a single 5-min training session is already enough to evoke changes in the abdominal breathing patterns toward the requested breathing frequency (0.1 Hz), we compared the baseline and the postmeasurement block with regard to the spectral power of this frequency in both groups. There were no main effect of measurement block, no effect of group, and no interaction between block and group (see Table 8 for details).

Comparable results were found for the SNR. There was no main effect of measurement block, no main effect of group, and no interaction between group and block (see Table 9 for details, see also Table 4 for mean and SD).

Table 8. Results of the analysis of variance for the power of the recommended frequency in the postmeasurement block.

Factor	*F* (*df*)	*P* value
Group	0.49 (1,37)	.49
Block	3.60 (1,37)	.07
Group × block	0.58 (1,37)	.45

Table 9. Results of the analysis of variance for the signal-to-noise ratio for the recommended frequency in the postmeasurement block.

Factor	*F* (*df*)	*P* value
Group	1.33 (1,37)	.26
Block	1.60 (1,37)	.21
Group × block	0.07 (1,37)	.80

Discussion

Principal Findings

The *Breathing-Mentor* app combines effective visualization of the instruction [19] with biofeedback on deep abdominal breathing, based on the mobile phone's accelerometers. We conducted a first pilot study with 2 versions of the app to receive the user's feedback and investigate breathing behavior during the initial 5 min of contact. To reveal possible effects of the biofeedback, both versions contained the same visual instruction, but only the full version included additional biofeedback.

Effects of the Visual Instruction

The baseline block revealed that both groups were comparable before the breathing training regarding their ability to breathe deeply and slowly with the abdomen. Breathing frequencies near the normal breathing frequency (0.2 Hz) were more prominent in both groups compared with slower frequencies. This shows that participants were rather novices for slow abdominal breathing exercises. This finding agrees with the results from the questionnaire on previous experience with breathing exercises for relaxation purposes. Although most participants were aware that abdominal breathing can be used for relaxation exercises, only few participants actually reported practicing such exercises. Thus, the participants were representative of users who could benefit from a training app for diaphragmatic breathing [19].

Indeed, both versions of Breathing-Mentor (visual instruction only and visual instruction with additional biofeedback) enabled the users to realize the requested breathing frequency of 0.1 Hz more accurately compared with the baseline, as reflected by the spectral power and the SNR. This was expected, as both conditions include the wave-based visual instruction, which has already been shown to be very effective for mobile breathing training [19]. Moreover, SNR increased in both groups after the first minute and remained at a stable level. This fast adaptation of the breathing pattern toward the instructed frequency supports the effectivity of the user interface [19] and goes in line with the high subjective ratings of the ease of use (see Table 7 for details). There is, however, no further

improvement within the 5-min training block. Moreover, the postmeasurement block revealed that breathing performance returns to the baseline performance in both groups, when the visual instructions are removed again. Both findings show that the 5-min training block is not enough to trigger transfer learning. Participants remain dependent on the app during the breathing exercise. However, the protection of the users' autonomy has been identified as an important factor in a recent stress management app [11]. Therefore, additional blocks with terminal feedback without the visual instruction might be 1 possibility to counteract dependency upon the interface and to trigger transfer learning [31].

Effects of Additional Biofeedback

The main research question of this study was, how additional biofeedback in a mobile app, as implemented in Breathing-Mentor (see the Methods section for details), influences people's ability to follow the visual breathing instruction and their subjective usage experience.

Although the spectral power of the desired frequency band did not result in significant group differences, the SNR was higher for the biofeedback training group (see the Results section for details). This means that abdominal breathing at the desired frequency was not more prominent compared with the CG without biofeedback, but the occurrence of undesired frequency bands was reduced for the biofeedback group, resulting in enhanced SNR values. These findings support the effectiveness of the additional biofeedback on breathing behavior.

This benefit in performance was, however, combined with lower subjective ratings regarding the effectiveness of the app to reduce stress and ease with which app instructions could be followed for the biofeedback training. This result could be a consequence of increased cognitive workload and attention resources that are required to interpret and modulate the biofeedback graph [32]. Nevertheless, ratings for ease of use and task difficulty were high in both groups. This suggests that workload was not excessive during the training. The physiological stress level and cognitive processing during the training should be addressed more deeply in future studies, as they are expected to change with proficiency level. The role of the relaxation level could be addressed by including additional objective psychophysiological parameters [25-27] to complete the subjective ratings. Cognitive measures could also be targeted with psychophysiological parameters from electroencephalography [33] or eye tracking [34].

Limitations and Outlook

To summarize, Breathing-Mentor seems to be a useful tool to teach specific abdominal breathing patterns. An immediate improvement of the user's relaxation state should, however, not be expected, especially for persons who are inexperienced with breathing tasks. With further experience, tools such as the BellyBio Interactive Breathing app might be more useful, as the auditory feedback allows to close the eyes and to focus more intensively on the body, which are facilitating factors for deep relaxation [35]. Such auditory tools might also be useful for people with age-related visual impairments. A multimodal approach could be considered to extend the app to older people.

Finally, the frame that is used to hold the mobile phone at a stable position is 1 limitation factor. Although there were no user complaints concerning the usability of this approach, the correct positioning of the mobile phone was guaranteed by the calibrating procedure and the principal investigator in this study. Other fixing solutions should be considered for everyday use.

Conclusions

In summary, it should be noted that participants were rapidly able to adjust their breathing pattern to the instruction (within 1 min). This result supports the feasibility and usefulness of biofeedback in mobile breathing apps based on the mobile phone's accelerometers, especially for people who are unfamiliar with breathing techniques. Immediate effects on the user's relaxation state should, however, not be expected.

Acknowledgments

The authors would like to thank Sabrina Defren and Nils Beese for supporting the data collection. The junior research group wearHEALTH is funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, reference number: 16SV7115).

Conflicts of Interest

None declared.

References

1. Stansfeld S, Candy B. Psychosocial work environment and mental health--a meta-analytic review. Scand J Work Environ Health 2006 Dec;32(6):443-462 [FREE Full text] [doi: 10.5271/sjweh.1050] [Medline: 17173201]
2. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. J Am Med Assoc 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
3. Lazarus RS. The psychology of stress and coping. Issues Ment Health Nurs 2010 Nov 26;7(1-4):399-418. [doi: 10.3109/01612848509009463]
4. Zhu B, Hedman A, Feng S, Li H, Osika W. Designing, prototyping and evaluating digital mindfulness applications: a case study of mindful breathing for stress reduction. J Med Internet Res 2017 Jun 14;19(6):e197 [FREE Full text] [doi: 10.2196/jmir.6955] [Medline: 28615157]

5. Christmann CA, Hoffmann A, Bleser G. Stress management apps with regard to emotion-focused coping and behavior change techniques: a content analysis. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e22 [FREE Full text] [doi: 10.2196/mhealth.6471] [Medline: 28232299]
6. Meuret AE, Wilhelm FH, Roth WT. Respiratory biofeedback-assisted therapy in panic disorder. Behav Modif 2001 Sep;25(4):584-605. [doi: 10.1177/0145445501254006] [Medline: 11530717]
7. Blödt S, Pach D, Roll S, Witt CM. Effectiveness of app-based relaxation for patients with chronic low back pain (Relaxback) and chronic neck pain (Relaxneck): study protocol for two randomized pragmatic trials. Trials 2014 Dec 15;15:490 [FREE Full text] [doi: 10.1186/1745-6215-15-490] [Medline: 25511185]
8. Hoffmann A, Faust-Christmann CA, Zolynski G, Bleser G. Assessing the quality of a gamified pain management app: an expert study (epub ahead of print). JMIR Mhealth Uhealth 2019. [doi: 10.2196/13170]
9. 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]
10. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
11. Christmann CA, Hoffmann A, Zolynski G, Bleser G. Stress-Mentor: Linking Gamification and Behavior Change Theory in a Stress Management Application. In: Proceedings of the International Conference on Human-Computer Interaction. 2018 Presented at: HCI'18; July 15-20, 2018; Las Vegas, NV, USA p. 387-393. [doi: 10.1007/978-3-319-92279-9_52]
12. Christmann CA, Zolynski G, Hoffmann A, Bleser G. Towards More Interactive Stress-Related Self-Monitoring Tools to Improve Quality of Life. In: Proceedings of the Congress of the International Ergonomics Association. 2018 Presented at: IEA'18; August 26-30, 2018; Florence, Italy p. 121-130. [doi: 10.1007/978-3-319-96098-2_17]
13. Luxton DD, Hansen RN, Stanfill K. Mobile app self-care versus in-office care for stress reduction: a cost minimization analysis. J Telemed Telecare 2014 Dec;20(8):431-435. [doi: 10.1177/1357633X14555616] [Medline: 25316037]
14. Chittaro L, Vianello A. Evaluation of a mobile mindfulness app distributed through on-line stores: a 4-week study. Int J Hum-Comput St 2016 Feb;86:63-80. [doi: 10.1016/j.ijhcs.2015.09.004]
15. Ebben MR, Kurbatov V, Pollak CP. Moderating laboratory adaptation with the use of a heart-rate variability biofeedback device (StressEraser). Appl Psychophysiol Biofeedback 2009 Dec;34(4):245-249. [doi: 10.1007/s10484-009-9086-1] [Medline: 19418214]
16. Hazlett-Stevens H, Craske MG. Breathing retraining and diaphragmatic breathing techniques. In: O'Donohue WT, Fisher JE, editors. Cognitive Behavior Therapy: Applying Empirically Supported Techniques in Your Practice. Second Edition. Chichester, New York: John Wiley & Sons; 2009:68-74.
17. 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 [FREE Full text] [doi: 10.2147/PRBM.S15375] [Medline: 22114534]
18. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
19. Chittaro L, Sioni R. Evaluating mobile apps for breathing training: the effectiveness of visualization. Comput Human Behav 2014 Nov;40:56-63. [doi: 10.1016/j.chb.2014.07.049]
20. Liu GZ, Huang BY, Wang L. A wearable respiratory biofeedback system based on generalized body sensor network. Telemed J E Health 2011 Jun;17(5):348-357 [FREE Full text] [doi: 10.1089/tmj.2010.0182] [Medline: 21545293]
21. Mitchell E, Coyle S, O'Connor NE, Diamond D, Ward T. Breathing Feedback System with Wearable Textile Sensors. In: Proceedings of the International Conference on Body Sensor Networks. 2010 Presented at: BSN'10; June 7-9, 2010; Singapore p. 56-61. [doi: 10.1109/BSN.2010.31]
22. Al Rihawi RG, Ahmed B, Gutierrez-Osuna R. Dodging Stress With a Personalized Biofeedback Game. In: Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play. 2014 Presented at: CHI PLAY'14; October 19-21, 2014; Toronto, Ontario, Canada p. 399-400. [doi: 10.1145/2658537.2661310]
23. Zafar MA, Ahmed B, Gutierrez-Osuna R. Playing With and Without Biofeedback. In: Proceedings of the 5th International Conference on Serious Games and Applications for Health. 2017 Presented at: SeGAH'17; April 2-4, 2017; Perth, WA, Australia p. 1-7. [doi: 10.1109/SeGAH.2017.7939272]
24. Parnandi A, Ahmed B, Shipp E, Gutierrez-Osuna R. Chill-Out: Relaxation Training through Respiratory Biofeedback in a Mobile Casual Game. In: Proceedings of the International Conference on Mobile Computing, Applications, and Services. 2013 Presented at: MobiCASE'13; November 7-8, 2013; Paris, France p. 252-260. [doi: 10.1007/978-3-319-05452-0_18]
25. Dillon A, Kelly M, Robertson IH, Robertson DA. Smartphone applications utilizing biofeedback can aid stress reduction. Front Psychol 2016;7:832 [FREE Full text] [doi: 10.3389/fpsyg.2016.00832] [Medline: 27378963]
26. Morrow AM, Burton KL, Watanabe MM, Cloyd BH, Khut GP. Developing brighthearts: a pediatric biofeedback-mediated relaxation app to manage procedural pain and anxiety. Pain Pract 2018 Jul;18(6):698-708. [doi: 10.1111/papr.12655] [Medline: 29080245]
27. de Ridder B, van Rompaey B, Kampen JK, Haine S, Dilles T. Smartphone apps using photoplethysmography for heart rate monitoring: meta-analysis. JMIR Cardio 2018 Feb 27;2(1):e4. [doi: 10.2196/cardio.8802]

28. Hailstone J, Kilding AE. Reliability and validity of the Zephyr™ BioHarness™ to measure respiratory responses to exercise. Meas Phys Educ Exerc Sci 2011 Oct;15(4):293-300. [doi: 10.1080/1091367X.2011.615671]
29. Villar R, Beltrame T, Hughson RL. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. Appl Physiol Nutr Metab 2015 Oct;40(10):1019-1024. [doi: 10.1139/apnm-2015-0140] [Medline: 26360814]
30. Holt M, Yule B, Jackson D, Zhu M, Moraveji N. Ambulatory Monitoring of Respiratory Effort Using a Clothing-Adhered Biosensor. In: Proceedings of the International Symposium on Medical Measurements and Applications. 2018 Presented at: MeMeA'18; June 11-13, 2018; Rome, Italy p. 1-6. [doi: 10.1109/MeMeA.2018.8438678]
31. Sigrist R, Rauter G, Riener R, Wolf P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychon Bull Rev 2013 Feb;20(1):21-53. [doi: 10.3758/s13423-012-0333-8] [Medline: 23132605]
32. Sweller J. Cognitive load theory. In: Mestre J, Ross B, editors. Cognition in Education. Volume 55. Amsterdam, Netherlands: Elsevier; 2011:37-76.
33. Antonenko P, Paas F, Grabner R, van Gog T. Using electroencephalography to measure cognitive load. Educ Psychol Rev 2010 Apr 29;22(4):425-438. [doi: 10.1007/s10648-010-9130-y]
34. Lai ML, Tsai MJ, Yang FY, Hsu CY, Liu TC, Lee SW, et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ Res Rev 2013 Dec;10(1):90-115 [FREE Full text] [doi: 10.1016/j.edurev.2013.10.001]
35. Benson H. The relaxation response: its subjective and objective historical precedents and physiology. Trends Neurosci 1983 Jan;6:281-284. [doi: 10.1016/0166-2236(83)90120-0]

Abbreviations

CG: control group
EG: experimental group
SNR: signal-to-noise ratio

Edited by G Eysenbach; submitted 14.02.19; peer-reviewed by K Paldan, J Lortz, J Weerdmeester; comments to author 28.04.19; revised version received 13.06.19; accepted 29.06.19; published 30.09.19.

Please cite as:
Faust-Christmann CA, Taetz B, Zolynski G, Zimmermann T, Bleser G
A Biofeedback App to Instruct Abdominal Breathing (Breathing-Mentor): Pilot Experiment
JMIR Mhealth Uhealth 2019;7(9):e13703
URL: http://mhealth.jmir.org/2005/9/e13703/
doi:10.2196/13703
PMID:31573919

Original Paper

Mobile Apps for the Care Management of Chronic Kidney and End-Stage Renal Diseases: Systematic Search in App Stores and Evaluation

Abu Bakkar Siddique[1*], MPP; Mary Krebs[2*], MS; Sarai Alvarez[1*], BSc; Iris Greenspan[1*], BSc; Amit Patel[2*], PhD; Julianna Kinsolving[1*], BS; Naoru Koizumi[1*], PhD

[1]Schar School of Policy and Government, George Mason University, Arlington, VA, United States

[2]University of Massachusetts Boston, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Abu Bakkar Siddique, MPP
Schar School of Policy and Government
George Mason University
Founders Hall
3351 Fairfax Dr, Suite 526
Arlington, VA, 22201
United States
Phone: 1 5712745328
Email: anirbanju36@gmail.com

Abstract

Background: Numerous free and low-cost mobile apps for the care management of kidney disease have become available in recent years. Although these appear to be promising tools, they have not been evaluated comparatively based on standard mobile app metrics, and thus, limited evidence is available regarding their efficacy. This study systematically cataloged and assessed mobile apps designed to assist medication compliance and nutrition tracking that are useful to the chronic kidney disease (CKD) and the end-stage renal disease (ESRD) patients who are on dialysis.

Objective: The objective of this study was to comprehensively evaluate mobile apps used for medication compliance and nutrition tracking for possible use by CKD and ESRD patients.

Methods: A systematic review framework was applied to the search, screening, and assessment of apps identified and downloaded from the iOS and Android app stores. We selected apps using 13 relevant search terms, narrowed down based on a set of inclusion and exclusion criteria, and then used the Mobile App Rating Scale (MARS), a widely adopted app evaluation tool to assess the effectiveness of apps. The internal consistency and interrater reliability were tested using Cronbach alpha and interclass correlation coefficients (ICCs), respectively.

Results: The MARS total score had excellent internal consistency (Cronbach alpha=.90) and a moderate level of interrater reliability (2-way mixed ICC 0.65). Overall, 11 out of the 12 reviewed apps met the minimum acceptable score of 3.0 in MARS rating. The 3 apps with the highest combined scores were *My Kidneys, My Health Handbook* (MARS=4.68); *My Food Coach* (MARS=4.48); and *National Kidney Foundation Malaysia* (MARS=4.20). The study identified 2 general weaknesses in the existing apps: the apps fell short of accommodating advanced interactive features such as providing motivational feedback and promoting family member and caregiver participations in the app utilization.

Conclusions: The MARS rating system performed well in the app evaluation. The 3 highest ranked apps scored consistently high across the 5 dimensions specified in MARS. These apps were developed in collaboration with reputable organizations and field experts, demonstrating the importance of expert guidance in developing medical apps.

(JMIR Mhealth Uhealth 2019;7(9):e12604) doi:10.2196/12604

KEYWORDS

mobile apps; mhealth; digital health; kidney patient; nutrition tracking; systematic assessment; Mobile App Rating Scale

Introduction

Background

The prevalence of chronic kidney disease (CKD), particularly end-stage renal disease (ESRD) or the final stage (stage 6) of CKD, has been growing steadily in the United States for the last decade [1]. The increase is primarily attributed to the rising health conditions linked to CKD, such as diabetes, hypertension, and obesity, as well as aging [1]. The most recent statistic reports that CKD affects 14% of the US adult population compared with, for instance, diabetes mellitus (DM), which affects 12% of the population [2]. The treatment costs for CKD, particularly for ESRD, have been a large driver of overall health care spending. The 3 most common comorbidities consisting of CKD, DM, and congestive heart failure together share the highest expenditure for Medicare reimbursement [1]. With the expected further increase in health care costs, improvement in the care for CKD and ESRD patients and the prevention of disease progression have been considered one of the highest priorities in the US health care [3].

Slowing the disease progression requires significant personal involvement for CKD/ESRD patients. The patients face complex recommendations on medication adherence, lifestyle modification, and nutritional adaptation [3]. Meeting specific nutrition guidelines is particularly challenging for ESRD patients on dialysis [4,5]. The burdens of complying with these guidelines are considerable, not only for the patients themselves but also for their families and caregivers (ibid). Previous literature suggests that enhanced knowledge could improve self-management skills in chronic disease [5-7]. However, CKD/ESRD patients are not often satisfied with their actual ability to connect with their health care providers and are mostly unaware of their diagnoses and the implications [8,9].

Objective

Information technology (IT) tools for monitoring, training, and self-management have been identified as an effective tool to empower patients [4]. The development of IT tools provides patients with access to numerous apps and portals for health information. A vast array of medical reference materials is available to patients through the internet and mobile apps, offering them a better understanding of their diseases and best practices. These apps can not only reduce costs and burdens on others but can also assist in tracking diet and nutrition, recommend healthy foods nearby, supplement medical intervention through drug information, identify pills, check drug interactions, and record personal medication. They can also estimate kidney function, provide diagnostic tests and information on disease signs and symptoms, function as medical calculators, and help manage the progression of CKD. Although these appear to be promising tools, physicians and patients are often overwhelmed with the profusion of these low-cost technologies, which limits their utilization of such innovations [10]. Concurrently, very few apps have a Food and Drug Administration (FDA) clearance or any clinical validation. Thus, not much is known about the effectiveness of these apps, especially those aiming at managing a disease or condition (ibid).

Bailey et al [11] systematically reviewed mobile apps available to patients to support outpatient medication self-management and found that hundreds of apps exist in the marketplace with a variety of quality, content, and functions. They recommended that determining optimal capabilities and clinical benefits as well as evaluating the utility of these existing mobile apps are necessary. Although there are studies that assessed the effectiveness of mobile apps supporting DM [12,13], mental health [14], bipolar disorder [15], suicide prevention [16], and asthma [17], those that support dialysis patients have not yet been assessed. Hence, no evidence is currently available regarding the effectiveness of these mobile apps that solely support dialysis patients who indeed follow more stringent diet than others. To fill this knowledge gap, this study performed a systematic review of existing mobile apps supporting CKD/ESRD patients who are on dialysis.

Methods

App Search Strategy

A team of reviewers consisting of 3 undergraduate students, a doctoral candidate, and 2 faculty members downloaded the apps and tested the usability of the apps between August 2016 and September 2017. The systematic review methodological framework was applied to the search, screening, and assessment of health-related mobile apps, except for a few instances where the guidelines are not applicable for app reviews.

For the search, we defined search strings developed specifically for nutrition and medication tracking for CKD and ESRD patients. The strings included “kidney” or “kidney care” or “kidney transplant” or “kidney nutrition” or “renal nutrition tracking” or “dialysis” or “dialysis diet” or “renal diet” or “CKD” or “kidney medication” or “kidney medication tracking” or “kidney water tracking” or “kidney transplant medication.” These search terms were derived through an iterative review process encompassing interactions with the app stores, expert physician inputs, and team consensus over the course of several months. This strategy incorporates the medical phrase synonyms for renal failure, as well as the layperson alternatives to the identified terms, as the affected population of interest may not use the technical terms of the kidney disorder.

During a 1-week window in July 2016, the 13 search terms were used by 3 reviewers to identify publicly accessible apps supportive of nutrition and medication tracking for renal patients. Each of the 3 reviewers utilized different but commonly used devices: (1) an iOS iPhone 5 (Apple Inc), (2) an Android Optimus (Samsung), and (3) a first-generation iPad (Apple Inc). The apps considered were those displayed by the US Google Play Store for Android-based and Apple App Store for iOS-based devices. One member of the team screened the Google Play Store, whereas 2 members screened the Apple App Store to compile cursory descriptions of available apps. The apps were also initially searched through various sites including the FDA medical device website and the *mHealth Database* developed by United States Agency for International Development and African Strategies for Health to verify that both the Google Play Store and Apple App Store were adequate sources for apps and that we do not need to expand our search

to other databases [18]. Our final search was performed only on Google Play Store and Apple App Store.

App Selection Strategy

Apps were selected based on the information included in app name, publisher's description, and price. Apps were selected for inclusion based on the following criteria: (1) available in English language, (2) free of charge, (3) smartphone app, (4) available for download from the official app stores of Apple or Google, (5) targets patients with kidney disease, and (6) targets patients with renal failure (as intended by the publisher). Duplicate apps extracted on the same device platform were disqualified, whereas different versions of the same app that appeared across platforms were retained for comparison [15]. The remaining apps were then screened for inclusion criteria, and those apps that do not target patients with kidney disease (CKD/ESRD, renal failure, dialysis, etc) were removed.

The team then met to discuss the apps that require additional scrutiny and to jointly determine the final set of apps to be downloaded and installed. This process removed apps with too few features to be considered for an effective care management system. Additional apps were disqualified subsequent to the installation based on several exclusion criteria, as specified in the Results section. Although rankings of apps in stores are in constant flux and updated and rated by app stores and their users [19], the apps selected by our team reflect those most visible to users seeking assistance at the time of selection. This approach is consistent with a representative user experience where the renal patient is subjected to unsystematic availability of apps for support.

App Data Analysis

This study used a rigorous assessment framework, known as the Mobile App Rating Scale (MARS), developed by researchers at the Queensland University. Broadly speaking, MARS developed by Stoyanov et al [20] offers a promising evaluation scheme for classifying and rating the quality of mobile health apps. MARS is built upon the existing body of scholarship between January 2000 and January 2013 and comprises 4 broad categories of *objective quality* criteria, including app *engagement*, *functionality*, *aesthetics*, and *information*, and 1 *subjectivequality* scale based on the 20 to 23 items of MARS subcriteria. Although *engagement* addressed questions such as "is it fun, interesting, customizable, interactive, well-targeted to the audience?," *functionality* measured app functioning level, reflecting how easy it is to learn or navigate, flow logic, and gestural design of apps. The score of *aesthetics* assessed the apps' graphic design, such as overall visual appeal, color scheme, and consistent style. The quality and quantity of information, credibility of the sources of information, evidence, etc, built the score of *information* criterion. The app's *subjectivequality* assessed overall satisfaction level and whether the app is worth recommending, stimulates repeat use, etc. MARS has demonstrated excellent internal consistency and interrater reliability (ibid).

To evaluate the final set of apps, 3 reviewers performed scoring for the apps on a 5-point scale, producing a comprehensive final mean score for each app. The internal consistency of the MARS quality subscales and total quality score were calculated using Cronbach alpha. This alpha coefficient indicates the degree (correlations) to which items measuring the same general construct produce similar scores. Interclass correlation coefficients (ICCs) were calculated using 2-way mixed effects for agreement [21]. There were 6 apps that were available on more than 1 device. We used independent scores by each reviewer to calculate ICC. For those 6 apps, reviewers had disagreements in their individual scores. However, reviewers reached to a consensus and produced an agreed-upon score after extensive deliberations for each scale for these 6 apps.

Results

Overall Assessment

Figure 1 presents the flow chart of the app selection process. The initial search based on the 13 search terms captured 431 apps. Of these apps, 235 were removed because of being duplicates, which left 196 apps for further evaluation. An additional 162 apps were removed because of not targeting CKD/ESRD patients. Of the remaining 34 apps, a final pool for assessment was determined through discussion and consensus among the team members. This excluded an additional 18 apps that had fairly limited functions. Most of these apps were mere calculators of some sort (estimating calorie, water, and phosphorous intakes), whereas other apps were simple appointment reminders or goal trackers. The remaining 16 apps were subsequently downloaded for comprehensive evaluation. Additional scrutiny of the downloaded apps removed 4 apps because of veiled matters of language, invitation for access, and concealed monetary motivation by publishers (as detailed in Multimedia Appendix 1). Consequently, the final set of 12 apps went through the final evaluation using MARS.

The initial assessment of the 12 apps revealed that their interface designs typically include data entry, goal and reminder settings, and graphing and analytics of achievements. In terms of the functions, these apps typically included calorie intake calculation, goal setting, reminders, access to social networks, and game elements such as rewards or competition among users. The MARS total score had excellent internal consistency (Cronbach alpha=.90) and was highly correlated with the MARS star rating item (#23), r (12)=0.88; P<.001. Internal consistencies of the MARS subscales were also very high except for 1 subscale (Cronbach alpha=.40-.83; median .70). Independent ratings on the overall MARS total score of the 6 apps by 2 raters demonstrated moderate level of interrater reliability (2-way mixed ICC 0.65; 95% CI −4.45 to 0.97), and interrater reliabilities of subscales were fair (ICC −0.61 to 0.88; median .48). Detailed item and subscale statistics are presented in Multimedia Appendix 2.

Table 1 presents the mean and median scores of the 12 apps in each of the 5 MARS dimensions. The apps, taken together, scored higher than the minimum acceptable score of 3.0 in all 5 dimensions.

Figure 1. App selection flow diagram.

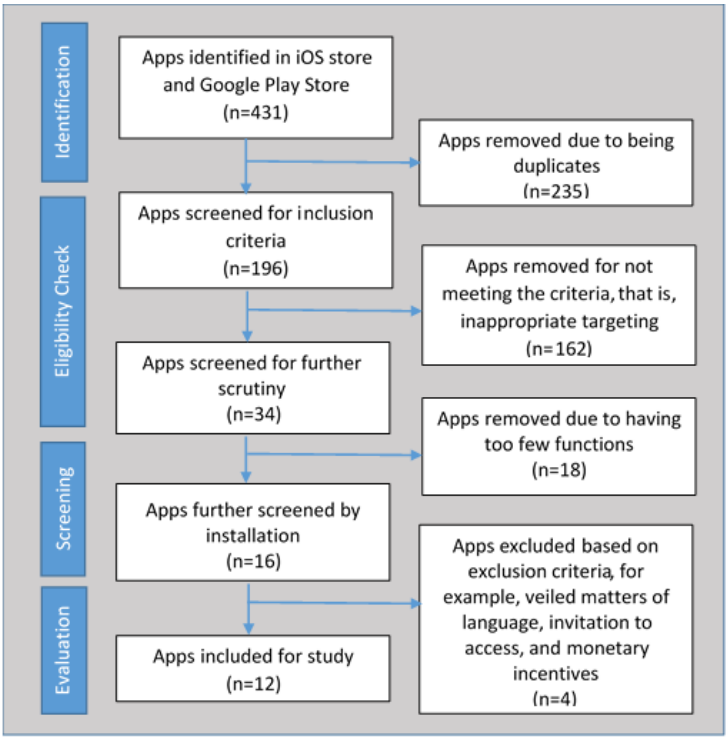


Table 1. The Mobile App Rating Scale's app quality ratings (1-5).

Criteria	Mean (SD)	Median (interquartile range)	Minimum and maximum
Engagement	3.35 (0.78)	3.40 (1.08)	1.6 and 4.3
Functionality	4.29 (0.63)	4.45 (0.85)	2.9 and 5.0
Aesthetics	3.85 (0.63)	3.90 (0.93)	2.7 and 5.0
Information	3.88 (0.84)	4.25 (1.43)	2.5 and 4.8
Subjective quality	3.60 (0.87)	3.60 (0.75)	1.9 and 5.0

The apps scored relatively low in the engagement dimension primarily because of the lack of interactive feature (mean 3.35; median 3.40). The apps scored high in the functionality dimension with the mean and median scores of 4.29 and 4.45, respectively. Here, the scores were consistently high across the subcriteria except for a few apps that scored low in all dimensions.

Table 2 exhibits the ranking of the reviewed apps and their individual comprehensive mean scores. The comprehensive mean scores ranged from as low as 2.98 to as high as 4.68. The median of the 12 mean scores and the interquartile range (IQR) were 3.70 and 0.78, respectively. All apps except for *Phosphorus Tracker* (MARS=2.98) met the minimum acceptable score of 3.0. The 3 apps receiving the highest combined scores were *My Kidneys, My Health Handbook* (MARS=4.68); *My Food Coach* (MARS=4.48); and *National Kidney Foundation Malaysia* (MARS=4.20). *H2O Overload*, the app ranked fourth best (MARS=4.18), was ranked very close to *National Kidney Foundation Malaysia*. *Phosphorus Tracker* and *Wholesome* consistently demonstrated poor scores across most criteria. In between these extremes, *Care After Kidney Transplant*, *CKD Go!*, *AAKP myHealth Nutrition Guide*, *Kidney APPetite*, Oxalator, and *D-Track - Dialysis Tracker* had medium effectiveness.

Table 2. Mobile apps and the Mobile App Rating Scale's comprehensive scores.

Ranking	Apps' name	Mobile apps rating scale protocol[a], mean	Subjective quality score[b]	App-specific score[c] (perceived impact)
1	My Kidneys, My Health handbook	4.68	5.00	4.20
2	My Food Coach	4.48	4.88	4.20
3	National Kidney Foundation Malaysia	4.20	3.67	4.00
4	H2O Overload	4.18	5.00	4.00
5	CKD Go!	4.10	3.38	4.10
6	Care After Kidney Transplant	4.08	3.75	4.70
7	AAKP myHealth Nutrition Guide	3.63	4.00	2.80
8	Kidney APPetite	3.63	3.25	3.40
9	Oxalator	3.58	2.50	1.90
10	D-Track - Dialysis Tracker	3.55	3.50	3.10
11	Wholesome	3.25	3.25	2.00
12	Phosphorus Tracker	2.98	1.88	2.20

[a]The mean of the mobile apps rating scale's protocol scores for the 4 criteria: (1) engagement, (2) functionality, (3) aesthetics, and (4) information. It includes items 1 to 19.

[b]The subjective quality score includes items 20 to 23.

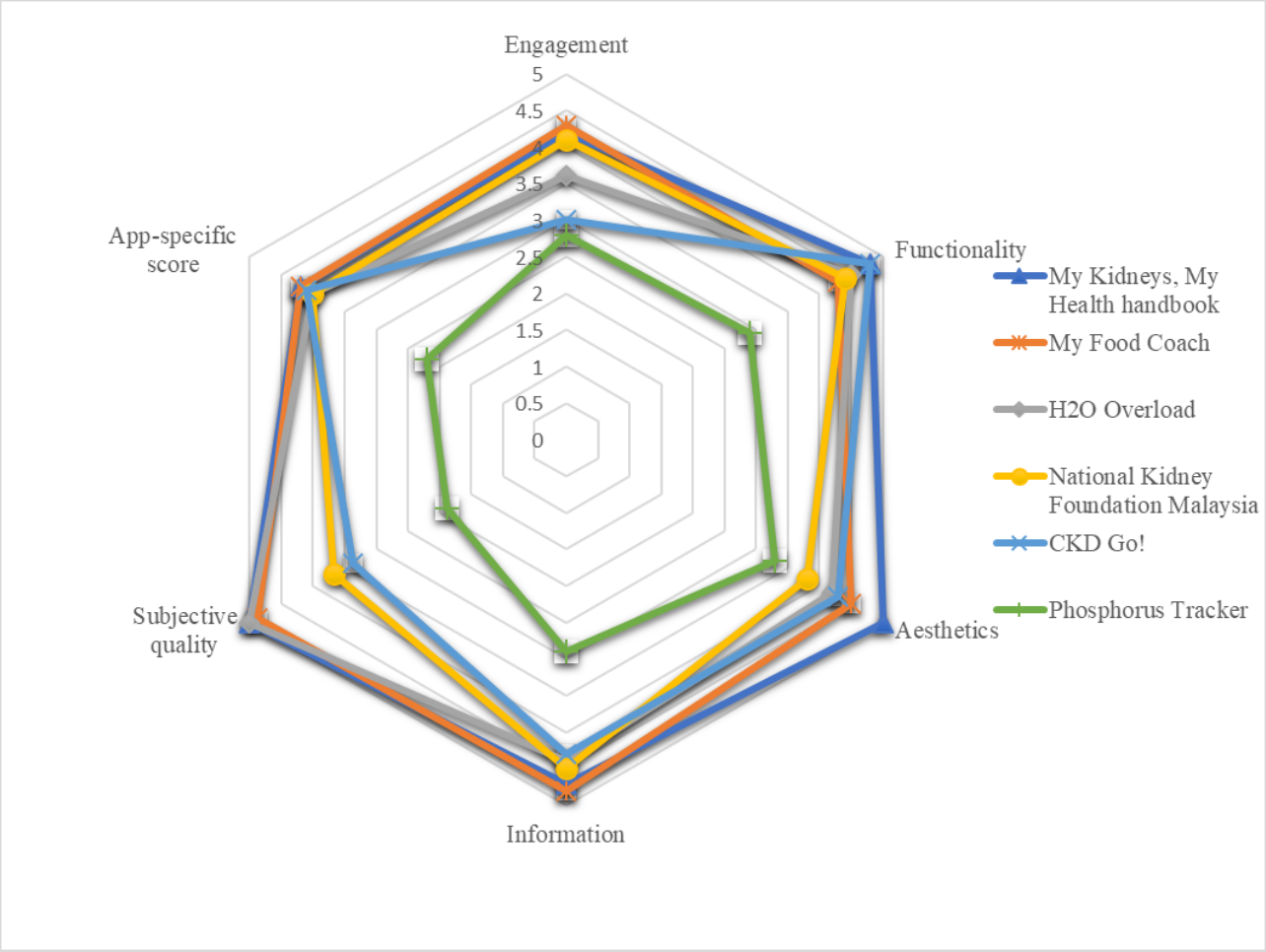
[c]App-specific score includes the scores for awareness, knowledge, attitudes, intention to change, help seeking, and behavior change.

Individual App Assessments

Figure 2 displays the scores of the top 5 and the worst apps in terms of their assessment criteria. The radar chart demonstrates how these individual apps scored in each criterion such as app engagement, functionality, aesthetics, information, subjective quality, and app-specific score. *Phosphorus Tracker* was ranked far below compared with the top 5 apps across all criteria, but especially in terms of subjective quality and app-specific score. The app was severely penalized for broken links and technical difficulties, that is, crashes and bugs. It is worth noting that the other 5 apps presented in the chart were all developed in collaboration with reputable expert organizations such as the National Kidney Foundation (NKF) in the United States (*My Food Coach*) and in Malaysia (*National Kidney Foundation Malaysia*) as well as the National Health and Medical Research Council (NHMRC) in Australia (the other 3 apps). The high information scores of these apps can be partially attributable to the involvements of these organizations, as credibility and legitimacy of the app source is a subcriterion in the assessment of the MARS information dimension.

The top 2 apps (*My Kidneys, My Health Handbook* and *My Food Coach*) demonstrated similar scores for all criteria, except that *My Kidneys, My Health Handbook* was slightly superior in terms of functionality and aesthetics. The reviewers noted that the app stood out aesthetically with the quality, color coordination, and resolution of the graphics as well as the stylistically consistent interface. The app was also highly interactive and allowed direct access to various support groups. *My Food Coach* received a high score in engagement primarily because of the GPS locator for restaurants and their menus. Similar to *My Kidneys, My Health Handbook*, it was also highly interactive, supporting access to a registered dietitian. *H2O Overload* was considered superior to *My Food Coach* in terms of functionality, whereas it was inferior in all other areas, but especially in terms of engagement and subjective quality. The reviewers noted that the app was memorable for allowing multiple functions such as appointment entry/reminder, medication entry, notepad for questions to physicians, and email to physicians with progression graphs (blood pressure, weight, and fluid intakes). *National Kidney Foundation Malaysia* generally scored similar to *H2O Overload* but stood out for its highly interactive nature, allowing users to set specific and measurable goals and providing feedback. This resulted in its high engagement score. *Care After Kidney Transplant* demonstrated inconsistent quality across criteria: although the app received high scores for functionality, information, and app-specific criteria, engagement and subjective quality criteria were deemed poor. The app excelled in terms of simple and intuitive functionality, links to abundant information, and overall sense of professionalism, but it lacked the customization and interactive features, which affected the engagement score.

Figure 2. Radar chart of the top 5 and the worst ranked apps.



Discussion

Principal Findings

This study performed a systematic review of the 12 apps that assist CKD and ESRD patients with their care management. To our knowledge, this is the first study that evaluated apps supporting CKD/ESRD care. The evaluation of these apps was performed using MARS. MARS has been successfully applied in recent years to evaluate apps that assist mindfulness development [22], heart failure symptom monitoring and self-management [23], weight management [24], palliative care among pediatric patients [25], epilepsy self-management [26], and drug-drug interaction checking [27]. Our evaluation demonstrated that the overall scores of our apps were comparable with those of apps in the other fields, with the median score of 3.70 (IQR 0.78). All of our apps exceeded the minimum acceptable score of 3.0, except for 1 app (*Phosphorus Tracker*). Overall, the reviewers observed that those apps scoring high (or low) in 1 dimension in the MARS tend to score high (or low) in other dimensions as well. Previous app reviews using MARS note correlations between MARS scores in several dimensions [27]. Such correlations are considered to be particularly evident between the aesthetics and engagement dimensions [28-30]. Although the tendency was also observed in our review, it appeared that the overall similarities in the scores across dimensions were more attributable to the overall professionalism (or lack thereof) of the app developers. The only exception was *Care After Kidney Transplant*, which scored relatively low in the engagement dimension because of the limited customization and interactive features but scored high in all other dimensions.

Most apps had a combination of the functions that support self-management such as appointment/medication reminder and water/weight/phosphorus calculation and monitoring. However, none of these apps incorporated more advanced support functions such as providing motivational feedback to their progress or response to the measured water/weight/phosphorus levels. The relative unavailability of such functions is also noted in previous studies [23,24,26]. Furthermore, the apps reviewed in this study solely focused on CKD/ESRD patients and not their caregivers, and there was no explicit involvement of family members or caregivers in the use of the apps. Apps that promote participation of caregivers/parents are available and are known to be effective in pediatric palliative care [25]. Given the considerable roles played by family members and caregivers in the CKD/ESRD care management, more participatory approach in the design of the apps would be beneficial.

Our findings also indicate that guidelines for app developers are needed. Currently, there are limited resources and information available for the developers to refer to as they develop apps for health services. A large body of the free or low-cost medical mobile apps was made available by unknown

publishers without the participation or inputs from reputable domain experts. For instance, 1 app claimed that it could heal kidney problems with sound frequency therapy. In contrast, all high-ranked apps were developed in collaboration with the reputable organizations/field experts such as the NKF in the United States and Malaysia as well as the NHMRC in Australia.

Several observations were made on the app quality and the performance of devices and platforms used to find appropriate apps. Each device (ie, iOS iPhone 5, an Android Optimus smartphone, and a first-generation iPad) identified a varied number of apps, with the first-generation iPad producing the least number of kidney care–related apps and the Android Optimus smartphone producing the largest number of kidney care–related apps. Regarding the platform, reviewers unanimously agreed that the Apple store produced a robust and narrowly refined selection of apps, whereas finding apps in the Google Play store was numerically overwhelming. In particular, irrelevant gaming and entertainment apps were included in the Google Play search. For instance, the app "Crazy XMas Santa Doctor Mania" was included because the description stated, "Be the crazy surgeon doctor to rescue Christmas Santa from… kidney problem…" Thousands of similar apps were captured initially during the Android search. This observation suggests that the Google Play store search-inclusion algorithm could be improved to aid a kidney failure patient in finding an appropriate quality app. In terms of the app quality, MARS scores exhibited no striking differences in terms of quality between apps downloaded from the iOS versus Android app stores.

Finally, there are limited monitoring and regulatory authorities to oversee the fraudulent apps. In the United States, the FDA provides guidance on which mobile apps they regulate and how to regulate. However, the monitoring takes a risk-based approach and is applied only to those apps that meet the regulatory definition of *device*, which operates as an accessary to a regulated medical device. As such, an authority providing guidelines or performing systematic evaluation of apps is warranted.

Limitations and Future Research

Although this study performed a consistent evaluation of the apps based on the consensus among the team members, an internal critique of the MARS rating scheme offers limitations for consideration. Self-reported limitations by the Queensland developers identify a lack of peer-reviewed literature on which to base the evaluation. There are few other tools to evaluate apps such as App Chronic Disease Checklist and Royal College of Physicians checklist. Cross-checking using these available tools may provide different findings from MARS. Such work could also provide stronger evidence to support CKD/ESRD patients and their physicians.

The interrater reliability of the reviews was fair but differed significantly across subscales. The interrater reliabilities for the engagement and functionality subscales were notably low despite the fact that all the reviewers received a Web-based MARS tool training a priori, followed by a team discussion utilizing a consensus approach to discrepancies in rating. The low interrater reliabilities for engagement and functionality may reflect inherent drawbacks in the MARS instruments, that is, whether one finds the app *interesting* or *engaging* depends heavily on his or her background, whereas individual aptitude for and experience in using apps in general could heavily influence perceived *ease of use and navigation*.

Another limitation is that the reviewers of apps in this study were not real CKD/ESRD patients. Rather, they pretended to be the patients. In addition, the reviewing team was quite small with a smaller number of devices, which may result in biased reviews. Future research should exploit a bigger team, more devices, and clinical trials with actual ESRD patients. A randomized control trial approach may also produce a more reliable result. Finally, as apps are perpetually being developed, so are the apps that support CKD/ESRD patients. A follow-up app search performed in early 2019 revealed that there are a few new apps available in the Apple and Google Play stores. However, these apps are rated relatively low compared with those reviewed in this study, with a low number of downloads. Furthermore, 5 of the 12 apps reviewed in this study appear to be unavailable in the app stores (*National Kidney Foundation Malaysia*, *Care After Kidney Transplant*, *AAKP myHealth Nutrition Guide*, *Kidney APPetite*, and *Phosphorus Tracker*). Although the 2 apps that scored the highest in this study seem to remain as the leading apps in the field, these findings indicate that evaluation of existing apps should be a continuous effort.

Conclusions

There has been an explosion of free mobile apps for tracking health in recent years. Although these apps appear to be promising tools, there has been a limited number of studies that systematically evaluate these apps, thereby burdening potential users of these apps with the responsibility to identify the apps that fit to their purposes. For some of these potential users, spending time and other resources to find the best app itself can be a challenging task, not to mention mastering the skills to use these apps effectively. With this in mind, this study conducted an evaluation of apps designed to assist CKD/ESRD patients who are under strict dietary and medication controls and tend to have limited recourses and capacity to explore numerous apps on their own. We used the MARS evaluation tool and identified the top 3 mobile apps supporting CKD/ESRD patients: *My Kidneys*, *My Health Handbook*; *My Food Coach*; and *National Kidney Foundation Malaysia*.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of apps excluded from the review.

[PDF File (Adobe PDF File), 88 KB - mhealth_v7i9e12604_app1.pdf]

Multimedia Appendix 2
Interrater reliability and internal consistency of the mobile apps rating scale items.
[PDF File (Adobe PDF File), 83 KB - mhealth_v7i9e12604_app2.pdf]

References

1. United States Renal Data System. 2014. USRDS Annual Data Reportpidemiology of Kidney Disease in the United States URL: https://www.usrds.org/2014/view/ [accessed 2019-01-15]
2. Centers for Disease Control and Prevention. 2016. Chronic Kidney Disease (CKD) Surveillance System URL: http://nccd.cdc.gov/CKD/detail.aspx [accessed 2019-01-15]
3. National Institute of Diabetes and Digestive and Kidney Diseases. 2017. Kidney Disease Statistics for the United States: About Kidney Disease URL: https://www.niddk.nih.gov/health-information/health-statistics/Pages/kidney-disease-statistics-united-states.aspx [accessed 2019-01-20]
4. Diamantidis CJ, Becker S. Health information technology (IT) to improve the care of patients with chronic kidney disease (CKD). BMC Nephrol 2014 Jan 9;15:7 [FREE Full text] [doi: 10.1186/1471-2369-15-7] [Medline: 24405907]
5. Curtin RB, Sitter DC, Schatell D, Chewning BA. Self-management, knowledge, and functioning and well-being of patients on hemodialysis. Nephrol Nurs J 2004;31(4):378-86, 396; quiz 387. [Medline: 15453230]
6. Dobbels F, Duerinckx N. Wise decisions on renal replacement therapy require knowledgeable patients and good self-report knowledge scales*. Transpl Int 2013 Dec;26(12):1161-1163 [FREE Full text] [doi: 10.1111/tri.12200] [Medline: 24118445]
7. Cavanaugh KL, Wingard RL, Hakim RM, Elasy TA, Ikizler TA. Patient dialysis knowledge is associated with permanent arteriovenous access use in chronic hemodialysis. Clin J Am Soc Nephrol 2009 May;4(5):950-956 [FREE Full text] [doi: 10.2215/CJN.04580908] [Medline: 19389825]
8. Nunes JA, Wallston KA, Eden SK, Shintani AK, Ikizler TA, Cavanaugh KL. Associations among perceived and objective disease knowledge and satisfaction with physician communication in patients with chronic kidney disease. Kidney Int 2011 Dec;80(12):1344-1351 [FREE Full text] [doi: 10.1038/ki.2011.240] [Medline: 21832984]
9. Greer RC, Cooper LA, Crews DC, Powe NR, Boulware LE. Quality of patient-physician discussions about CKD in primary care: a cross-sectional study. Am J Kidney Dis 2011 Apr;57(4):583-591 [FREE Full text] [doi: 10.1053/j.ajkd.2010.08.027] [Medline: 21131116]
10. Kuehn BM. Is there an app to solve app overload? J Am Med Assoc 2015 Apr 14;313(14):1405-1407. [doi: 10.1001/jama.2015.2381] [Medline: 25871657]
11. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]
12. 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]
13. Darby A, Strum MW, Holmes E, Gatwood J. A review of nutritional tracking mobile applications for diabetes patient use. Diabetes Technol Ther 2016 Mar;18(3):200-212. [doi: 10.1089/dia.2015.0299] [Medline: 26950679]
14. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
15. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
16. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
17. 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 Mar 23;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]
18. African Strategies for Health. 2019. mHealth Database URL: http://www.africanstrategies4health.org/mhealth-database.html [accessed 2019-02-01]
19. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mhealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
20. 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. [doi: 10.2196/mhealth.3422] [Medline: 25760773]

21. 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]
22. 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]
23. 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]
24. Bardus M, van Beurden SB, Smith JR, Abraham CA. 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]
25. Weekly T, Walker N, Beck J, Akers S, Weaver M. A review of apps for calming, relaxation, and mindfulness interventions for pediatric palliative care patients. Children (Basel) 2018 Jan 26;5(2):E16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]
26. Escoffery C, McGee R, Bidwell J, Sims C, Thropp EK, Frazier C, et al. A review of mobile apps for epilepsy self-management. Epilepsy Behav 2018 Apr;81:62-69. [doi: 10.1016/j.yebeh.2017.12.010] [Medline: 29494935]
27. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
28. Sonderegger A, Sauer J. The influence of design aesthetics in usability testing: effects on user performance and perceived usability. Appl Ergon 2010 May;41(3):403-410. [doi: 10.1016/j.apergo.2009.09.002] [Medline: 19892317]
29. Lindgaard G, Dudek C, Sen D, Sumegi L, Noonan P. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Trans Comput-Hum Interact 2011 Apr 1;18(1):1-30. [doi: 10.1145/1959022.1959023]
30. Hartmann J, Sutcliffe A, de Angeli A. Towards a theory of user judgment of aesthetics and user interface quality. ACM Trans Comput-Hum Interact 2008 Nov 1;15(4):1-30. [doi: 10.1145/1460355.1460357]

Abbreviations

CKD: chronic kidney disease
DM: diabetes mellitus
ESRD: end-stage renal disease
FDA: Food and Drug Administration
ICC: interclass correlation coefficient
IQR: interquartile range
IT: information technology
MARS: Mobile App Rating Scale
NHMRC: National Health and Medical Research Council
NKF: National Kidney Foundation

Edited by G Eysenbach; submitted 25.10.18; peer-reviewed by M Bardus, A Powell; comments to author 03.01.19; revised version received 05.03.19; accepted 19.06.19; published 04.09.19.

Please cite as:
Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, Koizumi N
Mobile Apps for the Care Management of Chronic Kidney and End-Stage Renal Diseases: Systematic Search in App Stores and Evaluation
JMIR Mhealth Uhealth 2019;7(9):e12604
URL: https://mhealth.jmir.org/2019/9/e12604/
doi:10.2196/12604
PMID:31486408

Original Paper

Mobile Apps for Medication Management: Review and Analysis

Katarina Tabi[1,2], PharmD; Abnashi Singh Randhawa[1], BSc; Fiona Choi[1,3], PhD; Zamina Mithani[1], BSc; Friederike Albers[4], BSc; Maren Schnieder[5], Dipl- Ing; Mohammadali Nikoo[1], MD; Daniel Vigo[1,6,7], MD, DrPH; Kerry Jang[1], PhD; Regina Demlova[2], MD, PhD; Michael Krausz[1,3], MD, PhD

[1]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

[2]Department of Pharmacology, Masaryk University, Brno, Czech Republic

[3]Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, BC, Canada

[4]Center of Human Movement Sciences, University of Groningen, Groningen, Netherlands

[5]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, United Kingdom

[6]Faculty of Health Sciences, Simon Fraser University, Vancouver, BC, Canada

[7]Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Katarina Tabi, PharmD
Department of Psychiatry
University of British Columbia
430-5950 University Boulevard
Vancouver, BC, V6T1Z3
Canada
Phone: 1 604 875 3525
Email: katarina.tabi@gmail.com

Abstract

Background: Pharmacotherapy remains one of the major interventional strategies in medicine. However, patients from all age groups and conditions face challenges when taking medications, such as integrating them into the daily routine, understanding their effects and side effects, and monitoring outcomes. In this context, a reliable medication management tool adaptable to the patient's needs becomes critical. As most people have a mobile phone, mobile apps offer a platform for such a personalized support tool available on the go.

Objective: This study aimed to provide an overview of available mobile apps, focusing on those that help patients understand and take their medications. We reviewed the existing apps and provided suggestions for future development based on the concept *understand and manage*, instead of the conventional *adhere to medication*. This concept aims to engage and empower patients to be in charge of their health, as well as see medication as part of a broader clinical approach, working simultaneously with other types of interventions or lifestyle changes, to achieve optimal outcomes.

Methods: We performed a Web search in the iOS Apple App Store and Android Google Play Store, using 4 search terms: medication management, pill reminder, medication health monitor, and medication helper. We extracted information from the app store descriptions for each eligible app and categorized into the following characteristics: features, author affiliation, specialty, user interface, cost, and user rating. In addition, we conducted Google searches to obtain more information about the author affiliation.

Results: A total of 328 apps (175 Android and 153 iOS) were categorized. The majority of the apps were developed by the software industry (73%, 11/15), a minority of them were codeveloped by health care professionals (15%, 3/20) or academia (2.1%; 7/328). The most prevalent specialty was diabetes (23 apps). Only 7 apps focused on mental health, but their content was highly comprehensive in terms of features and had the highest prevalence of the education component. The most prevalent features were reminder, symptom tracker, and ability to share data with a family member or doctor. In addition, we highlighted the features considered innovative and listed practical suggestions for future development and innovations.

Conclusions: We identified detailed characteristics of the existing apps, with the aim of informing future app development. Ultimately, the goal was to provide users with effective mobile health solutions, which can be expected to improve their engagement in the treatment process and long-term well-being. This study also highlighted the need for improved standards for reporting on app stores. Furthermore, it underlined the need for a platform to offer health app users an ongoing evaluation of apps by health professionals in addition to other users and to provide them with tools to easily select an appropriate and trustworthy app.

(JMIR Mhealth Uhealth 2019;7(9):e13608) doi:10.2196/13608

KEYWORDS

eHealth; mHealth; drugs; pharmaceuticals; therapy

Introduction

With the population aging worldwide [1] and increasing lifespans, the rates of chronic health conditions are accelerating [2]. Consequently, the regular use of medication to manage these conditions is increasing [3,4]. Moreover, there is also an increase in the number of individuals taking several medications simultaneously, known as polypharmacy [5,6]. The US National Health Survey has reported that the prevalence of polypharmacy (ie, taking ≥5 medications) rose from 8.2% to 15% between 1999 and 2012 [5]. This increase was found in all age groups, with the greatest increase among young adults aged 20 to 39 years, in which polypharmacy had grown from 0.7% to 3.1% [5].

Taking several medications simultaneously or starting to take a medication that is prescribed for the first time may bring with it challenges for many users [7,8]. These challenges may include integrating the medication into the daily routine at the right time, understanding the medication and its effects and side effects [9], or dealing with concerns about medication safety [10] and efficacy. Some patients want to organize their medications, keep a history or a list of currently used medications readily available, or track their symptoms in relation to the treatment [11]. To achieve these goals, a reliable and easy-to-use medication management (MM) tool that facilitates patient engagement becomes critical. The literature confirms that the patient's compliance with the treatment is significantly improved when the patient is engaged in the treatment [12].

Mobile apps can offer users (patients or health consumers) tools to facilitate their engagement in the treatment and well-being [13,14]. In addition, mobile apps represent one of the solutions for challenges with MM [15] simply because 95% of US adults in 2018 owned a cellphone [16,17], and it allows MM solutions to become highly personalized. Moreover, people carry cellphones with them for the majority of the day, so MM apps can provide assistance *on the go* and in the required time [16,18].

App users as well as developers are realizing this potential, which is reflected in the mobile app market growth, with increasing numbers of MM apps available [15]. However, there are a limited number of studies that provide detailed characteristics of the currently available solutions [15]. The term electronic health (eHealth) solution is a general term that refers to software, telecommunication, and virtual technology related to health [19]. Some of the earlier studies focused on apps with selected features, such as reminders and other medication adherence strategies [20-22], or on a specific population [23]. Therefore, to inform further app development, it is crucial to explore the existing selection of solutions available in the market, review their key characteristics, and identify the needs and directions for future development.

This study aimed to provide an overview of the existing solutions in the mobile app marketplace, with a focus on apps that help users understand and take their medications. We reviewed the existing apps and provided suggestions for future development based on the concept *understand and manage*, as opposed to the conventional *adherence to medication*. This concept aims to engage users and empower them to be in charge of their health and well-being as well as to see medication as part of a broader clinical approach, working simultaneously with other types of interventions or lifestyle changes, to achieve optimal outcomes.

We outlined the detailed characteristics of the available apps and sought to (1) identify who designed each app and whether health experts were involved in the development and design; (2) identify key features defined as (a) features that are most prevalent and (b) features that are novel or innovative and have the potential to address the concept *understand and manage*; and (3) create a framework to categorize the mobile health solutions. We believe that this framework may contribute to the understanding of the current landscape of MM apps and to the development of future solutions.

Methods

Overview

The approach used in this study was based on the previous literature [15,24,25]. The app search was conducted on the 2 main mobile app stores: Android Google Play Store (from here on referred to as Android store) and iOS Apple App Store (from here on referred to as Apple store). Worldwide mobile phones using these 2 operating systems (ie, iOS and Android) currently account for 96% of the mobile operating system market share, and in North America and Europe, they account for 99% of the market share [26]; the market share numbers for newly purchased devices are even higher [27]. The app stores were searched in January 2017. The search query was typed into the search engine Start page (details described in the section *Eliminating Personalization During the Search*). The query consisted of the search term and the app store, for example, *medication management site://play.google.com/store/apps/*. The search terms used were *medication management, pill reminder, medication health monitor,* and *medication helper*. The selection of search terms was based on a panel discussion among the research team and terms used in previous studies. The final search terms were confirmed based on a preliminary search, when the selected terms showed greater relevance of results. Table 1 lists the search terms together with the number of apps found with them in each app store.

Table 1. Search terms and the number of eligible apps found with them in each app store.

Search term	Number of apps found that met the inclusion criteria	
	Apple store	Android store
Medication health monitor	69	70
Medication management	53	61
Pill reminder	49	94
Medication helper	53	54

Selection of Eligible Solutions

Each term was searched separately in each app store, and 2 researchers performed identical searches independently. For all search results, the information available in the app store description was reviewed. The inclusion criteria were as follows: (1) user population consisting of patient and health consumers, including family or caregivers, and the eligible solution was written in lay language; (2) purpose—helping users take or understand their medication; and (3) the app was available in English. Apps were excluded from further review if professional language was used, they were intended for health care experts, and they did not have a clear focus on helping take medications and instead focused on other issues, such as purchasing medications or ordering refills.

Solutions that met the eligibility criteria were selected by the researcher and their URLs were saved. After the independent search, the 2 researchers compared the lists of apps they identified as eligible and performed analyses to determine agreement between the 2 sets of results. Interrater reliability was determined based on Cohen kappa (0.81). In accordance with Landis and Koch statistics guidelines, the strength of the agreement between researchers was almost perfect (0.81-1.00).

Only those apps that were considered ineligible by both researchers were excluded from further review. Discrepancies between the 2 researchers were reviewed (third opinion) and resolved. Subsequently, both researchers agreed on the final list of results for each search term [15].

Duplicates on the same device platform were identified using Microsoft Excel and removed, whereas solutions that appeared across 2 different platforms were not considered duplicates [28]. Subsequently, the final list of unique solutions meeting the eligibility criteria was created.

Eliminating Personalization During the Search

We applied the following strategies to decrease the extent to which the search results were affected by personalization, app store search algorithm, and cookies. We used (1) Start page as the search engine, which uses results from Google while offering increased anonymity through Secure Sockets Layer encryption and limited data collection; (2) a different Web browser than is usually used on the computer; and (3) incognito mode or a private window of the browser [29,30].

Data Extraction and Solution Assessment

Before data extraction, a preliminary framework of app characteristics was designed, based on a panel discussion of the research team and the previous literature [15,24,31]. Later, the framework was shaped according to information that accumulated during the selection of eligible solutions. The final framework of the solution characteristics is outlined in Table 2.

Table 2. Final framework for app characteristics categorization, including description for selected subcategories.

Categories of app characteristics and subcategories	Description
Feature or purpose	
Medication reminder	Reminds user to take medication dose in real time
Shares data and reports with others	When chosen, ability to share health data managed in the app with other people (family member and health care professional)
Education about medication	Information about medication (eg, benefits, side effects, interactions, and use)
Identifies pills	Identifies unknown pill, usually through the phone's camera and the appearance
Checks for drug interactions	After entering two or more medications (or medication with food or alcohol), the checker will evaluate the risk of their interactions
Tracks symptoms, side effects, health data, and vitals	Tracks user's measurements (eg, blood glucose, blood pressure, weight, pulse, temperature, mood, and sleep patterns), symptoms of disease, and side effects
Manages profiles of multiple users	Within one app, there is an option to have medication profiles of several people (eg, other family members)
Synchronization with other apps or devices	The app has the ability to synchronize itself and the entered data with another mobile app or device (eg, Apple Watch)
Data privacy and security	Different forms of information privacy (eg, password) and security; privacy policies are transparent and easy to find; and advanced protection of health data
Medication management (MM) is not the primary aim of the app	Medication-related functions are not the primary focus of the app, for example, many fitness apps have mainly other features, and the medication component simply represents a minor piece
Other features	Describing any other features, not listed above
Author Affiliation	
Academia	Developed by or in affiliation with university or other forms of academia
Health care professionals	Developed by or in affiliation with health care professionals, health centers, and hospitals
Software industry	Including software companies or independent software developers
Other	Affiliations not listed above; usually charities, nonprofits, and government organizations; pharmaceutical companies and insurance companies; various level of health care professional involvement; therefore, the apps under this subcategory may have input from health experts
Insufficient information	The affiliation cannot be found
Specialty in Medication	
Diabetes	App specialized in MM for diabetes-specific medication
Women's reproductive health	App specialized in the management of contraceptive pills or women's reproductive health-specific medication
Cardiovascular health	App specialized in MM for heart and blood circulation–specific medication
Lifestyle management	App specialized in lifestyle management, often offering MM as one of the several features
Neurology	App specialized in MM for neurology-specific medication
Mental health	App specialized in MM for mental health–specific medication
Oncology	App specialized in MM for cancer-specific medication
Hematology	App specialized in MM for hematology-specific medication
Lungs, allergy, and immunity	App specialized in MM of drugs for allergy, asthma, lung disease, and immune system–specific medication
Veterinary medicine	App specialized in MM for pets
Digestive system	App specialized in MM for gastrointestinal-specific medication
Other specialty	Specialty not included in the list
Without specialty	General MM, for use with any medication

Categories of app characteristics and subcategories	Description
User Interface	
Static only	Provides minimal interaction with the user—does not work with individual user's data, just displays the same content for everybody; only available interaction is for navigation and settings; usually presents information
Single dynamic feature	Provides one interactive component; compared with the multiple dynamic category, apps in this category often have only a basic pill reminder and no other features
Multiple dynamic features	Solutions in this category are more comprehensive, with more than one dynamic feature. Dynamic user interface is defined as providing the opportunity to input one's individual data into the app and being able to interact with the app (eg, tracking health data, warning if set parameters are exceeded, games, and communication)
Cost	
Free	The app itself was available for free. However, some of the apps in this category had the option of purchasing extra content or features (referred to as in-app purchases). These apps were coded as free and the cost for the optional in-app purchases was listed at the same time—under the next column cost
Price	Includes price for the app itself or price for the in-app purchases (available also in free apps)
Currency or country	Encoded to determine the currency that was used; later, all the prices can be converted into the same currency for comparison and descriptive statistics
Notes	Explains details related to the listed price—whether the price is for the app itself, regular subscriptions, or in-app purchases
User Rating	
Number of stars	1 star-5 stars
Number of ratings per app	Total number of ratings in the app store available for the given app; if the information was presented by the app store

Subsequently, the 2 researchers assessed and categorized the apps. The first 20.1% (66/328) of the apps were categorized by the 2 researchers together. During this process, they came to an agreement in understanding the scope of the subcategories. Subsequently, they divided the remaining sample of apps (262) and each researcher extracted information for half of the sample. They assessed the apps at the same time and discussed any issue or uncertainty that arose.

Information was extracted from the app store description and from the available screenshots. An exception was the category author affiliation. The developers rarely described their credentials or affiliation within the text of the app store description. Moreover, app stores provided only the name of the developer in most cases. Therefore, to find out more information about the author's affiliation, we conducted Google searches using the name and the website provided by the app store.

The app features were assessed in the following way. Features that were not part of the list of prevalent features were manually entered in the database as *Other features*. At the same time, researchers flagged those features, which they considered novel, interesting, and innovative. Upon completion of the solution assessment, the team of coauthors reviewed the flagged features and agreed on the final list, which is included in the study as the *novel and innovative features*.

Data Analysis

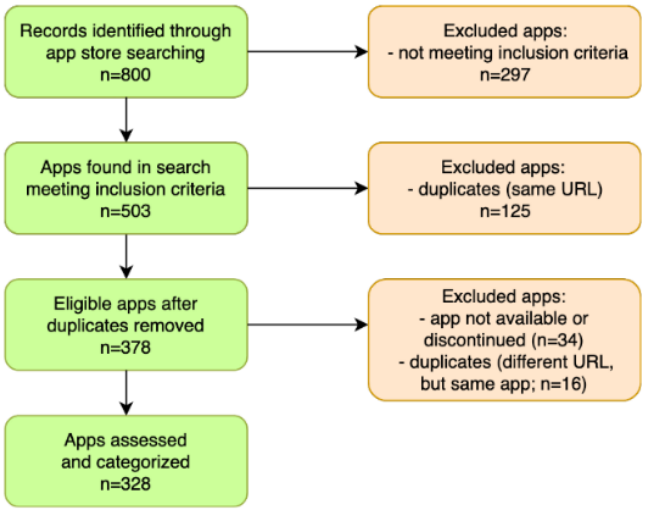
Once all the apps were coded, descriptive statistics were computed for each variable. Cohen kappa and selected descriptive statistics were computed using SPSS (IBM), version 25. The rest of the descriptive statistics and charts were prepared in Microsoft Excel, version 16.1.1 and Euler online application. The visualization through Euler diagram and related descriptive statistics were presented to add to the understanding of who authored the apps and which professionals collaborated during the app development.

Results

Search and Categorization

The initial app store search yielded 800 records, of which 297 were excluded based on the app store description because they did not meet the inclusion criteria. From the eligible apps, other apps were excluded later because they were duplicates or they were no longer available. A total of 328 apps were included in the final assessment. Categorization was carried out in December 2017. Between the search and categorization, 34 apps were discontinued. Therefore, the apps that were assessed and are the subject of this review are the ones that remained available on the app stores for a year. Figure 1 demonstrates the various stages of the review process.

Figure 1. Diagram illustrating the review process and exclusion of apps at various stages of the study.



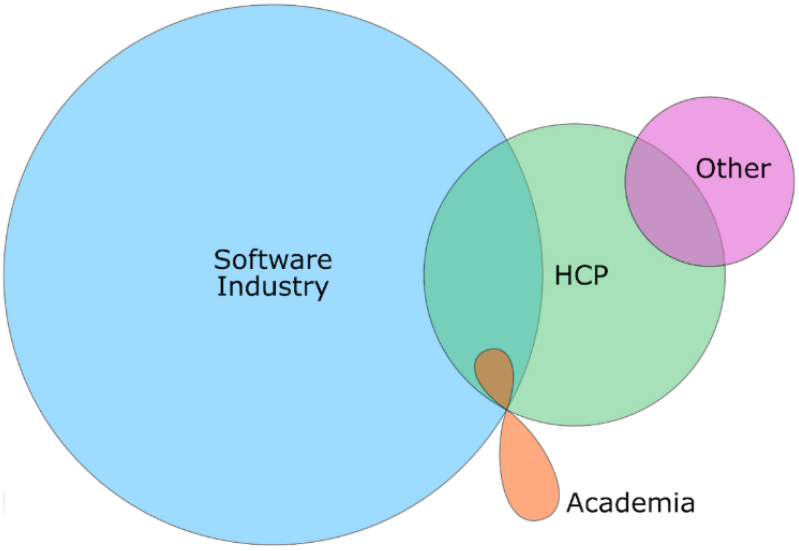
General Characteristics of Included Apps

Both marketplaces were represented in the final list of apps: the Android store accounted for 53.8% (175/328) of the apps, and 46.6% (153/328) of the apps were from the Apple store. In terms of cost, the majority of the apps were available for free (86.3%; 283/328), some of which had options for in-app purchases (9.5%; 31/328) for extra content or features. The average price of paid apps (13.7%; 45/328) was US $3.77. The price range for the paid apps or in-app purchases was between US $0.99 and $24.55.

Authors and Affiliations

The majority of the mobile apps were developed by the software industry (72.9%; 239/328), which consisted of software companies (219/239) and independent software developers (20/239). In addition, 14.6% (48/328) were developed with the involvement of health care professionals. Very few apps were developed by, or produced in collaboration with, academic institutions (2.1%; 7/328). Furthermore, 5.2% (17/328) were developed by *Others*, which consisted mostly of governmental organizations and nonprofits; for a longer list of institutions, see Table 2. In 56 cases (17.1%; 56/328), there was insufficient information available about author affiliation. The distribution of different authors or developers, including the combinations, is illustrated in Figure 2.

Figure 2. Diagram illustrating the distribution of different authors involved in the development of the medication management apps. The overlapping regions indicate the apps codeveloped by 2 or 3 different author affiliations (combinations). HCP: health care professionals.



Features

The most prevalent features of all reviewed apps are listed in Table 3. From the rest of the features, those identified as novel or innovative, are listed in Table 4. From the apps included in the study, 77 apps did not have MM as their main purpose. The main focus of these apps was usually overall health management, and a feature related to medication was not one of their primary features. Some of these apps were supporting healthy life style choices and a broader clinical approach to health, including tracking physical activity, food intake, sleep, emotions, and others.

Table 3. The most prevalent features of the mobile apps for medication management.

Features	Number of apps that have the feature
Medication reminder	282
Tracks symptoms, side effects, health data, and vitals	152
Shares data and reports with others	135
Synchronization with other apps or devices	70
Education	63
Manages multiple user profiles	60
Data privacy and security	45
Checks for drug interactions	12
Identifies pills	11
Medication management is not the primary aim of the app	77

Table 4. Novel or innovative features of the apps.

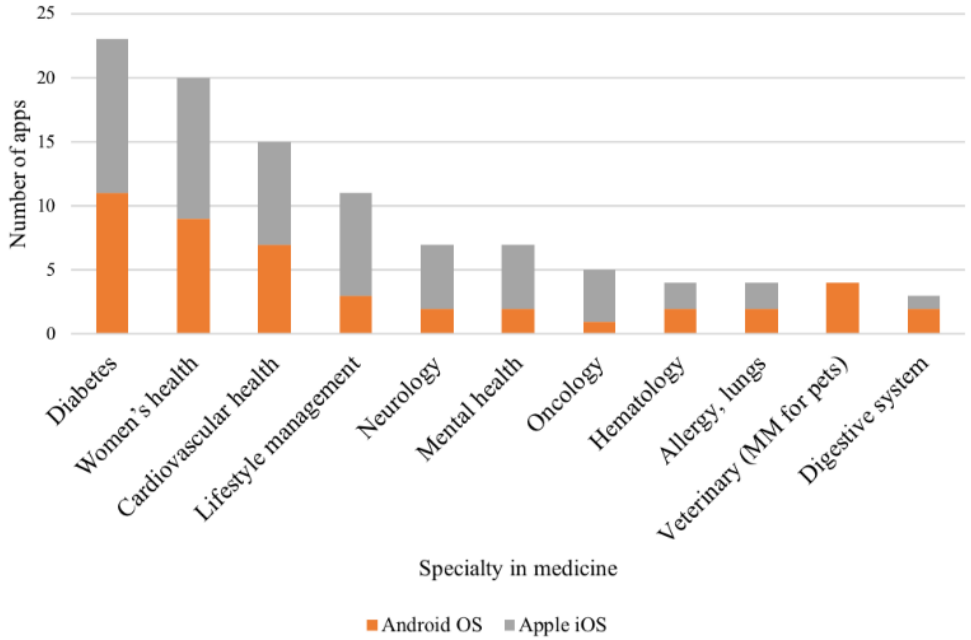
Feature description	Examples
Novel ways of data entry, instead of typing	Scanning barcode, taking a picture of the package, and voice entry
Correlations based on entered data	Medication-blood pressure, food-glucose, and medication-sleep
Communication system with the health care professionals	Connecting user with pharmacist, doctor, nurse, or other professionals
Goal setting	For physical activity, diet, sleep hygiene, social interactions, and other lifestyle choices
Journaling	Accomplishments, emotions, triggers, pain, sex drive, mood and thought journal, and other
Reminder noticing the change of time zone	Support during traveling
Reminder with personalized voice	Voice of grandchild reminds the grandmother
Warning if safe dosage exceeded	App includes information about the recommended maximum daily dosage and warns if that is exceeded
uBox (place where the medication can be physically stored) synchronized with the app	Option to set: care contacts receive text message when patient misses a dose
Safety plan for acute situations	Prepared ahead when the health condition is stable or prepared together with the health care professional
Emergency button	911 and counselor

Specialty in Medicine

Of the apps focusing on certain medical condition and its specific MM (36.0%; 118/328), most apps focused on diabetes (n=23), women's reproductive health (n=20), and cardiovascular health (n=15; see Figure 3). Mental health was not 1 of the 3 most prevalent specialties. Of the 7 apps focusing on MM for mental health, 5 were from the Apple store, whereas only 2 apps were found in the Android store. Most apps included general support for mental health treatment; 2 apps were more specialized, focusing on attention-deficit/hyperactivity disorder and bipolar disorder.

Overall, 210 apps did not specify a specialty (64.0%; 210/328), which means they were designed to help with general MM. These general MM apps may differ from the specialized apps by nature of their features, which are designed to address a wide spectrum of medications and health conditions. For example, the tracker of health data in a general MM app usually offers tracking of weight, heart rate, sleep, or exercise. Conversely, a specialized MM app for mental health may track some of the previously listed health data in addition to mood, thoughts, and social events. Some users may prefer the general MM apps. For example, a user taking several medications across a range of specialties can manage all their medications within one app and customize the settings accordingly.

Figure 3. Distribution of medication management (MM) apps focusing on certain medical conditions or specialties in medicine.



Another way of looking at the apps' features is to see the relationship between features and medical specialty. Table 5 shows what features were prevalent in apps for diabetes, cardiovascular health, and so on and compares them with other specialties. For example, the table shows that 61% (14/23) of the diabetes apps had reminders, 57% (13/23) of them had the option to share data with others, 22% (5/23) had an education component, 4% (1/23) could check interactions, 91% (21/23) tracked symptoms and vitals, and so forth. Specifically, Table 5 shows that the diabetes apps had the greatest variety of features (9 different features are represented). In contrast, the veterinary apps had the lowest variety, with only 3 different features. Not surprisingly, mental health MM apps had the highest prevalence (43%; 3/7) of the education component.

Table 5. Distribution of medication management apps by feature and specialty in medicine.

Specialty	Number of apps having the given feature within the specialty, n1 (%)[a]									Number of all apps within the specialty, n2
	Rem[b]	Sha[c]	Edu[d]	Int[e]	Tra[f]	Mul[g]	Syn[h]	Not[i]	Sec[j]	
Diabetes	14 (61)	13 (57)	5 (22)	1 (4)	21 (91)	4 (17)	8 (35)	16 (70)	1 (4)	23
Women's health	18 (90)	1 (5)	2 (10)	0 (0)	9 (45)	0 (0)	3 (15)	3 (15)	5 (25)	20
Cardiovascular health	12 (80)	6 (40)	1 (7)	0 (0)	12 (80)	2 (13)	6 (40)	7 (47)	1 (7)	15
Lifestyle management	11 (100)	7 (64)	1 (9)	0 (0)	11 (100)	1 (9)	6 (55)	7 (64)	1 (9)	11
Neurology	6 (86)	4 (57)	0 (0)	0 (0)	7 (100)	0 (0)	5 (71)	5 (71)	0 (0)	7
Mental health	6 (86)	5 (71)	3 (43)	0 (0)	6 (86)	0 (0)	1 (14)	4 (57)	2 (29)	7
Oncology	4 (80)	2 (40)	2 (40)	0 (0)	4 (80)	0 (0)	1 (20)	2 (40)	0 (0)	5
Hematology	3 (75)	3 (75)	0 (0)	0 (0)	4 (100)	0 (0)	2 (50)	2 (50)	1 (24)	4
Allergy, lungs	3 (75)	3 (75)	0 (0)	0 (0)	3 (75)	2 (50)	1 (25)	2 (50)	0 (0)	4
Veterinary—medication management for pets	4 (100)	0 (0)	0 (0)	0 (0)	2 (50)	1 (25)	0 (0)	0 (0)	0 (0)	4
Digestive system	1 (33)	0 (0)	0 (0)	0 (0)	2 (67)	0 (0)	1 (33)	2 (67)	0 (0)	3
Other specialty	11 (73)	7 (47)	4 (27)	0 (0)	7 (47)	0 (0)	4 (27)	10 (67)	0 (0)	15

[a]Percentage (%) calculated as (n1/n2) × 100.

[b]Rem: Reminder.

[c]Sha: Share data.

[d]Edu: Education and information.

[e]Int: Interactions.

[f]Tra: Tracks symptoms and health data.

[g]Mul: Multiple profiles management.

[h]Syn: Sync with other apps or devices.

[i]Not: Medication management not primary aim.

[j]Sec: Data security.

User Interface

Apps with only a static user interface accounted for 0.6% (2/328) of the sample. A definition of different user interfaces is provided in Table 2. The remaining apps had a dynamic user interface. We defined this as providing an opportunity to input one's individual data into the app (medication type, use times, tracking health outcomes, or other individual information), and the possibility to interact with the feature. Of the dynamic apps, 99 (30.2%; 99/328) had only 1 dynamic feature (usually a basic pill reminder), whereas 227 (69.2%; 227/328) apps had multiple dynamic features available, and the complexity of these apps was higher.

Rating

The user rating scale was slightly different for the Android and Apple stores. The Android store showed ratings with an accuracy of 0.1 star (eg, 3.5 and 3.6), whereas the Apple store reported ratings with an accuracy of 0.5 star (eg, 3.5 and 4). In both stores, 5 stars was the best possible score. Of the 275 rated apps, the average rating was 3.84 stars. The number of raters per app ranged between 1 and 152,011, with an average of 2658 raters. However, 50 apps in the Apple store and 3 apps in the Android store did not have rating information available.

Discussion

Principal Findings

This review describes a comprehensive range of characteristics of available mobile solutions for MM, which help users understand and take their medication. An emphasis is placed on the functional features the current solutions offer and how this can inform future app development. Another main finding is related to the low transparency of information available in the app store description. We suggest that it is difficult for the user to identify the author's affiliation and professional credentials, as reporting this information is not a standard in the app store descriptions.

Market Overload—Difficult for Users to Navigate

When searching for MM apps, we found over 300 unique apps on the Apple and Android stores. A study by Martínez-Pérez examining mobile apps for the most prevalent health conditions reported that there were more than 1000 apps focusing on diabetes or depression [32]. This represents a large number of apps that the user must filter through to find the right one.

When one looks for an online solution for their health condition, well-established requirements for health interventions, such as

providing evidence-based content and clinical effectiveness become pivotal, in addition to the qualities that are commonly sought out in any apps, such as user-friendliness and engagement [33,34].

Low Indication of Author Affiliation

Another aspect of the health apps that would be helpful for the user to know during the process of choosing an app is the topic-related credibility of the author. However, during the solution assessment, we observed that the majority of apps did not mention the credentials of the author or professionals who helped codevelop the app in the app store description. Currently, the 2 main app stores require app authors to list their name (termed *seller* in the Apple store and *developer* in the Android store) and website if available. However, listing the qualifications or affiliations of the author is currently not a store requirement. Therefore, very few apps include this information in the app store description. Similarly, many other studies reviewing mobile apps for a variety of health conditions described that author affiliation and reporting of content sources in the app store description were infrequent [24,25,35,36]. The earlier Towards Evaluation and Certification of Telematics Services for Health project focusing on quality and safety in health informatics, recommends that the identity and qualification of the health care professional responsible for the clinical element of the software design should be one of the labeling requirements for clinical software.

We continued to look further for the authors' credentials and conducted Google searches using the website and the name listed in the app store. We did so to find more information on the subject and examine the authors' affiliations for the purpose of this study. However, as this is time consuming, we do not predict users would routinely do the same.

We found a lack of health care professional involvement (14.6%; 48/328) in app development. Furthermore, even fewer apps were codeveloped by academia. In addition, a few apps were authored by other organizations such as charities, nonprofits, government organizations, insurance companies, or pharmaceutical companies. Conversely, the majority of solutions were developed by software developers or tech companies. The previous literature widely indicates that there is a lack of involvement of health professionals or academics in the development of health apps [37-39].

Features

One of the most prevalent features was symptom or vital tracker. It provides a unique record of health outcome details over time that can prove helpful to both users and their health care providers. We suggest that tracking symptoms may also be beneficial within a broader treatment approach, where medication is an integrated component of a comprehensive strategy (including, eg, diet, physical exercise, psychotherapy, or other, depending on the health issue). In such cases, tracked health outcomes provide insight into the effects and synergies of the interventions and may help optimize the treatment for each user. Future development could offer alternatives to automatically track data through wearables or new upcoming technologies, increasing and facilitating utilization of tracking features as a result.

There is an increasing number of apps (21.0%; 69/328) with the option to synchronize data from the MM app with a different app or device, including wearables (eg, Apple Watch). A similar review from 2013 did not report synchronization as a prevalent feature [15], which may be because of the focus of the previous study or market changes. We suggest that future development can further build on the synchronization efforts and reflect usage of wearables or other devices prevalent in society. This may improve the functionality of MM apps, including the tracking options described in the previous paragraph, as well as engagement of users in the treatment.

A majority of the reviewed apps offered a medication reminder. A study by Ahmed et al focusing on adherence apps has also described other adherence strategies such as gamification, in addition to the common reminder [21]. Another common feature was the option to share data with another person, for example, a doctor or a family member.

In addition to features helping users take their medication, almost one-fifth of the apps included an education component to help users understand the medication or the treatment process. Examples include information about correct usage, side effects, and interactions. For future development, we suggest also covering common patient concerns (legitimate vs myths) and desired treatment outcomes based on the evidence. The education section could be written in an easy-to-understand language and provide an added value compared with the patient information sheet, which is usually provided together with a prescription medication.

Some of the interesting novel features identified in this study were easier data entry for new medication (eg, barcode scanning and voice entry), correlations based on data entry, providing warnings if safe dosage exceeded, communication with health professional, and safety plan for acute or emergency situations. In addition, an increasing number of apps included in their tracking and journaling features options supporting healthy lifestyle choices such as physical activity; social interactions; food diary; awareness of emotions, mood, and thoughts; accomplishments; and triggers. Many of the novel features could inform future development, especially ones that make the use of mobile MM increasingly effortless, for example, easier data entry options, and features empowering the user, for example, individualized and improved tracking features.

Other Characteristics

Although most of the apps were not specialized for MM of a particular drug or condition, there were several apps with a focus on a specific medical condition or specialty. The greatest number of apps were for MM of diabetes and apps helping women keep track of their reproductive health, including reminders for contraceptive pills.

When looking at the distribution of prevalent features by different specialties, diabetes apps had the greatest variety of features. When this fact is taken into consideration alongside the advancement in engagement and gamification [40,41], MM apps for diabetes can serve as a potential innovation model for

other specialties. Although only a few apps focused on mental health MM, they were some of the most comprehensive ones in terms of features and content and had the highest prevalence of the education component.

Initiatives for User Navigation

This study offers an overview of the app market for professionals in the field of eHealth and can inform development of new innovative solutions. On the other hand, the perspective of the user (the patient or the health consumer) remains crucial. For them, a highly dynamic approach including up-to-date transparent overviews of characteristics and evaluations of solutions would be very beneficial and would empower them to *understand and manage* their medication with the support of a mobile app. The pivotal first step for the user is to select an appropriate app.

There are several initiatives aiming to help users navigate the app market, such as Apps Library by the National Health Library in the United Kingdom, RankedHealth or Psyberguide [42-44]. Health On The Net Foundation provides a certification (HONcode) to health-related websites providing reliable health information and meeting the standards for ethics and transparency [45,46]. The Ontario Telemedicine Network has launched a website called Practical Apps, which publishes reviews on specific health topics and can be used by public or health care providers recommending apps to their patients [47]. Previous literature states that developing such an initiative remains a challenge with little success thus far [24,48]. Deshpande and Jadad suggest a *crowd-sourcing collaborative approach*, similar to Wikipedia, as a potential solution [24,48].

A future direction could be an online registry and review aggregation website, similar to Metacritic, which is used for film and other media. The future platform could contain structured overviews of various characteristics, reviews by professionals (trained health care *critics*) including an *evidence-based score*, and reviews by users, all in one place.

The professional reviews could follow some of the standardized frameworks for quality assessment of mobile apps, for example, Mobile App Rating Scale [49], suitable for a wide range of health apps; a recently published assessment framework specifically for mental health apps [50]; or other published frameworks [33,51-54].

Limitations

One of the potential limitations of this review is using the information available in the app store description, without downloading and fully testing the apps out. The apps themselves potentially may have additional functionalities not described in the app store description and that may have impacted the results presented in this study. However, this way of reviewing the current apps and collecting their characteristics mimics the experience of the user when they are deciding whether to download or purchase an app or not, and thus, it bridges the gap between reality and research. This is of particular value because the fact whether an app will be actually used or just be in the app store with minimal downloads is based in the first place on the information available in the app store description. Developers are aware of this and strongly motivated to list their apps' features. If a major feature is not mentioned in the app store description and visible at first sight, usually the app naturally gets passed over. In addition, the proposed study focused on collecting the objective characteristics of the apps and did not aim to do a qualitative evaluation of the solutions.

A second limitation is related to the time difference between data collection (January 2017) and time of submission of the publication, considering the fast-evolving landscape of the app market. Moreover, there was also a 1-year time difference between the search and categorization. The disadvantage of this is that 34 apps were discontinued during this period. The advantage might be that the rest of the apps (328) that were categorized, and their characteristics are outlined in this study, are the sustained ones that remained available over a year. Third, identification of app duplicates was limited to the duplicates within one app store. As store requirements are different between the Android and Apple store, we were not able to reliably identify duplicates across 2 different platforms.

The inclusion criteria included English language. Therefore, the findings are not representative of all MM apps available in the global market offered in languages other than English. Moreover, both researchers were physically present in Canada during the search process. Even though substantial effort was put into increasing the anonymity during the search, a search in another country may have returned different results. Similarly, despite these efforts, the search results may also be affected by the app stores' algorithms. Specifically, the ranking of the results may be affected by various factors, including but not limited to how many people have clicked on the app title before, the user rating, increase of downloads in the past month, or the number of downloads per day.

Conclusions

Mobile apps have the potential to empower patients with personalized immediate support and improve compliance with the treatment and engagement in long-term well-being. However, searching for an appropriate MM app is challenging because the marketplace offers hundreds of solutions which makes it difficult for the user to filter through. To find a trustworthy app, it would be beneficial for the user to have the information about the content source or author affiliation easily available during the search. However, this study found a lack of reporting of the author affiliation in app store descriptions. Therefore, the study highlights the need for improved standards for reporting on app stores. In addition, it underlines the need for a platform to offer users of health apps an ongoing evaluation by health professionals in addition to other users. The aim of such efforts would be to provide users with tools to readily assess the credibility of eHealth solutions before making their choice.

The study reports that prevalent features of available MM apps were reminder, symptom tracker, and ability to share data with a family member or doctor. The results presented in the study help inform the theoretical and practical approach for app development, in particular, decisions related to the content selection process. Building on the existing experience, it is important to start working on the next generation of solutions, which will improve engagement of users in the treatment

process. Future development may include improved tracking features to optimize treatment for each user based on their exact treatment outcome data, as well as components making the use of mobile MM increasingly effortless (eg, easier data entry options). Moreover, emphasis could be put on a broader comprehensive treatment approach framing medication as its integrated component instead of a stand-alone intervention.

Acknowledgments

This paper was written with the support of the Specific University Research (MUNI/A/1132/2017) provided by Ministry of Education, Youth and Sport of the Czech Republic, Freemovers program, and the Faculty of Medicine at Masaryk University.

Conflicts of Interest

None declared.

References

1. United Nations. 2017. World Population Ageing: Highlights URL: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf [WebCite Cache ID 75hxBj0RD]
2. Raghupathi W, Raghupathi V. An empirical study of chronic diseases in the United States: a visual analytics approach. Int J Environ Res Public Health 2018 Mar 1;15(3):pii: E431 [FREE Full text] [doi: 10.3390/ijerph15030431] [Medline: 29494555]
3. Zhang F, Mamtani R, Scott FI, Goldberg DS, Haynes K, Lewis JD. Increasing use of prescription drugs in the United Kingdom. Pharmacoepidemiol Drug Saf 2016 Jun;25(6):628-636 [FREE Full text] [doi: 10.1002/pds.3947] [Medline: 26710965]
4. Gu Q, Dillon CF, Burt VL. Prescription drug use continues to increase: US prescription drug data for 2007-2008. NCHS Data Brief 2010 Sep(42):1-8 [FREE Full text] [doi: 10.1037/e665492010-001] [Medline: 20854747]
5. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in prescription drug use among adults in the United States from 1999-2012. J Am Med Assoc 2015 Nov 3;314(17):1818-1831 [FREE Full text] [doi: 10.1001/jama.2015.13766] [Medline: 26529160]
6. Payne RA, Avery AJ, Duerden M, Saunders CL, Simpson CR, Abel GA. Prevalence of polypharmacy in a Scottish primary care population. Eur J Clin Pharmacol 2014 May;70(5):575-581. [doi: 10.1007/s00228-013-1639-9] [Medline: 24487416]
7. Aston J, Wilson KA, Terry DR. The treatment-related experiences of parents, children and young people with regular prescribed medication. Int J Clin Pharm 2019 Feb;41(1):113-121 [FREE Full text] [doi: 10.1007/s11096-018-0756-z] [Medline: 30478490]
8. Millar E, Gurney J, Stanley J, Stairmand J, Davies C, Semper K, et al. Pill for this and a pill for that: a cross-sectional survey of use and understanding of medication among adults with multimorbidity. Australas J Ageing 2019 Jun;38(2):91-97. [doi: 10.1111/ajag.12606] [Medline: 30556358]
9. Mallet L, Spinewine A, Huang A. The challenge of managing drug interactions in elderly people. Lancet 2007 Jul 14;370(9582):185-191. [doi: 10.1016/S0140-6736(07)61092-7] [Medline: 17630042]
10. Hughes RG, Blegen MA. Medication administration safety. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Maryland, United States: Agency for Healthcare Research and Quality; 2008.
11. Wilcox L, Woollen J, Prey J, Restaino S, Bakken S, Feiner S, et al. Interactive tools for inpatient medication tracking: a multi-phase study with cardiothoracic surgery patients. J Am Med Inform Assoc 2016 Jan;23(1):144-158 [FREE Full text] [doi: 10.1093/jamia/ocv160] [Medline: 26744489]
12. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34 [FREE Full text] [doi: 10.2196/publichealth.8923] [Medline: 29712626]
13. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
14. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
15. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]
16. Statista. Number of Mobile Phone Users Worldwide From 2015 to 2020 (in Billions) URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/ [WebCite Cache ID 75ehS4Oy0]
17. Pew Research Center. 2019. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [WebCite Cache ID 75hk7VXrt]

18. 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]
19. Lee A, Sandvei M, Asmussen HC, Skougaard M, Macdonald J, Zavada J, et al. The development of complex digital health solutions: formative evaluation combining different methodologies. JMIR Res Protoc 2018 Jul 16;7(7):e165 [FREE Full text] [doi: 10.2196/resprot.9521] [Medline: 30012548]
20. Park JY, Li J, Howren A, Tsao NW, de Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
21. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Danish HS, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
22. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
23. Stuck RE, Chong AW, Mitzner TL, Rogers WA. Medication management apps: usable by older adults? Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):1141-1144 [FREE Full text] [doi: 10.1177/1541931213601769] [Medline: 29158662]
24. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
25. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
26. StatCounter Global Stats. Mobile Operating System Market Share Worldwide: June 2018 - June 2019 URL: http://gs.statcounter.com/os-market-share/mobile/worldwide [accessed 2019-01-26] [WebCite Cache ID 75hkF0M31]
27. IDC: The premier global market intelligence firm. Smartphone Market Share URL: https://www.idc.com/promo/smartphone-market-share/os [WebCite Cache ID 75hkKCGCs]
28. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
29. Preibusch S. The value of web search privacy. IEEE Secur Priv 2015;13(5):24-32 [FREE Full text] [doi: 10.1109/MSP.2015.109]
30. Winkler S, Zeadally S. An analysis of tools for online anonymity. Int J of Pervasive Compt Commun 2015 Nov 2;11(4):436-453. [doi: 10.1108/IJPCC-08-2015-0030]
31. Endl R, Jäschke T, Diana CT, Wickinghoff V. eHealth Suisse. 2015. mHealth im Kontext des Elektronischen Patientendossiers: Eine Studie im Auftrag von eHealth Suisse URL: https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2015/F/150302_rapport_fhsg_mhealth_alemand.pdf [WebCite Cache ID 75hoIwMPV]
32. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
33. Fiore P. How to evaluate mobile health applications: a scoping review. Stud Health Technol Inform 2017;234:109-114. [doi: 10.3233/978-1-61499-742-9-109] [Medline: 28186025]
34. Rigby M, Forsström J, Roberts R, Wyatt J. Verifying quality and safety in health informatics services. Br Med J 2001 Sep 8;323(7312):552-556 [FREE Full text] [doi: 10.1136/bmj.323.7312.552] [Medline: 11546703]
35. Brouard B, Bardo P, Bonnet C, Mounier N, Vignot M, Vignot S. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? Ann Med 2016 Nov;48(7):509-515. [doi: 10.1080/07853890.2016.1195010] [Medline: 27348761]
36. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
37. Pereira-Azevedo N, Osório L, Cavadas V, Fraga A, Carrasquinho E, de Oliveira EC, et al. Expert involvement predicts mhealth app downloads: multivariate regression analysis of urology apps. JMIR Mhealth Uhealth 2016 Jul 15;4(3):e86 [FREE Full text] [doi: 10.2196/mhealth.5738] [Medline: 27421338]
38. Carter T, O'Neill S, Johns N, Brady RR. Contemporary vascular smartphone medical applications. Ann Vasc Surg 2013 Aug;27(6):804-809. [doi: 10.1016/j.avsg.2012.10.013] [Medline: 23535521]
39. Subhi Y, Bube SH, Bojsen SR, Thomsen AS, 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]
40. Baghaei N, Nandigam D, Casey J, Direito A, Maddison R. Diabetic Mario: designing and evaluating mobile games for diabetes education. Games Health J 2016 Aug;5(4):270-278. [doi: 10.1089/g4h.2015.0038] [Medline: 27304882]
41. Lazem S, Webster M, Holmes W, Wolf M. Games and diabetes: a review investigating theoretical frameworks, evaluation methodologies, and opportunities for design grounded in learning theories. J Diabetes Sci Technol 2015 Sep 2;10(2):447-452 [FREE Full text] [doi: 10.1177/1932296815604634] [Medline: 26337753]

42. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
43. 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]
44. Lipczynska S. Psyberguide: a path through the app jungle. J Ment Health 2019 Feb;28(1):104. [doi: 10.1080/09638237.2017.1417574] [Medline: 29265897]
45. Health On the Net. HONcode Certification URL: https://www.hon.ch/HONcode/ [accessed 2019-01-26] [WebCite Cache ID 75hp8B3Fu]
46. Khazaal Y, Chatton A, Zullino D, Khan R. HON label and DISCERN as content quality indicators of health-related websites. Psychiatr Q 2012 Mar;83(1):15-27. [doi: 10.1007/s11126-011-9179-x] [Medline: 21547515]
47. Practical Apps: Home. Our Approach URL: https://practicalapps.ca/approach/ [accessed 2019-01-26] [WebCite Cache ID 75hpBrrdl]
48. Deshpande A, Jadad AR. Trying to measure the quality of health information on the internet: is it time to move on? J Rheumatol 2009 Jan;36(1):1-3. [doi: 10.3899/jrheum.081101] [Medline: 19208527]
49. 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]
50. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 9;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
51. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
52. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013 Dec 30;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]
53. 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]
54. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]

Abbreviations

eHealth: electronic health
MM: medication management

Edited by C Dias; submitted 04.02.19; peer-reviewed by E Alagoz, J Ford Ii, A Gulliver; comments to author 25.03.19; revised version received 31.05.19; accepted 22.06.19; published 11.09.19.

Please cite as:
Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, Nikoo M, Vigo D, Jang K, Demlova R, Krausz M
Mobile Apps for Medication Management: Review and Analysis
JMIR Mhealth Uhealth 2019;7(9):e13608
URL: http://mhealth.jmir.org/2019/9/e13608/
doi:10.2196/13608
PMID:31512580

Original Paper

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:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/4/e18411/

Abstract

Background: The efficacy of smartphone apps for improving medication adherence in type 2 diabetes is not well studied in Asian populations.

Objective: This study aimed to determine the feasibility, acceptability, and clinical outcomes of using a smartphone app to improve medication adherence in a multiethnic Asian population with type 2 diabetes.

Methods: We block randomized 51 nonadherent and digitally literate patients with type 2 diabetes between the ages of 21 and 75 years into two treatment arms (control: usual care; intervention: usual care+Medisafe app) and followed them up for 12 weeks. Recruitment occurred at a public tertiary diabetes specialist outpatient center in Singapore. The intervention group received email reminders to complete online surveys monthly, while the control group only received an email reminder(s) at the end of the study. Barriers to medication adherence and self-appraisal of diabetes were assessed using the Adherence Starts with Knowledge-12 (ASK-12) and Appraisal of Diabetes Scale (ADS) questionnaires at baseline and poststudy in both groups. Perception toward medication adherence and app usage, attitude, and satisfaction were assessed in the intervention group during and after the follow-up period. Sociodemographic data were collected at baseline. Clinical data (ie, hemoglobin A_{1c}, body mass index, low-density lipoprotein, high-density lipoprotein, and total cholesterol levels) were extracted from patients' electronic medical records.

Results: A total of 51 (intervention group: 25 [49%]; control group: 26 [51%]) participants were randomized, of which 41 (intervention group: 22 [88.0%]; control group: 19 [73.1%]) completed the poststudy survey. The baseline-adjusted poststudy ASK-12 score was significantly lower in the intervention group than in the control group (mean difference: 4.7, *P*=.01). No changes were observed in the clinical outcomes. The average 12-week medication adherence rate of participants tracked by the app was between 38.3% and 100% in the intervention group. The majority (>80%) of the participants agreed that the app was easy to use and made them more adherent to their medication.

Conclusions: Our feasibility study showed that among medication-nonadherent patients with type 2 diabetes, a smartphone app intervention was acceptable, improved awareness of medication adherence, and reduced self-reported barriers to medication adherence, but did not improve clinical outcomes in a developed Asian setting.

(JMIR Mhealth Uhealth 2019;7(9):e14914) doi:10.2196/14914

KEYWORDS

smartphone apps; mobile phone apps; medication adherence; type 2 diabetes; feasibility trial; pilot study

Introduction

Medication nonadherence is a complex, costly, and multidimensional problem that involves the patient, his/her health care provider, and the process of taking/using the medication [1]. Patient education, medication management, reminders, and incentives to promote adherence are interventions that have been successful in improving medication adherence worldwide [2]. Despite measures to improve medication adherence, approximately one-third to half of the people with diabetes are still not adherent to their medication [3,4]. People with type 2 diabetes have poorer medication adherence if they do not believe in the safety and efficacy of the medication, which is common in asymptomatic diseases [5]. Poor adherence to diabetes medication results in suboptimal glycemic control [6,7], which increases the risk of diabetes-related complications [8,9], leading to more hospitalization and emergency department visits [10,11].

Smartphone apps are increasingly used as a complementary tool for diabetes self-management (which includes medication management) in recent years. A pooled analysis on the effect of smartphone apps for diabetes self-management found an overall 0.5% reduction in hemoglobin A_{1c} (HbA_{1c}) levels [12]. Despite emerging positive evidence on the efficacy of apps in diabetes self-management [13,14], gaps exist in the utility of apps' features in meeting users' needs [15-17]. There is a paucity of studies on the efficacy and implementation of smartphone apps in supporting medication taking [12], with only a small number of randomized controlled trials investigating medication adherence in people with high blood pressure [18,19]. Furthermore, diabetes and medication adherence app interventions are not well studied in Asian populations. Asians constitute 60% of people with diabetes globally and are likely to have different cultural beliefs toward disease and medication management [20,21]. This represents missed opportunities to benefit up to 250 million people with diabetes [20]. Given the acceleration of mobile connectivity in the Asia Pacific region in recent years [22], it is important to investigate the receptivity and usage of apps for diabetes medication management in Asian populations with high mobile penetration.

Population-based interventions involving smartphone apps are often complex and multifaceted due to their challenges in controlling the environment [23]. These challenges are amplified when population characteristics are not well understood. In view of the challenges with evaluating complex health interventions, a feasibility and piloting phase to optimize study design and evaluation is warranted [23,24]. We aimed to determine the feasibility, effectiveness, acceptability, and clinical outcomes of using a smartphone app to improve medication adherence in a multiethnic Asian population with type 2 diabetes through a pilot study. Our objectives were to assess the recruitment rate, changes in self-reported barriers to medication adherence, diabetes-related health outcomes, app usage behavior, and satisfaction levels. We referred to the Consolidated Standards of Reporting Trials (CONSORT) eHealth checklist for feasibility trials [25] and the mobile health evidence reporting and assessment (mERA) checklist for mobile health [26] to report the findings of our study.

Methods

Study Design

We used a randomized two-arm pre-posttest control group design with a 12-week follow-up period. All participants received usual care, while the intervention group participants additionally downloaded and used the Medisafe app [27] on their personal smartphones during the study.

Study Setting

Participants were recruited over 10 weeks from September to November 2018 at a tertiary diabetes specialist outpatient center, which is part of a 1000-bed public hospital in (the Eastern region of) Singapore. The center serves subsidized and private patients and nonresidents of Singapore. Patients were self-referred or referred from primary care general practitioners, other departments in the same hospital, or other hospitals. Usual care provided by the center comprises clinic appointments every 3-6 months. At each clinic appointment, patients have their blood pressure and body weight taken, undergo blood tests to monitor their blood glucose and lipid levels, review diabetes management with their endocrinologist, and collect their prescribed medications from the hospital pharmacy. Consultations with the podiatrist, dietitian, or other specialists (ie, ophthalmologist, cardiologist, and renal specialist) were arranged on an ad hoc basis (ie, usually once a year for foot and eye examination). Patients are expected to self-manage their diabetes (following

their treatment plan) outside the hospital setting between these scheduled clinic appointments.

Singapore has one of the highest smartphone penetration rates in the world, with 150% mobile subscriptions (one person with two or more mobile subscriptions) and 85% smartphone ownership [28,29].

Participant Recruitment and Eligibility Criteria

Potential participants were referred by four endocrinologists using a recruitment pamphlet. To be referred by the endocrinologist, participants were at or above the age of 21 years (the legal age for study consent in Singapore), diagnosed with type 2 diabetes according to the American Diabetes Association guidelines, on insulin or oral hypoglycemic agents, and English speakers.

Participants were excluded from the study if they were pregnant, cognitively impaired or diagnosed with psychological issues, prisoners, diagnosed with type 1 diabetes, bed bound and undergoing tube feeding, or prescribed medication for the first time.

Referred patients who consented to participate in the study were asked to complete a baseline questionnaire, which also served as a screening tool to identify eligible patients for randomization. To prevent the "ceiling effect," participants who were adherent to their medications were screened out of the study. Participants were considered nonadherent to their medication if they answered, "Strongly Agree" or "Agree" to the question, "I forget to take my medicines some of the time" or if they answered "In the last week/month/3 months" to the question, "Have you taken a medicine more or less often than prescribed?"(ask-12-Q8) in the Adherence Starts with Knowledge-12 (ASK-12) questionnaire [30]. To screen participants who were not digitally literate, participants must have responded "Yes" to the question, "Have you used any phone apps in the past 2 weeks?" Lastly, to screen participants who were already using an app to manage their medication, participants must have responded "No" to the question, "Have you used any smartphone app to manage your medications in the past 2 weeks?"

Hence, secondary inclusion criteria for randomization into the study were self-reported medication nonadherence, digital literacy, and nonuse of a medication management app in the past 2 weeks.

Study Procedures

Patients with type 2 diabetes attending their scheduled clinic appointments, who met the referral inclusion criteria, were referred to the researchers by their endocrinologist. Interested patients proceeded to provide informed consent. The patient was termed a research participant once the informed consent document was signed. At the point of consent seeking, researchers explained to potential participants that they may or may not be selected for the study, depending on their eligibility, which can only be determined after they respond to the baseline questionnaire. Informed consent was collected with printed hardcopy forms.

Study data were collected and managed using Research Electronic Data Capture (REDCap) by Nanyang Technological University [31,32]. 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. Upon receiving informed consent, REDCap generates a unique survey code linked to each participant's predefined email for the baseline and subsequent follow-up surveys. All data in REDCap are deidentified apart from participants' email address. Baseline data were collected using an iPad on the day of recruitment prior to randomization.

Intervention

The intervention group participants were asked to download and use the Medisafe app to help them manage their medications for 12 weeks. Participants were assisted by the researchers to download the Medisafe app on their personal smartphone, to set their medication schedule and reminder on the app, and to use the app. Participants were asked to use the app freely outside the health care setting and add the research group as a "Medfriend" for their medication-taking patterns to be observed.

Medisafe is a commercial, free medication management app available on both Android and iOS platforms. Its features include medication scheduling, reminder, tracking, data sharing, and medication adherence assessments. We selected a commercial app with evidence supporting its effectiveness [13,33] to assess the feasibility of a smartphone app in promoting medication adherence in patients with type 2 diabetes.

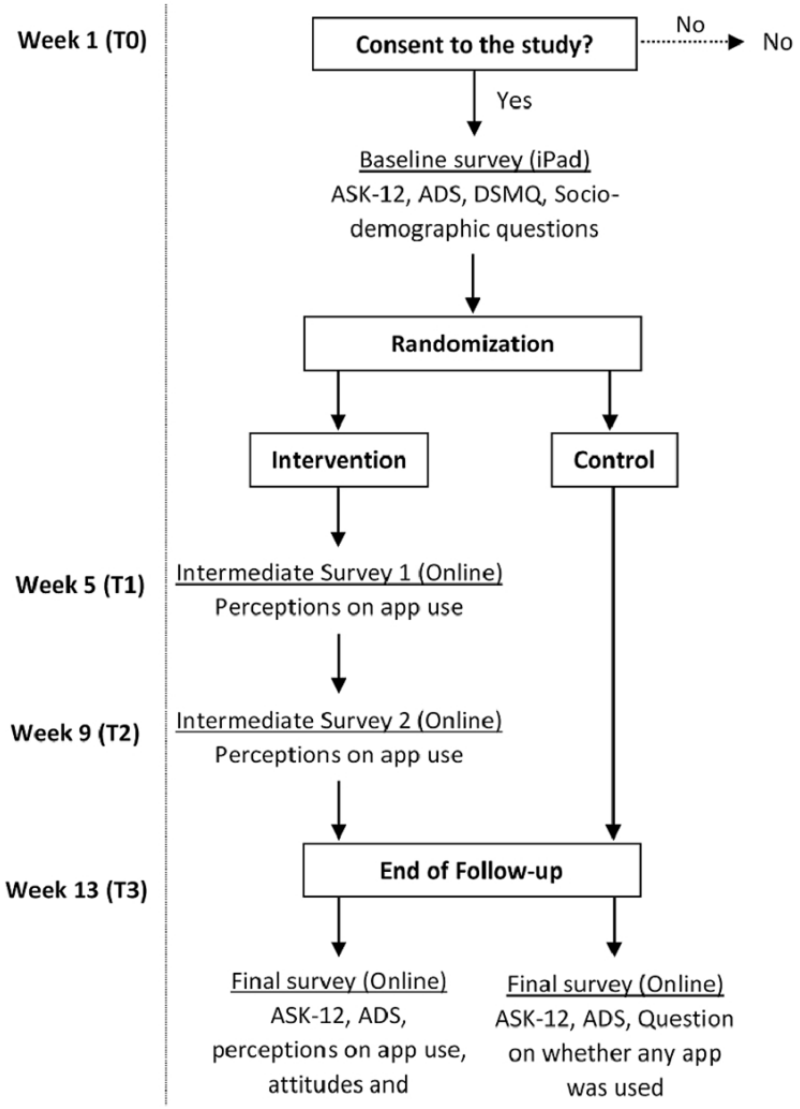
The intervention group participants were reminded via email to complete two intermediate and one final online survey at 4-week intervals during the 12-week follow-up period. Control group participants were instructed to complete only one online survey at the end of the 12-week follow-up period. All follow-up surveys were conducted online via a unique link sent to the participants' email address. Each unique survey link was accessible for a maximum of 14 days or until the participant completed the survey. Participants in both groups were reminded by calling them on their mobile phone to complete the final survey if no response was received a week after the survey was sent out. Participants were given supermarket vouchers on completion of each online survey. Voucher rewards were consolidated and collected from the diabetes center by participants at the end of the study. We collected some participant feedback with the online satisfaction survey and while handing out vouchers to participants who completed the online survey(s).

Outcomes

Primary outcomes were the feasibility, effectiveness, and acceptability of using a smartphone app to improve medication adherence in a multiethnic Asian population with type 2 diabetes (Figure 1). Feasibility was determined from the recruitment/enrolment rate (percentage of people who reject the study, ie, the number of patients who consented to the study divided by the total number of clinic sessions). Another measure of feasibility is adherence to trial participation, which was

assessed by observing intervention group participants' interaction with the app throughout the intervention through the "Medfriend" feature of the app. The research team, as a "Medfriend," did not interact with participants during the follow-up period. Reports on the medication-taking status of participants were generated at the end of the intervention (T3) through the app.

Figure 1. Schedule of outcome measurements. ASK-12: Adherence Starts with Knowledge-12, ADS: Appraisal of Diabetes Scale, DSMQ: Diabetes Self-Management Questionnaire.



Effectiveness was measured with self-reported barriers to medication adherence and assessed at baseline (T0) and poststudy (T3) in both groups by using the ASK-12 questionnaire [30]. The Appraisal of Diabetes Scale (ADS) [34] was administered concurrently with ASK-12 to account for changes (if any) in self-appraisal of diabetes. Acceptability of app use intervention was determined by self-reported perceptions, attitudes, and satisfaction in using the app. Perception toward medication adherence and app usage were assessed at all three time points (T1, T1, and T3), while attitude and satisfaction were only assessed poststudy (T3). Control group participants were asked for the app they used to manage their medication(s) (if any) in the past 3 months in order to assess the level of contamination in the control arm.

Secondary outcomes were diabetes-related health outcomes. Data for assessing secondary outcomes such as anthropometric measures, blood glucose level, and lipid measurements were extracted from clinical records. The following data were also collected for participant profiling and baseline adjustments: data on medications and history of diabetes-related complications from clinical records; sociodemographic data; and responses from a 16-item Diabetes Self-Management Questionnaire (DSMQ) [35] collected at baseline (T0).

Sample Size

A minimum of 12 participants per treatment arm is necessary to assess the objectives of the study in a two-arm trial [36], and 25 participants per arm is sufficient to account for a dropout rate of about 40% [37]. Therefore, we aimed to recruit and randomize a minimum of 25 participants per arm in 10 weeks of recruitment.

Randomization

Block randomization (blocks of four) was conducted to ensure a balanced allocation, since we could not anticipate the final sample size. Eligible participants were asked to draw a card from a box with two "intervention" and two "control" cards, which were reset after all four cards were drawn.

Blinding

The clinical care team was blinded from the study. Participants were only partially blinded, as we had to explain the purpose of the study before randomization. The name of the app was not revealed to participants unless they were randomized into the intervention group or screened out of the study.

Data Analyses

The intention-to-treat approach was used to analyze the data. We excluded participants who did not complete the final survey due to the lack of poststudy data for pre-post comparison. Intervention group participants who stopped using the app during the study and control group participants who used an app to manage their medications during the study follow-up period were included in the analysis. Scores for the ASK-12, ADS, and DSMQ surveys were computed in accordance to the method suggested by the original authors [30,34,35]. Descriptive analyses were used for baseline comparisons, and linear regressions, controlled for baseline imbalances, were used to compare the pretest and posttest change scores. All statistical assumptions were checked to ensure the accuracy of analyses. Statistical significance was set at $P<.05$. SPSS (version 22; IBM Corp, Armonk, NY) was used for all statistical analyses.

Ethical Considerations

This study was approved by the SingHealth Centralised Institutional Review Board (Reference: 2018/2563) and the Nanyang Technological University Institutional Review Board (Reference: IRB-2018-09-029) in Singapore. Licenses and permission to use published questionnaires were obtained from the original authors and relevant institutions prior to data collection. We did not prospectively register the trial, as this was a feasibility study.

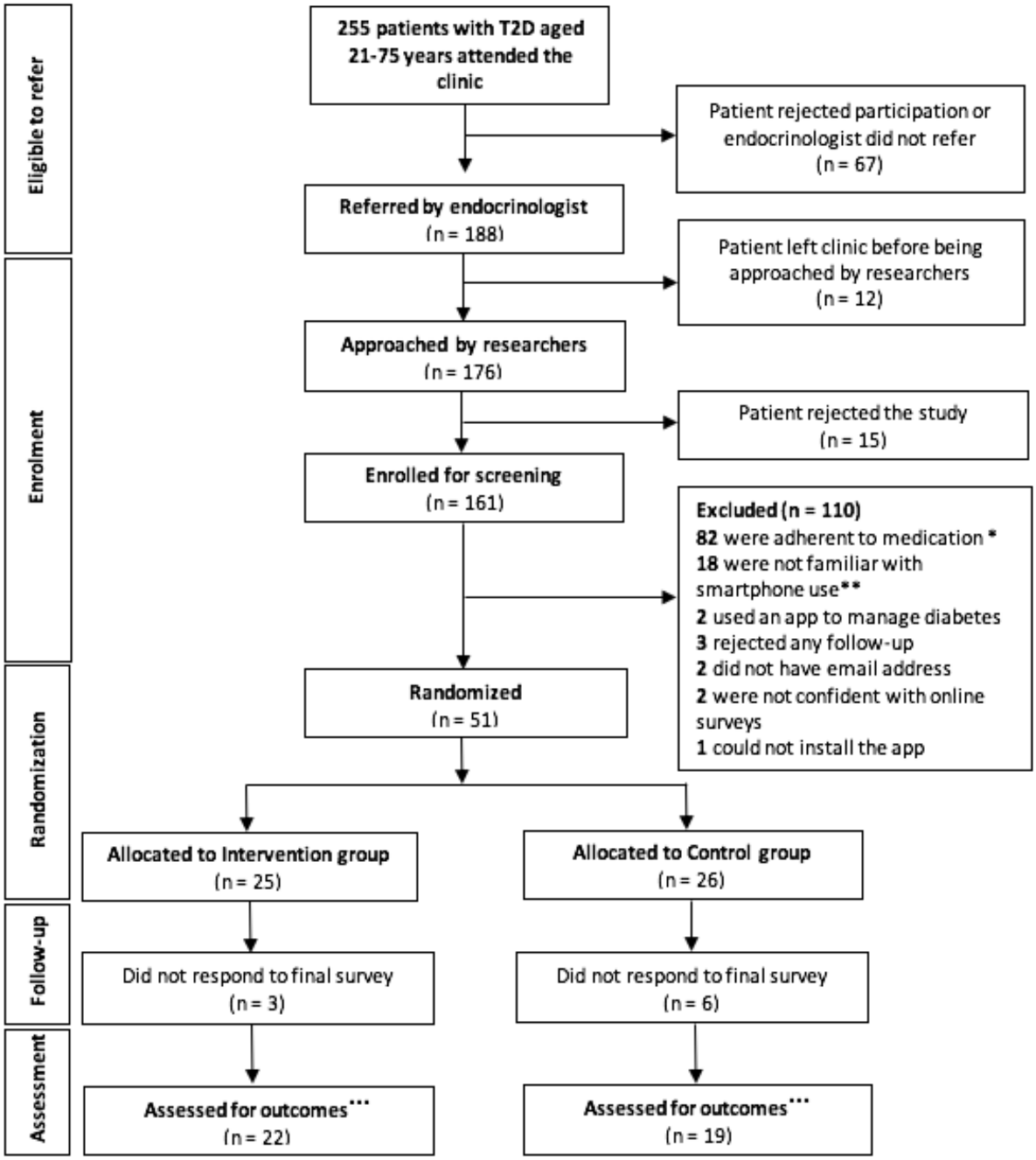
Results

Recruitment

A total of 176 patients were referred and approached for recruitment over 48 three-hour clinic sessions. Overall, 15 patients (8.5%) rejected study participation, which yielded an enrolment rate of approximately 3 (161/48) patients per clinic session. Reasons for rejecting study participation included concerns over the collection of personal data, pressed for time, and refusal to complete the baseline survey. Of the 161 enrolled participants, 110 were not eligible for randomization: 82 (50.9%) self-reported that they were adherent to their medications; 18 (11.2%) were not familiar with smartphone use; 7 (4.3%) refused participation, did not have an email address, or were not confident with completing the online surveys; 2 (1.2%) were already using a smartphone app to complement diabetes management; and 1 (0.6%) could not install the app.

A total of 51 (31.7%) participants met the inclusion criteria and were randomized to the intervention (n=25) or control (n=26) group, of which 22 (88.0%) and 19 (73.1%) in the intervention and control group, respectively, completed the postintervention survey (Figure 2). Three intervention group participants (3/22) indicated that they stopped using the app, and two control group participants (2/19) indicated that they used a diabetes self-management app during the follow-up period.

Figure 2. Diagram of participant flow. T2D: type 2 diabetes. *Participants were considered adherent if they answered "disagree/neutral" to the question, "I forget to take my medicines some of the time" or any option within 3 months to the question, "Have you taken a medicine more or less often than prescribed?" in the Adherence Starts with Knowledge-12 questionnaire. **Patients who were not confident of using a new app. ***Three intervention group participants stopped using the app; two control group participants started using an app to manage diabetes during the follow-up period.



Randomization

The baseline characteristics of patients included in the analysis are shown in Table 1. Randomization was successful, as there are no statistically significant differences at baseline between groups for sociodemographic and clinical characteristics (eg, blood test results, diabetes-related complications, and anthropometric measurements) and baseline questionnaires (eg, DSMQ and ADS), apart from the number of years with diabetes and the pretest total ASK-12 score. Control group participants lived on an average of 7 years more with diabetes (P=.005) and had a lower total ASK-12 score (intervention group: 28.6; control group: 25.5; P=.044) compared with the intervention group. Higher ASK-12 scores represent higher barriers to medication adherence.

Table 1. Baseline characteristics of patients included in the analyses.

Characteristics	Intervention group (n=22)	Control group (n=19)	*P* value
Sociodemographic characteristics			
Age, median (min-max)	51.5 (22-69)	52 (28-67)	.85[a]
Sex, n (%)			.28
Male	9 (40.9)	11 (57.9)	
Female	13 (59.1)	8 (42.1)	
Ethnicity, n (%)			.26
Chinese	10 (45.5)	12 (63.2)	
Non-Chinese	12 (54.5)	7 (36.8)	
Highest education, n (%)			.49
Secondary school and below	11 (50.0)	6 (31.6)	
Junior college/diploma	4 (18.2)	5 (26.3)	
University	7 (31.8)	8 (42.1)	
Housing (number of rooms), n (%)			.52
≤3	2 (9.1)	4 (21.1)	
4-5	12 (54.5)	10 (52.6)	
≥5	8 (36.4)	5 (26.3)	
Household income (US $), n (%)			.17
<4000	6 (30.0)	9 (47.4)	
4000-6999	4 (20.0)	6 (31.6)	
≥7000	10 (50.0)	4 (21.1)	
Clinical characteristics			
Number of years with diabetes, median (SD)	11.1 (7.1)	18.3 (8.4)	.005[b]
Number of different types of medications, median (min-max)	4 (1-9)	4 (1-13)	.47[a]
Type of medications, n (%)			
Insulin	7 (31.8)	9 (47.4)	.31
Antihypertensive medication	11 (50.0)	5 (26.3)	.12
Cholesterol-lowering medication	8 (36.4)	5 (26.3)	.49
Medication intensity, n (%)			
Oral medications only	7 (31.8)	9 (47.4%)	.051
Insulin only	0 (0.0)	3 (15.8%)	
Oral and insulin	15 (68.2)	7 (36.8%)	
Anthropometric data, median (min-max)			
Body mass index	28.7 (20.2-49.2)	28.3 (21.1-35.6)	.66[b]
Diabetes-related complications, n (%)			
Proliferative diabetic retinopathy	4 (18.2)	3 (15.8)	>.99[c]
Peripheral vascular disease	2 (9.1)	3 (15.8)	.65[c]
Chronic kidney disease (≥stage 3)	3 (13.6)	4 (21.1)	.70[c]
History of major cardiovascular events	5 (22.7)	3 (15.8)	.69[c]
Blood glucose level, median (min-max)			
Hemoglobin A_{1c} (%), preintervention	8.2 (5.9-14.8)	8.5 (6.4-11.8)	.57[a]

Characteristics	Intervention group (n=22)	Control group (n=19)	*P* value
Lipid profile, median (min-max)			
Low-density lipoprotein (mmol/L)	2.7 (2.0-6.6)	2.4 (1.3-4.3)	.30[a]
High-density lipoprotein (mmol/L)	1.1 (0.9-1.7)	1.0 (0.7-2.0)	.09[a]
Total cholesterol (mmol/L)	4.1 (3.2-8.2)	4.1 (2.5-6.9)	.56[a]
Baseline questionnaires			
Appraisal of Diabetes Scale, mean (SD)[d]			
Total score (baseline)	19.7 (3.7)	19.0 (3.8)	.57
Diabetes Self-Management Scale score, mean (SD)[e]			
Total score (baseline)	2.0 (0.4)	2.0 (0.3)	.69
Adherence Starts with Knowledge-12 medication adherence barrier survey, median (SD)[f]			
Total score (baseline)	28.6 (5.2)	25.5 (4.4)	.04[b]

[a]*P*<.05.

[b]Mann-Whitney *U* test for continuous variables.

[c]Fisher exact test for categorical variables with small sample sizes.

[d]Scores (min=7, max=35) are summed up (questions 2 and 6 are reverse scored). Lower scores signify more positive appraisal of diabetes.

[e]Scale scores are computed (min=0, max=4), as there are responses that cannot be scored (eg, “Not part of my treatment”). Items 5, 7, 10, 11, 12, 13, 14, 15, and 16 are reverse scored. Scale scores can be computed as Total_Sum(All)/(16-missing). Higher scores signify better diabetes self-management.

[f]Scores are summed up with reverse scoring for Inconvenience (questions 1-3) and Behavior (questions 8-12). Higher scores signify higher barriers to adherence.

Outcomes

The mean ASK-12 (adherence barrier) score decreased in the intervention group but increased in the control group. Higher ASK-12 scores signify higher barriers to medication adherence. After baseline adjustment with “years with diabetes” and “baseline ASK-12 score,” the ASK-12 pre-post “change score” was statistically significant (*P*=.01), with the intervention group having a 4.7-point (1.2-8.2) lower mean score than the control group (Table 2).

There were no statistically significant mean differences between groups for baseline-adjusted regression in ADS score, HbA_{1c}, lipids, and body mass index (Table 2). Although the mean HbA_{1c} level increased slightly in both groups, the intervention group participants had an average of 0.5% lower increment compared with the control group.

Table 2. Adjusted mean differences between treatment groups.

Outcome measure	Intervention		Control		Adjusted mean difference (95% CI)[a]	*P* value
	Baseline	Poststudy	Baseline	Poststudy		
Self-reported questionnaires, mean (SD)						
Number of participants	22	22	19	19	N/A[b]	N/A
Adherence Starts with Knowledge-12 scale score[c]	28.6 (5.2)	27.2 (5.8)	25.5 (4.4)	28.5 (7.0)	–4.73 (–8.26 to –1.21)	.01[d]
Appraisal of Diabetes scale score[e]	19.7 (3.7)	19.4 (3.5)	19.0 (3.8)	19.4 (4.3)[f]	–0.48 (–1.82 to 2.78)	.43[g]
Clinical measurements						
Blood glucose level						
Number of participants	22	19	19	15	N/A	N/A
Hemoglobin A_{1c} (%)	8.7 (2.4)	9.0 (1.6)	8.6 (1.5)	9.4 (2.4)	–0.42 (–1.89 to 1.06)	.57[g]
Lipids						
Number of participants	21	17	19	12	N/A	N/A
Low-density lipoprotein (mmol/L), mean (SD)	3.1 (1.2)	3.1 (0.7)	2.7 (1.0)	2.7 (0.8)	0.11 (–0.20 to 0.06)	.75[g]
High-density lipoprotein (mmol/L), mean (SD)	1.2 (0.3)	1.2 (0.3)	1.1 (0.3)	1.2 (0.3)	–0.09 (–0.56 to 0.77)	.14
Total cholesterol (mmol/L), mean (SD)	4.5 (1.2)	4.6 (0.8)	4.2 (1.0)	4.1 (1.1)	–0.02 (–0.69 to 0.72)	.052[g]
Anthropometric data						
Number of participants	22	18	19	13	N/A	N/A
Body mass index (kg/m^2), mean (SD)	29.4 (7.3)	25.2 (12.5)	28.0 (4.0)	27.5 (4.2)	0.02 (–1.13, 1.10)	.98[g]

[a]Adjusted variables for linear regressions: mean baseline ASK-12 score, years with diabetes, baseline of outcome variable.

[b]N/A: not applicable.

[c]Scores are summed up with reverse scoring for Inconvenience (questions 1-3) and Behavior (questions 8-12). Higher scores signify higher barriers to adherence.

[d]$P<.05$

[e]Scores (min=7, max=35) are summed up (questions 2 and 6 are reverse scored). Lower scores signify more positive appraisal of diabetes.

[f]One missing value, n=18.

[g]Normality assumption is violated due to a small number of outliers and small sample sizes per group.

Adherence to Trial Participation

Three intervention group participants did not complete the final survey, of which, two had intermittent app usage, while one did not use the app from the start. Three other participants who completed the final survey indicated that they stopped using the app between 2 weeks and 2 months into the study, as they did not find the app useful or found it distracting. Two participants who indicated that they were still using the app at the end of the study did not have their medication-taking status tracked, as they were unfamiliar with the app-based medication-logging process. The average individual 12-week medication adherence rate tracked by the app was 38.3%-100% for the remaining 17 participants. 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.

The medication adherence rates tracked by the app also reflect the app usage patterns of the participant. Despite differences in app usage patterns between participants, the aggregated weekly medication adherence tracked by the app did not fall below 50% over the 12 weeks (Figure 3). The graphs in Figure 3 show actual examples of one aggregated and three typical app usage patterns observed in the participants. Medication adherence and health outcomes improved for Participant W who was still occasionally nonadherent to the medication but highly adherent to app usage. Participant X had waning app usage, as perception of the app became less positive over time. Medication adherence and health outcomes did not improve, as participant X ran out of medication in week 7. Several participants exhibited similar cyclical app usage behavior to Participant Y where medication adherence increases when they receive emailed survey reminders. This cyclical pattern was also observed in the aggregated weekly medication adherence tracked by the app.

Acceptability of the Medication Management Smartphone App

The perception, attitude, and satisfaction of app use (Table 3) show the acceptability of a smartphone app in supporting medication management in the feasibility trial. These surveys were related to participants' experiences in app use and therefore only administered to the intervention group.

Figure 3. Weekly medication adherence over 12 weeks, extracted from participants' "Medisafe" reports.

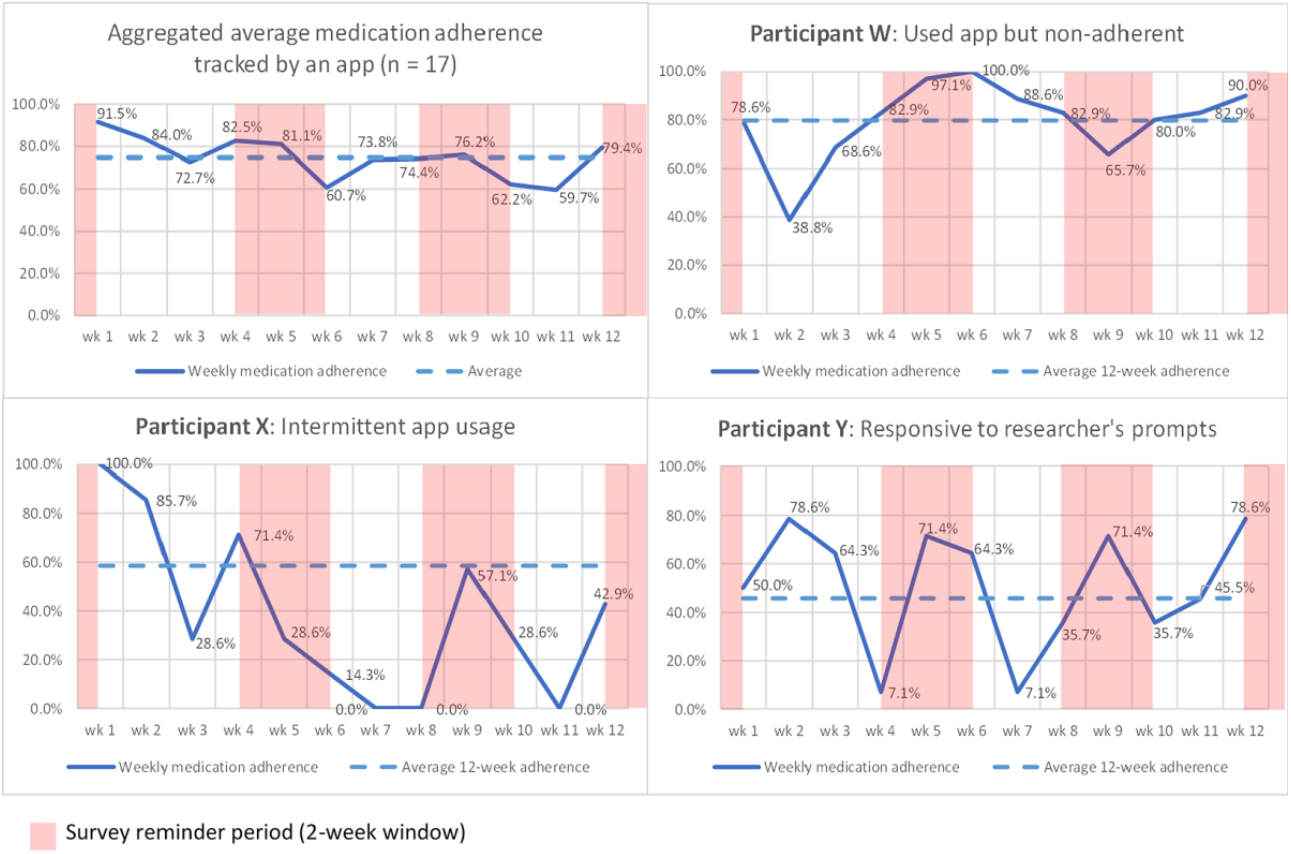


Table 3. Perception, attitude, and satisfaction of app use in the intervention group.

Survey topic	Value
Perception of app usage[a,b] , n (%)	
Made you more aware of your adherence to medication (Agree; n=21)	19 (90.5)
Made you more adherent to your medication (Agree; n=21)	17 (81.0)
Made you more confident in managing your medication (Agree; n=21)	17 (81.0)
Reduces the stress in managing your medication (Agree; n=21)	14 (66.7)
Is easy to use (Agree; n=22)	20 (90.9)
Annoys you when the notification goes off (Agree/neutral; n=20)	16 (80.0)
Attitude toward app use, n (%)	
Would you recommend Medisafe to another person with the same condition? (Yes)	21 (95.5)
Would you trust your doctor to recommend an app for you to manage your condition? (Yes)	21 (95.5)
Will you continue to use the Medisafe app after today? (Yes)	19 (86.4)
Satisfaction, median (min-max)	
On a scale of 1 to 10, with 10 being very satisfied, how would you rate your experience in using an app for managing your medication?	8 (1-10)

[a]There is a "Not applicable" option for "Perception on app usage" questions, which caused the denominator to differ.

[b]In response to the question, "Thinking about the past few days, how far do you agree that the app?".

Perception of App Usage

The perception of app usage was generally positive among respondents, with the majority (>80%) agreeing that the app made them more aware of the importance of medication adherence, more confident in managing their medication, and more adherent to their medication. For 90.9% of the respondents, the app was easy to use. However, use of the app did not reduce medication management stress in 34% of the respondents, and

80% of the respondents found the reminder notification annoying.

Attitude Toward App Use

The attitude toward app use was generally positive, with 95.5% of the respondents answering "Yes" to recommending the app to another person with the same condition and trusting their doctor to recommend an app for them to manage type 2 diabetes. The majority of respondents (86.4%) indicated they would continue to use the app after the study.

Satisfaction

General satisfaction was high, with a median score of 8 on a scale of 1-10. Participants who stopped using the app provided lower scores. For example, one participant who stopped using the app provided a score of 1/10.

Participant Feedback

Two participants would have liked to add their spouses as a "Medfriend" but could not do so, as the free version only allowed the addition of one "Medfriend" (ie, the study team). Other feedback include suggestions to incorporate the doctor's appointment scheduling and other diabetes self-management features, simplifying the app interface, educating participants on manipulating the settings, and integrating some of the hospital's services with the app. Although Medisafe is a third-party app, patients would prefer integrating all health services into a one-stop reliable and personalized platform.

Discussion

Principal Findings

We established the feasibility of using a smartphone app to improve medication adherence in patients with type 2 diabetes managed at a public diabetes specialist outpatient center in Singapore through a pilot study. The medication nonadherence rate determined by the study (49.1%) falls within the range of rates reported by other studies in Singapore and internationally using a variety of measurement tools [6,38]. We observed significantly lower self-reported barriers to medication adherence in the intervention group compared with the control group but no improvement in the HbA_{1c} level. This concurs with the findings of a similar US study, which observed improvement in self-reported medication adherence but no change in blood pressure over 12 weeks [18].

The control group had slightly lower HbA_{1c} level, barriers to medication adherence, and more positive appraisal of diabetes at baseline compared with the intervention group. This observation was reversed 12 weeks later when the intervention group had slightly better outcomes in all three measurements. Improvement in barriers to medication adherence in the intervention group is likely attributed to medication-taking reinforcements by the app and monthly email reminders to complete the online surveys. Adherence reinforcements will likely lead to short-term improvement in medication adherence [39].

We observed increased HbA_{1c} levels in both groups, which is attributed to the follow-up period falling within a few holiday seasons (ie, Diwali, Christmas, and Chinese New Year) where festive feasting in Asian cultures (ie, Singapore) is likely [40]. A different intervention period may change the study outcomes, although we acknowledge that the degree of medication nonadherence, personal motivation, and response to treatment can affect the HbA_{1c} levels and add complexity to the interpretation of outcomes [41].

We observed various factors that influenced study feasibility. First, physician advocacy is important in encouraging the uptake of new health interventions. The majority (>85%) of patients referred by their endocrinologists were willing to provide informed consent and complete the baseline questionnaire. Most of the intervention group participants also indicated that they would trust their doctor to recommend an app to manage their condition. Second, the use of digital data collection tools (ie, REDCap) minimized data entry errors and human resources required for data collection.

Third, participants' digital literacy and the app's usability influence adherence to the intervention and satisfaction. Many older participants have difficulty adjusting the app settings, which caused the reminders to become a distraction instead. Fourth, reasons for medication nonadherence affect study feasibility and outcomes. For people with polypharmacy, an app may help to better organize medication-taking schedules. However, this does not solve barriers to medication adherence such as the inconvenience of taking multiple medications, medication side effects, or fear of injections. Lastly, the health-seeking behavior of participants will influence the study outcomes. For example, one motivated participant in the control group started using an app for diabetes management during study follow-up and achieved >0.5% HbA_{1c} improvement in 12 weeks.

There were limitations to the study. We were unable to observe app usage patterns of a few participants who changed smartphones during study follow-up. Medication adherence rates in the control group were also not tracked for comparison. Self-reported tools are subjective to a patient's own judgement and social desirability bias; hence, actual medication adherence may not be accurately reflected. We observed patients who over- and underreported their medication adherence status and problems with survey interpretation. For example, when the researchers verbally asked (at baseline), "How likely do you think your diabetes will worsen in the next few years?" a few participants answered "I hope it will not worsen" instead of choosing a Likert scale response. The study may not be generalizable to all people with diabetes, as tertiary specialist outpatient clinics are likely to manage more complex cases that cannot be managed in the primary care setting. Lastly, contamination may have occurred when the control group participants were exposed to the idea of using an app for type 2 diabetes medication management.

This study allowed us to better understand the impact of a health app on patients with type 2 diabetes and identify potential problems that could occur before scaling up the study. One registered trial using a self-developed smartphone app to improve the 6-month medication adherence among patients with type 2 diabetes in Singapore was withdrawn due to poor

patient recruitment [42]. Therefore, we conducted a pilot trial with a commercial app to first evaluate factors that are important for implementing a full trial. Our findings suggest that should a full randomized controlled trial be conducted, a five-fold scale-up is required to achieve full trial power under the same conditions. This can be achieved with the involvement of more physicians, more study sites, or a longer recruitment period. Future studies should assess factors that could enhance the usability of apps in older adults who are less technologically savvy. The app usage behavior of different patient subgroups and interaction between various diabetes app features can also be explored.

Conclusions

Our feasibility study found that a smartphone app intervention for medication nonadherent patients with type 2 diabetes in a developed Asian setting is feasible and acceptable, improved awareness of medication adherence, and reduced self-reported barriers to medication adherence. Digital literacy, health-seeking behavior, app usability, and the time period of the intervention are factors that influenced feasibility.

Acknowledgments

This study was funded by Nanyang Technological University, Singapore, through (1) the Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore; (2) the Institute for Health Technologies (HealthTech NTU), Nanyang Technological University, Singapore; and (3) Changi General Hospital, Singapore.

We wish to thank Dr Tay Tunn Lin, Dr Soh Shui Boon, Dr Khoo Joo Ching Joan, and the staff of Changi General Hospital Diabetes Center for their support in study recruitment. We would also like to acknowledge Ms My Linh Thai and Ms Audrey Nah for their help in participant recruitment and data collection, and Medisafe for providing technical support during the study.

Authors' Contributions

ZH conceptualized and set up the study, created the online surveys, trained the data collectors, collected and analyzed data, and drafted and revised the manuscript. ET co-conceptualized the study, supported study implementation, referred patients, supported extraction of clinical records data, and provided critical review of the draft manuscript. EL provided critical input into the conceptualization of the study, supported study implementation, and revised the manuscript. JC provided mentoring to ZH, obtained funding for the data collectors, provided critical input into all stages of the study, and critical review of the draft manuscript. BB and PS provided critical input into the study and critical review of the draft manuscript. All authors reviewed and approved the final version of the manuscript prior to submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File)325 KB - mhealth_v7i9e14914_app1.pdf]

References

1. Sabaté E. World Health Organization. 2003. Adherence to long-term therapies: evidence for action URL: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1 [accessed 2019-06-03] [WebCite Cache ID 78qiPrZyF]
2. Kini V, Ho PM. Interventions to Improve Medication Adherence: A Review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
3. Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. Diabet Med 2015 Jun;32(6):725-737. [doi: 10.1111/dme.12651] [Medline: 25440507]
4. Cramer JA. A systematic review of adherence with medications for diabetes. Diabetes Care 2004 May;27(5):1218-1224. [doi: 10.2337/diacare.27.5.1218] [Medline: 15111553]
5. Bailey CJ, Kodack M. Patient adherence to medication requirements for therapy of type 2 diabetes. International Journal of Clinical Practice 2011 Mar;65(3):314-322. [doi: 10.1111/j.1742-1241.2010.02544.x] [Medline: 21314869]
6. Lin L, Sun Y, Heng BH, Chew DEK, Chong P. Medication adherence and glycemic control among newly diagnosed diabetes patients. BMJ Open Diabetes Research & Care 2017;5(1):e000429 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000429] [Medline: 28878942]
7. Mayberry LS, Osborn CY. Family support, medication adherence, and glycemic control among adults with type 2 diabetes. Diabetes Care 2012 Jun;35(6):1239-1245 [FREE Full text] [doi: 10.2337/dc11-2103] [Medline: 22538012]
8. Stolar M. Glycemic control and complications in type 2 diabetes mellitus. Am J Med 2010 Mar;123(3 Suppl):S3-11. [doi: 10.1016/j.amjmed.2009.12.004] [Medline: 20206730]
9. American Diabetes Association. Abridged for Primary Care Providers. Clin Diabetes 2017 Jan;35(1):5-26 [FREE Full text] [doi: 10.2337/cd16-0067] [Medline: 28144042]

10. Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost. Medical Care 2005 Jun;43(6):521-530. [doi: 10.1097/01.mlr.0000163641.86870.af] [Medline: 15908846]
11. Rubens M, Saxena A, Ramamoorthy V, Khera R, Hong J, Veledar E, et al. Trends in Diabetes-Related Preventable Hospitalizations in the U.S., 2005-2014. Diabetes Care 2018 May;41(5):e72-e73. [doi: 10.2337/dc17-1942] [Medline: 29378777]
12. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
13. IQVIA. 2017. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-06-03] [WebCite Cache ID 78qkViDOa]
14. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
15. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision Support and Alerts of Apps for Self-management of Blood Glucose for Type 2 Diabetes. JAMA 2019 Dec 16;321(15):1530-1532. [doi: 10.1001/jama.2019.1644] [Medline: 30990543]
16. 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]
17. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
18. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a Smartphone Application With Medication Adherence and Blood Pressure Control: The MedISAFE-BP Randomized Clinical Trial. JAMA Intern Med 2018 Jun 01;178(6):802-809. [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
19. Monroe VD. Repository@TWU. Denton: Texas Women's University; 2018 May. Testing a smartphone application intervention to improve medication adherence in African American female clinic patients with unstable high blood pressure: A two-group randomized control trial URL: https://twu-ir.tdl.org/handle/11274/10148 [accessed 2019-08-22]
20. International Diabetes Federation. 2017. IDF Diabetes Atlas Eighth Edition URL: https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/134-idf-diabetes-atlas-8th-edition.html [accessed 2019-06-03] [WebCite Cache ID 78qo0zYWm]
21. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011 Jul;93(1):1-9. [doi: 10.1016/j.diabres.2011.02.002] [Medline: 21382643]
22. Bahia K. GSM Association. 2018 Aug. Connected Society: State of Mobile Internet Connectivity 2018 URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/09/State-of-Mobile-Internet-Connectivity-2018.pdf [accessed 2019-06-03] [WebCite Cache ID 78qsEvKVF]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
24. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
25. 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]
26. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
27. Medisafe App. URL: https://www.medisafeapp.com/ [accessed 2019-09-04]
28. Singapore Business Review. 2018 Jan 30. 4.83 million Singaporeans are now online URL: https://sbr.com.sg/information-technology/news/483-million-singaporeans-are-now-online [accessed 2019-06-03] [WebCite Cache ID 78qpG903R]
29. Aloysius L. CNET. 2014 Nov 03. Singapore has the highest smartphone adoption rate in the world URL: https://www.cnet.com/news/singapore-has-the-highest-smartphone-adoption-in-the-world/ [accessed 2019-06-03] [WebCite Cache ID 78qpXM9y3]
30. Matza LS, Park J, Coyne KS, Skinner EP, Malley KG, Wolever RQ. Derivation and validation of the ASK-12 adherence barrier survey. Ann Pharmacother 2009 Oct;43(10):1621-1630. [doi: 10.1345/aph.1M174] [Medline: 19776298]
31. 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]

32. 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]
33. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile Phone Apps to Improve Medication Adherence: A Systematic Stepwise Process to Identify High-Quality Apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
34. Carey MP, Jorgensen RS, Weinstock RS, Sprafkin RP, Lantinga LJ, Carnrike CL, et al. Reliability and validity of the appraisal of diabetes scale. J Behav Med 1991 Feb;14(1):43-51. [Medline: 2038044]
35. Schmitt A, Gahr A, Hermanns N, Kulzer B, Huber J, Haak T. The Diabetes Self-Management Questionnaire (DSMQ): development and evaluation of an instrument to assess diabetes self-care activities associated with glycaemic control. Health Qual Life Outcomes 2013;11:138 [FREE Full text] [doi: 10.1186/1477-7525-11-138] [Medline: 23937988]
36. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
37. Fincham JE. Response rates and responsiveness for surveys, standards, and the Journal. Am J Pharm Educ 2008 Apr 15;72(2):43-626 [FREE Full text] [doi: 10.5688/aj720243] [Medline: 18483608]
38. Lee CS, Tan JHM, Sankari U, Koh YLE, Tan NC. Assessing oral medication adherence among patients with type 2 diabetes mellitus treated with polytherapy in a developed Asian community: a cross-sectional study. BMJ Open 2017 Sep 14;7(9):e016317 [FREE Full text] [doi: 10.1136/bmjopen-2017-016317] [Medline: 28912194]
39. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A. Interventions for enhancing medication adherence. Cochrane Database of Systematic Reviews 2014 Nov 20:1465-1858 [FREE Full text] [doi: 10.1002/14651858.cd000011.pub4]
40. Hawkins RC. Circannual variation in glycohemoglobin in Singapore. Clin Chim Acta 2010 Jan;411(1-2):18-21. [doi: 10.1016/j.cca.2009.09.031] [Medline: 19800877]
41. Khattab M, Khader YS, Al-Khawaldeh A, Ajlouni K. Factors associated with poor glycemic control among patients with type 2 diabetes. J Diabetes Complications 2010 Mar;24(2):84-89. [doi: 10.1016/j.jdiacomp.2008.12.008] [Medline: 19282203]
42. ClinicalTrials.gov. 2017 Feb 10. A Smartphone Application to Improve Medication Adherence Among People With Type 2 Diabetes Mellitus in Singapore (iADHERE) URL: https://ClinicalTrials.gov/show/NCT02105597 [accessed 2019-05-13]

Abbreviations

ASK-12: Adherence Starts with Knowledge-12
ADS: Appraisal of Diabetes Scale
CONSORT: Consolidated Standards of Reporting Trials
DSMQ: Diabetes Self-Management Questionnaire
HbA_{1c}: hemoglobin A_{1c}
mERA: mobile health evidence reporting and assessment
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 03.06.19; peer-reviewed by L Ming, R Robbins, L Steinman; comments to author 22.07.19; revised version received 30.07.19; accepted 17.08.19; published 12.09.19.

Please cite as:
Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J
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
URL: http://mhealth.jmir.org/2019/9/e14914/
doi:10.2196/14914
PMID:31516127

Original Paper

Mobile Health Divide Between Clinicians and Patients in Cancer Care: Results From a Cross-Sectional International Survey

Rosanna Tarricone[1,2], MSc, PhD; Maria Cucciniello[1,2], MSc, PhD; Patrizio Armeni[2], MSc, PhD; Francesco Petracca[2], MSc; Kevin C Desouza[3], PhD; Leslie Kelly Hall[4,5], BA; Dorothy Keefe[6], MD, PhD

[1]Department of Social and Political Science, Bocconi University, Milan, Italy

[2]Centre for Research on Health and Social Care Management, SDA Bocconi School of Management, Bocconi University, Milan, Italy

[3]School of Management, QUT Business School, Queensland University of Technology, Brisbane, Australia

[4]Healthwise, Boise, ID, United States

[5]Engaging Patient Strategy, Boise, ID, United States

[6]School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, Australia

Corresponding Author:
Rosanna Tarricone, MSc, PhD
Department of Social and Political Science
Bocconi University
Via Roentgen, 1
Milan, 20136
Italy
Phone: 39 3351250616
Email: rosanna.tarricone@unibocconi.it

Abstract

Background: Mobile technologies are increasingly being used to manage chronic diseases, including cancer, with the promise of improving the efficiency and effectiveness of care. Among the myriad of mobile technologies in health care, we have seen an explosion of mobile apps. The rapid increase in digital health apps is not paralleled by a similar trend in usage statistics by clinicians and patients. Little is known about how much and in what ways mobile health (mHealth) apps are used by clinicians and patients for cancer care, what variables affect their use of mHealth, and what patients' and clinicians' expectations of mHealth apps are.

Objective: This study aimed to describe the patient and clinician population that uses mHealth in cancer care and to provide recommendations to app developers and regulators to generally increase the use and efficacy of mHealth apps.

Methods: Through a cross-sectional Web-based survey, we explored the current utilization rates of mHealth in cancer care and factors that explain the differences in utilization by patients and clinicians across the United States and 5 different countries in Europe. In addition, we conducted an international workshop with more than 100 stakeholders and a roundtable with key representatives of international organizations of clinicians and patients to solicit feedback on the survey results and develop insights into mHealth app development practices.

Results: A total of 1033 patients and 1116 clinicians participated in the survey. The proportion of cancer patients using mHealth (294/1033, 28.46%) was far lower than that of clinicians (859/1116, 76.97%). Accounting for age and salary level, the marginal probabilities of use at means are still significantly different between the 2 groups and were 69.8% for clinicians and 38.7% for patients using the propensity score–based regression adjustment with weighting technique. Moreover, our analysis identified a gap between basic and advanced users, with a prevalent use for activities related to the automation of processes and the interaction with other individuals and a limited adoption for side-effect management and compliance monitoring in both groups.

Conclusions: mHealth apps can provide access to clinical and economic data that are low cost, easy to access, and personalized. The benefits can go as far as increasing patients' chances of overall survival. However, despite its potential, evidence on the actual use of mobile technologies in cancer care is not promising. If the promise of mHealth is to be fulfilled, clinician and patient usage rates will need to converge. Ideally, cancer apps should be designed in ways that strengthen the patient-physician relationship, ease physicians' workload, be tested for validity and effectiveness, and fit the criteria for reimbursement.

(JMIR Mhealth Uhealth 2019;7(9):e13584) doi:10.2196/13584

KEYWORDS

mHealth; cancer; mobile phone; survey; mobile app; digital health; mhealth

Introduction

Background

Many would agree that mobile health (mHealth; the use of portable devices for medical purposes) holds considerable promise for improving health care and the quality of life for people with cancer [1]. With internet access growing worldwide and well over 70% of people in Europe and the United States owning smartphones, the potential market for mHealth is very large and is projected to continue to grow [2].

In 2017, there were 325,000 mHealth (health and fitness and medical) apps with over 3.7 billion downloads [3], reflecting over 30% growth compared with 2016. Although the number of apps for wellness management decreased by 18% from 2015 to 2017, the number for managing health conditions increased by 48% in the same period [3].

Although the potential benefits of mHealth seem particularly compelling for managing chronic conditions, where the overall efficacy largely depends on patient compliance that frequently occurs outside of the formal health care system, prolonged, regular, and intensive use still represents a major challenge [4]. Apps specifically developed for chronic disease management have had some initial success but have so far failed to live up to their expectations [5-7]. Among specific diseases, the greatest proportion of apps on the market is for mental health and behavioral disorders (28%), followed by diabetes (16%) and cardiovascular disorders (11%) [3]. Although nearly 1 in 6 deaths are due to cancer, which is indeed among the leading causes of mortality (with an estimated 9.6 million deaths in 2018 and approximately 14 million new cases worldwide every year, projected to increase up to 22 million within the next two decades) [8], few of the mHealth apps focus on cancer care (5%). Not only are mHealth cancer apps relatively few, but the action put forward thus far has not been steered in the right direction; the available cancer apps mostly focus on awareness raising and information provision [9] and appear to be used for limited purposes in the actual health care process, with a prevailing focus on self-management activities and the automation of structured and unstructured processes [10]. Although the improved cancer survival rates and outcomes have led to considering most cancers as chronic, their treatment is still accompanied by distressing symptoms and serious toxicities that affect functioning and quality of life [11]. To address these issues, mHealth has the potential to track the patient experience and collect patient-reported outcomes to personalize care, draw insights, and shorten the cycle from research to clinical implementation [12].

When patients are able to record their experiences in real time and combine them with passive data collection from sensors and mobile devices, this information can inform better care for each patient and contribute to the growing body of health data that can be used to draw insights for all patients.

Preliminary research has addressed the interest of cancer patients in the use of mobile technologies to manage their disease [13,14], whereas the influence of demographic factors on predicting the use of Web-based health information resources and its patterns has been mostly assessed with respect to electronic health technologies [15-17]. Furthermore, unlike other medical devices, to which mHealth technologies broadly belongs, mHealth performance mainly depends on whether both patients and clinicians are actively involved in its use [18-21]. However, current evidence has not addressed oncologists, and little is known about what incentivizes their use of mobile technologies; although oncologists have previously been shown to be open, in principle, to considering mHealth technology as part of patient care [22].

Objective

In summary, the interest in the use of mobile apps in cancer care is increasing, but there is little empirical insight into stakeholders' perceptions. Therefore, to gain insight into key stakeholders' perceptions of the value that mHealth app use creates, we distributed 2 surveys targeting 2 populations of mHealth app stakeholders—randomly selected cancer clinicians and patients who use internet-enabled mobile devices, such as smartphones. Through these surveys, we gathered data on the use of mHealth apps by patients and on how clinicians and cancer patients perceive the value of mHealth app use. In this study, we therefore aimed to describe the physician and patient population that utilizes mHealth in cancer care, the activities they perform, as well as the reasons for not using it.

Methods

Survey Design and Settings

To investigate the use of mHealth in cancer care by patients and clinicians and the reasons for its use, we conducted a cross-sectional, international survey from July 2015 to February 2016. The survey included the European Big 5, that is, France, Germany, Italy, Spain, and the United Kingdom, as well as the United States. These countries exhibit some of the highest smartphone ownership rates and have mobile broadband penetration rates above 75% [23]. Concurrently, although cancer is now on the rise in developing countries, the overall age-standardized cancer rate is still approximately 1.8 times higher in more developed countries [24]. Thus, the diffusion and the current performance of a health care innovation device such as mHealth can suitably be investigated in these countries, ensuring the ecological validity of the study. The first draft of the survey was based on existing literature and previous surveys and experiments on mHealth [25-27]. The survey questions were finalized by the authors and translated into Spanish, French, German, and Italian by professional medical editors in the different languages. To guarantee the accuracy of the translations, we pilot tested them with a group of clinicians and patient representatives. A final completeness check was implemented, and all essential items were made mandatory; when possible, a nonresponse option such as *not applicable* was

provided. In the final version of the survey, we asked clinicians and patients up to 37 and 32 questions, respectively (some on the questions were dependent on previous answers).

Study Population

Eligible clinicians included oncologists who used smartphones and other devices for internet access, although not necessarily for mHealth. Similarly, eligible patients were those diagnosed with any kind of cancer who owned a smartphone that could access the internet. The survey was administered through qualtrics^{XM}, a US-based company established in 2002 that allows researchers to conduct surveys in communities that are traditionally hard to reach. To construct the panel, potential respondents were recruited through the Web (using specific keywords) and outlined based on their characteristics, and a stratified sample was later invited to join the research panel.

In particular, patients were selected based on a population panel that provides recruitment via Web (Web banners, pay per click, natural optimization of research, affiliate marketing, email, and online public relations activities). Oncologists were reached out drawing on panels that are constructed by telephone recruitment or via recruitment portals starting from specialized databases—such as those of scientific communities.

The survey was sent to 1800 oncologists and 1800 cancer patients consisting of a random sample of panelists stratified by country and age group. Both clinicians and patients were invited to participate in the study via email through Qualtrics and were provided with a link leading to the survey. The main screen of the online questionnaire provided all respondents with the aim of the study, the investigator information, and the expected time length of the survey (approximately 10 min). The respondents' right to confidentiality was respected, and consent to participate in the survey was obtained.

One concern when using online recruitment panels is that subjects rush through the online questionnaire without properly reading the provided instructions and questions. To increase the statistical power and reliability of our dataset, we screened respondents based on several criteria. First, we included control questions to detect spammers. Those study participants who failed to answer the control questions, answered all the questions in the same way, or filled in boxes with no-sense comments were excluded from our sample (28 clinicians and 68 patients). Second, we examined the time subjects took to fill out the questionnaire (for clinicians, a mean of 6.31 min and for patients, 6.40 min). Extreme deviations from the average time to complete the questionnaire were treated as outliers and excluded from further analysis. Thus, respondents within the lowest 1% percentile (less than 2.5 min) in terms of total time till survey completion were excluded. Furthermore, we checked whether the subjects' internet protocol (IP) addresses overlapped. In such cases, duplicate entries from the same IP address were excluded from our analysis (12 respondents—patients—in total).

Variables

The survey instrument included 4 different domains in both the clinician and patient versions: (1) sociodemographic variables (age, sex, education, and salary level), (2) mHealth utilization, (3) mHealth activities performed, and (4) reasons for not using mHealth. Both clinicians and patients were asked about their use of mHealth technologies for the management of cancer. Users, namely, individuals who owned a smartphone or any other mobile device and who used it for cancer-related purposes, were then asked to report for what purposes they used mHealth by choosing from a list of activities. These activities related to different degrees of pervasiveness of the technology aimed at highlighting different user expertise levels, based on a previously designed framework by Nasi et al [10]. As a result, respondents were further classified as either basic (ie, those who used mHealth to schedule appointments, access personal health care information, or read test results only) or advanced users (ie, people who used mHealth to monitor treatment side effects and prevent further events). In contrast, respondents not using mHealth were asked to identify the reasons that had so far hindered them from adopting the technology using 5-point Likert scale items in the following format: 1=I completely disagree, 2=I disagree, 3=I neither disagree nor agree, 4=I agree, and 5=I completely agree.

Statistical Analysis

Descriptive statistics were used to report respondents' sociodemographic information and the degree of utilization of mHealth in managing cancer care. To measure the relation between specific sociodemographic information and the likelihood of being mHealth users, 2 possible sources of sample selection bias need to be addressed. First, as the survey was administered online, the results are influenced by the general digital divide in the population. Second, the survey was completed only by patients and clinicians using mobile technologies for any purpose, ie, if the respondent could access the internet but was not a user of mobile technologies (eg, smartphones or tablets), the survey was concluded, and no further questions were asked. In the first case, the sample selection bias is relevant but does not influence our results as the target population of mHealth technologies does not include people without basic technological endowments (eg, a computer with internet access). The second source of bias, instead, is more relevant because it refers to the population having access to the Web but whose mobile endowment is low. Ideally, we should not exclude these respondents as they are a part of the potential target of mHealth. In our sample, only 21 out of 2170 respondents reported not using mobile technologies for any given purpose. To account for this potential bias that could still have a potential effect on our results, we used 2 different statistical approaches, namely, a propensity score–based regression adjustment (PSBRA) with weighting and a Heckman probit selection model (HPSM). For both the propensity score equation in the PSBRA and the selection equation in the HPSM, the independent variables were the age group, nationality, and salary level of the respondent. The choice of using 2 different procedures was motivated by the necessity of testing the robustness of the obtained estimations because of the disproportion between censored and uncensored observations. Analyses were conducted with STATA software, version 14 (Stata Corp).

Workshop and Stakeholder Engagement

The main results from the survey were shared with several stakeholders to solicit input and feedback as well as develop policy recommendations for an appropriate spread of mobile technologies. An international workshop was organized in Milan, Italy, to facilitate interaction with over 100 stakeholders, including patients, clinicians, app developers, the pharmaceutical and medical technology industry, telecom industries, experts in medical communications and health education, payers, and policymakers.

We announced the international workshop through different channels: (1) the general way, that is, by using the website and social networks normally used by our university (Bocconi University) to promote events and (2) a more specific way, that is, by compiling a mailing list of all potential stakeholders at the international and national levels. Participation was free of charge and travel expenses were covered by participants.

The session was intended to focus on the discussion of the results arising from the survey. Specifically, 3 main questions were aimed toward participants: (1) why patients and clinicians do not use mHealth evenly, (2) what are the main barriers that have slowed the adoption of mHealth in cancer care, and finally (3) what is the likely effect of mHealth on clinicians' activity and on patients' quality of life. A member of the research team facilitated the workshop, ensuring the surfacing of diverse perspectives and a rich discussion of issues. The feedback from the workshop was used to develop a set of questions that we posed to an expert roundtable.

The roundtable consisted of 4 participants who represented 2 leading patient and clinician associations: European Cancer Patient Coalition and the European School of Oncology in Europe and Healthwise Organization and the Multinational Association of Supportive Care in Cancer in the United States. The discussion was moderated by a member of the research team. Both the workshop and the roundtable were recorded and professionally transcribed.

Results

Study Population

Valid responses were obtained from 1116 of the clinicians surveyed (62.00% response rate) and 1033 of the cancer patients interviewed (57.39%). The respondents' characteristics are summarized in Tables 1 and 2. The patients were mostly female (637/1033, 61.66%) and aged over 45 years (798/1033, 77.25%), whereas the clinicians were mostly male (795/1116, 71.24%) and evenly apportioned between the 2 age groups. With respect to education, 28.46% of patients (294/1033) had received no education or had only attended primary school, 37.37% (386/1033) had either completed secondary school or achieved an undergraduate degree, and the remaining 34.17% (353/1033) had completed graduate (18.0%) or postgraduate (16.2%) education. Approximately one-third (335/1033, 32.43%) of the patients were employed full time, 12.88% (133/1033) were employed part time, and about one-third (366/1033, 35.43%) were retired. Employed patients prevalently earned less than US $30,000 per year (178/1033, 17.23%) or between US $30,001 and US $50,000 per year (129/1033, 12.49%). In contrast, more than half of the clinicians (721/1116, 64.61%) earned over US $75,000 per year, with relevant observed differences between Germany, the United Kingdom, and the United States (62.7%, 70.4%, and 91.1%, respectively) and Mediterranean countries (30.5% in France, 26.0% in Italy, and 7.7% in Spain).

Table 1. Patient sample characteristics by country, 2016.

Patient characteristics	France (n=103)	Germany (n=101)	Italy (n=105)	Spain (n=102)	United Kingdom (n=111)	United States (n=511)	Total (N=1033)
Sex, n (%)							
Male	35 (34.0)	45 (44.6)	39 (37.1)	37 (36.3)	53 (47.7)	187 (36.6)	396 (38.33)
Female	68 (66.0)	56 (55.4)	66 (62.9)	65 (63.7)	58 (52.3)	324 (63.4)	637 (61.67)
Age group (years), n (%)							
Under 45	20 (19.4)	24 (23.8)	31 (29.5)	41 (40.2)	13 (11.7)	106 (20.7)	235 (22.75)
Over 45	83 (80.6)	77 (76.2)	74 (70.5)	61 (59.8)	98 (88.3)	405 (79.3)	798 (77.25)
Education level, n (%)							
No or primary education	29 (28.2)	57 (56.4)	56 (53.3)	8 (7.8)	28 (25.2)	110 (21.5)	288 (27.88)
Secondary or undergraduate education	26 (25.2)	24 (23.8)	8 (7.6)	34 (33.3)	32 (28.8)	262 (51.3)	392 (37.95)
Graduate	30 (29.1)	7 (6.9)	29 (27.6)	48 (47.1)	30 (27.0)	42 (8.2)	186 (18.01)
Postgraduate	18 (17.5)	13 (12.9)	12 (11.4)	12 (11.8)	15 (13.5)	97 (19.0)	167 (16.17)
Employment status, n (%)							
Full-time employed	32 (31.1)	29 (28.7)	43 (41.0)	57 (55.9)	28 (25.2)	146 (28.6)	335 (32.43)
Part-time employed	13 (12.6)	19 (18.8)	16 (15.2)	5 (4.9)	15 (13.5)	65 (12.7)	133 (12.88)
Unemployed	6 (5.8)	5 (5.0)	6 (5.7)	15 (14.7)	1 (0.9)	21 (4.1)	54 (5.23)
Not employed and not looking for work	4 (3.9)	5 (5.0)	1 (1.0)	7 (6.9)	9 (8.1)	26 (5.1)	52 (5.03)
Unable to work	6 (5.8)	9 (8.9)	7 (6.7)	4 (3.9)	7 (6.3)	50 (9.8)	83 (8.03)
Student	0	0	3 (2.9)	1 (1.0)	0	6 (1.2)	10 (0.97)
Retired	42 (40.8)	34 (33.7)	29 (27.6)	13 (12.7)	51 (45.9)	197 (38.6)	366 (35.43)
Salary level, n (%)							
≤US $30,000	23 (22.3)	20 (19.8)	41 (39.0)	42 (41.2)	19 (17.1)	33 (6.5)	178 (17.23)
US $30,001-US $50,000	16 (15.5)	14 (13.9)	13 (12.4)	18 (17.6)	14 (12.6)	54 (10.6)	129 (12.49)
US $50,001-US $75,000	6 (5.8)	10 (9.9)	4 (3.8)	0	6 (5.4)	52 (10.2)	78 (7.55)
>US $75,001	0	4 (4.0)	1 (1.0)	2 (2.0)	4 (3.6)	72 (14.1)	83 (8.03)
Missing or not applicable	58 (56.3)	53 (52.5)	46 (43.8)	40 (39.2)	68 (61.3)	300 (58.7)	565 (54.70)

Table 2. Clinician sample characteristics by country, 2016.

Sample characteristics	France (n=105)	Germany (n=150)	Italy (n=123)	Spain (n=104)	United Kingdom (n=108)	United States (n=526)	Total (N=1116)
Sex, n (%)							
Male	75 (71.4)	116 (77.3)	81 (65.9)	55 (52.9)	74 (68.5)	394 (74.9)	795 (71.24)
Female	30 (28.6)	34 (22.7)	42 (34.1)	49 (47.1)	34 (31.5)	132 (25.1)	321 (28.76)
Age group (years), n (%)							
Under 45	62 (59.0)	49 (32.7)	44 (35.8)	66 (63.5)	70 (64.8)	286 (54.4)	577 (51.70)
Over 45	43 (41.0)	101 (67.3)	79 (64.2)	38 (36.5)	38 (35.2)	240 (45.6)	539 (48.30)
Education level, n (%)							
No or primary education	0	0	0	0	0	0	0
Secondary or undergraduate education	0	0	0	0	0	0	0
Graduate	0	0	0	0	0	0	0
Postgraduate	105 (100.0)	150 (100.0)	123 (100.0)	104 (100.0)	108 (100.0)	526 (100.0)	1116 (100.00)
Employment status, n (%)							
Full-time employed	105 (100.0)	150 (100.0)	123 (100.0)	104 (100.0)	108 (100.0)	526 (100.0)	1116 (100.00)
Part-time employed	0	0	0	0	0	0	0
Unemployed	0	0	0	0	0	0	0
Not employed and not looking for work	0	0	0	0	0	0	0
Unable to work	0	0	0	0	0	0	0
Student	0	0	0	0	0	0	0
Retired	0	0	0	0	0	0	0
Salary level, n (%)							
≤US $30,000	3 (2.9)	4 (2.7)	13 (10.6)	6 (5.8)	1 (0.9)	5 (1.0)	32 (2.87)
US $30,001-US $50,000	28 (26.7)	14 (9.3)	40 (32.5)	48 (46.2)	10 (9.3)	4 (0.8)	144 (12.90)
US $50,001-US $75,000	41 (39.0)	34 (22.7)	37 (30.1)	42 (40.4)	20 (18.5)	31 (5.9)	205 (18.37)
>US $75,001	32 (30.5)	94 (62.7)	32 (26.0)	8 (7.7)	76 (70.4)	479 (91.1)	721 (64.61)
Missing	1 (1.0)	4 (2.7)	1 (0.8)	0	1 (0.9)	7 (1.3)	14 (1.25)

Patient and Clinician Usage of Mobile Health

Of the 2149 participants surveyed, 1153 (53.65%) had previously accessed some sort of mobile technology for cancer-related purposes. Different mHealth access rates were observed in the 2 end-user groups. Among patients, 28.46% (294/1033) were mHealth users: nonusers were the majority in all countries assessed, although there were between-country differences. Clinicians, in contrast, were most often mHealth users: 76.97% of the respondents (859/1116) utilized mobile technology in their daily activity or in the management of cancer patients. The highest percentage was observed in the United States (459/526, 87.26%). Regarding the intensity of use, we observed that among clinician respondents, 32.26% (360/1116) were advanced users and 44.71% (499/1116) were basic users, whereas 18.39% (190/1033) of patients reported being advanced users versus 10.16% (105/1033) of basic users (Table 3).

Table 3. Distribution of users and nonusers of mobile health in the analyzed countries, 2016 (N=2149).

Users	France	Germany	Italy	Spain	United Kingdom	United States	Total
Patients, total	**n=103**	**n=101**	**n=105**	**n=102**	**n=111**	**n=511**	**N=1033**
Users, n (%)	16 (15.5)	34 (33.7)	46 (43.8)	25 (24.5)	18 (16.2)	155 (30.3)	294 (28.46)
Basic users	8 (7.8)	10 (9.9)	15 (14.3)	9 (8.8)	6 (5.4)	56 (11.0)	104 (10.07)
Advanced users	8 (7.8)	24 (23.8)	31 (29.5)	16 (15.7)	12 (10.8)	99 (19.4)	190 (18.39)
Nonusers, n (%)	87 (84.5)	67 (66.3)	59 (56.2)	77 (75.5)	93 (83.8)	356 (69.7)	739 (71.54)
Clinicians, total	**n=105**	**n=150**	**n=123**	**n=104**	**n=108**	**n=526**	**N=1116**
Users, n (%)	72 (68.6)	104 (69.3)	72 (58.5)	60 (57.7)	92 (85.2)	459 (87.3)	859 (76.97)
Basic users	33 (31.4)	38 (25.3)	35 (28.5)	36 (34.6)	56 (51.9)	301 (57.2)	499 (44.71)
Advanced users	39 (37.1)	66 (44.0)	37 (30.1)	24 (23.3)	36 (33.3)	158 (30.0)	360 (32.26)
Nonusers, n (%)	33 (31.4)	46 (30.7)	51 (41.5)	44 (42.3)	16 (14.8)	67 (12.7)	257 (23.03)

Clinicians' and Patients' Mobile Health Activities

Among the patients classified as mHealth user, approximately half of the respondents used mobile technologies for automation and decision-making support into activities such as scheduling an appointment (157/294, 53.4%), accessing personal information (147/294, 50.0%), and reading test results (135/294, 45.9%). Only approximately one-third of users and, therefore, about one-tenth of total patient respondents, supported treatment and follow-up phases through mHealth by either monitoring side effects (108/294, 36.7% of users), helping prevent further events (85/294, 28.9%), or taking medications as prescribed (97/294, 33.0%; Table 4). With regard to clinicians, the majority accessed mHealth to carry out activities that pertain to the automation and interaction domains: 88.6% (761/859) used mobile apps to perform literature research, 66.9% (575/859) to interact with their colleagues, and 44.6% (383/859) to communicate directly with patients. Fewer clinician users utilized mHealth for decision-making purposes: 46.1% (396/859) used mobile apps to access patients' electronic health records, 44.0% (378/859) to collect test results, and a smaller number (324/859, 37.7%) of users used mHealth to support decision making for ordering further tests. A minority of users performed activities that support treatment and follow-up in the care process, such as side-effect management (318/859, 37.0%) and compliance monitoring (116/859, 13.5%; Table 5).

Table 4. Activities performed by patient users, by degree of pervasiveness of the technology.

Activities performed by patient users	Frequency of the activity among users (n=294), n (%)	Frequency of the activity among total respondents (N=1033), n (%)
Activities supporting automation and interaction		
Schedule an appointment with a physician	157 (53.4)	157 (15.20)
Activities supporting decision making processes		
Access personal health care information	147 (50.0)	147 (14.23)
Get test results	135 (45.9)	135 (13.07)
Activities supporting treatment and follow-up		
Monitor side effects (nausea, vomiting, and diarrhea)	108 (36.7)	108 (10.45)
Help prevent further events (cancer progression and recurrence)	85 (28.9)	85 (8.23)
Help in taking medications as prescribed	97 (33.0)	97 (9.39)

Table 5. Activities performed by clinician users, by degree of pervasiveness of technology.

Activities performed by clinician users	Frequency of the activity among users (n=859), n (%)	Frequency of the activity among total respondents (N=1116), n (%)
Activities supporting automation and interaction		
Literature research	761 (88.6)	761 (68.19)
Communicate directly with patients	383 (44.6)	383 (34.32)
Interact with colleagues for timely decision-making	575 (66.9)	575 (51.52)
Activities supporting decision-making processes		
Access patients' electronic health records	396 (46.1)	396 (35.48)
Get test results	378 (44.0)	378 (33.87)
Decision support for ordering further tests	324 (37.7)	324 (29.03)
Activities supporting treatment and follow-up		
To monitor compliance (principal treatment)	116 (13.5)	116 (10.39)
To manage side effects	318 (37.0)	318 (28.49)

Professional Mobile Health Divide Between Clinicians and Patients

Table 6 shows the marginal probabilities of use at means that were 69.8% for clinicians and 38.7% for patients using the PSBRA technique (69.5% and 38.7%, respectively, using HPSM). Other things being equal, clinicians use mHealth more than patients, thus, highlighting an inefficient activation of the complementarities between the two main actors involved in the process of care. Age and salary level influenced mHealth adoption in both end-user groups. Among clinicians, younger professionals exhibited an approximately 15 percentage point higher likelihood of being mHealth users (82.9% vs 64.6% using PSBRA), while this gap was even wider for patients, verging on 30 percentage points. Salary level had a similar impact, with more affluent respondents more likely to be mHealth users than the less well-off respondents. These variables also explain the width of the divide between clinicians and patients. With respect to age, the divide was significantly higher for old respondents (35.4% using PSBRA) than that for young respondents (24.5%), whereas in regard to salary level, the divide was lower for low-income (29.5% using PSBRA) and high-income categories (23.2%) and significantly higher for medium-income ones (between 36.4% and 43.7%). Further differences arose when country-level situations were addressed, with the divide being as high as nearly 50 percentage points in the United Kingdom (51.7% using PSBRA).

Table 6. Marginal probabilities of mobile health use (N=2149).

Statistical approaches[a,b]	PSBRA[c], %			HPSM[d], %		
	Clinicians	Patients	Divide	Clinicians	Patients	Divide
Main effect	71.8	40.2	31.6	71.2	39.6	31.6
Age (years)						
≤45	82.9	58.5	24.5	80	54.2	25.7
>45	64.6	29.2	35.4	66	30.6	35.4
Salary						
≤US $30,000	63.7	34.2	29.5	61.2	33.3	28
US $30,001-US $50,000	70.6	34.3	36.4	70.1	34.4	35.7
US $50,001-US $75,000	75.5	31.9	43.7	75.2	30.2	44.9
>US $75,000	74	50.8	23.2	73.7	50.5	23.1
Country						
France	62	27.5	34.5	60.6	25.5	35.1
Germany	60	43.4	16.6	64.4	47	17.4
Italy	50.6	51.4	−0.8[e]	55.2	55.5	−0.3[e]
Spain	46	26.3	19.7	49.5	29.3	20.2
United Kingdom	75.5	23.8	51.7	77	28	48.9
United States	84.7	46.2	38.5	82.1	41.6	40.5

[a]Marginal probabilities at both values of *clinician/patient* dummy are displayed.

[b]The regression used to estimate propensity scores had a pseudo-R-squared value of 0.15 and the goodness-of-fit test showed a Pearson chi-square value of 19.1. The logit model included the propensity score as covariate and as probability weight.

[c]PSBRA: propensity score–based regression adjustment with weighting.

[d]HPSM: Heckman probit selection model.

[e]Not significant.

Reasons That Hinder Greater Mobile Health Use

Participants who did not belong to the user category were asked about their concerns regarding mHealth use and answered 5-point Likert scale items (Table 7). On the patient side, the most diffused concerns pertained to the preference for traditional means of communication with their doctor (mean 4.26, SD 0.93), the lack of knowledge about the potentials of information technologies (mean 3.82, SD 1.17), and the doubts about the reliability and effectiveness of mHealth for medical purposes (mean 3.03, SD 1.07). For nonuser clinicians, the most substantial doubts were related to the preference for in-person visits (mean 4.13, SD 0.91) and the inability of patients to use smartphones (mean 3.44, SD 0.94).

Table 7. Barriers for mobile health (mHealth) use rated on a 5-point Likert-type scale by patient and clinician nonusers.

Participant, barrier	Mean (SD)
Patient nonusers	
I am worried about the protection of the confidentiality of my personal, medical and health information	2.65 (1.21)
I do not trust the technical reliability of the software	2.86 (1.06)
I think mobile technologies are not effective and reliable for medical purposes	3.03 (1.07)
I am not attracted by mHealth because I cannot use the devices properly	2.65 (1.08)
I prefer to communicate and meet my doctor in person	4.26 (0.93)
I was not aware of this possibility	3.82 (1.17)
I cannot afford the costs of mobile devices and connection	2.64 (1.23)
Clinician nonusers	
I am doubtful about providing mobile type of support because of data security concerns	2.89 (1.18)
I do not trust the technical reliability of the software	2.34 (0.95)
I am not interested in mHealth because I cannot use the devices properly	2.12 (0.97)
I was not aware of this potential use of mobile phones	3.02 (1.17)
I think mobile technologies are not effective and reliable for medical purposes	2.22 (0.96)
I realized patients are often not able to utilize mobile technologies	3.44 (0.94)
I still prefer to communicate and meet my patient in person	4.13 (0.91)
I think it would be uncomfortable mixing the face-to-face relationship with my patients with the virtual practice produced by mHealth	2.90 (1.13)

Qualitative Feedback

The large spectrum of stakeholders involved in the workshop helped identify further key themes. These were the generic nature of medical apps, the lack of user-friendliness because of integration into work and life contexts, the poor interaction interfaces, and the confusion about whether and when medical apps must be considered medical devices and whether they must meet evidential requirements or not. During the roundtable, the experts agreed that current apps are seldom developed with patients in mind and that, in many cases, the app functionalities do not meet patients' expectations and needs. Therefore, participants agreed that it would be extremely important to identify the target audience's wishes or expectations before designing and developing new apps. In particular, the participants emphasized that to define the content of apps, it would be fundamental to understand the characteristics of the main target population (ie, old/young user, type and stage of disease, and different familiarity levels with technology), the language (the simpler the better, avoiding scientific language, and making the app immediately easy to use), and the layout (ie, small fonts on a small screen are a barrier for old people).

Discussion

Principal Findings

With the aging of the population and the epidemics of chronic diseases, the financial sustainability of health care systems across the globe is at threat and calls for new paradigms where patients are empowered to stay healthy and/or to self-manage their conditions and hospitals only serve to treat the acute phases of diseases and to connect the community and patients' home to deliver long-term chronic care. In such a context, mHealth, leveraging on the increase in mobile smartphone subscribers (over 4.4 billion in 2017 [28], representing over 2 out of 3 adults on earth), is emerging as a viable solution to keep patients informed and empowered, to provide clinicians with timely data that can improve their capacity to assess patients' health status, and to help improve hospitals to reorganize their production function and management processes to better fit the evolving needs of the population [29].

Cancer survivors experience differing needs in terms of medical care, psychosocial support, and practical needs of daily living, and mHealth apps can provide access to information and health behavior interventions that are low cost, easy to access, and personalized to their specific needs [30]. Benefits can go as far as increasing a patient's chances of overall survival [31]. However, despite the largely acknowledged potential and the increase in artifact development, available evidence on the actual use of mobile technologies in cancer care and cancer supportive care is still scant. This study found a utilization rate of mHealth of less than 30% by cancer patients. These results are slightly higher, although in the same order of magnitude, compared with those of a cross-sectional survey administered at a University Hospital in Spain, according to which 20.3% of the surveyed hematology-oncology patients had a health app [13]. Clinicians, in contrast, exhibited a more widespread utilization of mHealth according to our survey results, with over three-quarters of users among those surveyed.

In any case, not all users accessed mHealth for the same purposes; our analysis identified a further gap, the one between basic and advanced users.

Clinicians tend to use mHealth mainly in isolation, without mHealth-based interaction with patients. Indeed, clinician users reported that they access mHealth extensively to perform activities that support automation and data collection (such as *Interaction with colleagues* and *Literature research*) and less often to support clinical decision making.

Our survey results are thus consistent with previously published studies that highlighted a limited focus of mHealth experimental studies and apps on treatment and follow-up activities in the oncology field [9,10].

Several barriers still halt wider adoption by both clinicians and patients, the main one being the preference for in-person communication and the related concern that mobile technologies might hinder the relationship of trust. Both nonuser groups support these apprehensions, and our results are akin to previous literature results that identified the wish for personal contact with the treating physician as the main reason for app refusal in a cancer patient survey [14] as well as the clinician fear that mHealth might jeopardize the patient-clinician relationship and increase their workload [29]. However, according to a broader systematic, narrative review, after adopting mHealth apps, patients felt empowered and perceived a positive impact on the relationship with their providers [32].

On a different note, our analysis confirmed that, as is seen for the use of internet and smartphones in general, age, education, and income play important roles in explaining the use of mHealth in cancer care by both clinicians and patients. However, other things being equal, we found that the use of mHealth technologies is significantly more common among clinicians than among patients and that factors such as age, income, and origin further contribute to modulating the extent of this divide. This divide might be present because mHealth, such as most types of health technologies (eg, medical devices), represents a work instrument for clinicians who, for the sake of improving their performance, normally are prone to and represent the natural target for technological innovation [33,34]. However, consumers do not normally encounter health technologies until they become patients and, in principle, would not care at all about them unless they happened to contract a disease or were prescribed the technology by their doctors. Much is known about the typical agency relationship that happens between patients and doctors together with the supplier-induced demand that makes patients' consumption of health care services highly dependent upon doctors' advice [13,35,36]. Moreover, there is evidence that the membership in interprofessional alliances and networks for change is instrumental to facilitate or hinder the diffusion process of new technologies [37]. Physicians participate in specific networks for change that place them in a privileged position in the diffusion of innovations. Until the sociodemographic evolution alleviates this trend, clinicians might play decisive roles in spreading mHealth utilization in cancer care and recruiting more and more patients to adopt it.

In fact, this professional divide represents a barrier to mHealth effectiveness in cancer treatment. If the promise of mHealth is to be fulfilled, clinician and patient usage rates will need to converge. There is merit in incentivizing oncologists to adopt cancer apps in routine practice to encourage patients to access mHealth at greater length.

Incentives for Greater Mobile Health Utilization

To enhance clinician use, several different layers can be approached. The first dimension pertains to artifact design; ideally, cancer apps would need to be designed in such a way that would strengthen the patient-clinician relationship, and they should be tested for validity, accuracy, and self-efficacy to help clinicians and patients orient themselves in what now seems to be an app overload [38].

App designers and developers must do more to bring their end users into the design process. Our findings point to the need for app developers to leverage toolkits to enable patients and physicians to more fully engage in the design and development process, each contributing with their own expertise [39]. This will enable the cocreation of solutions. In working toward the development of a sustainable Information Technology system, it is important to engage the final users, particularly the clinicians and patients, throughout the different phases from problem identification to the design and development phase, aligning the project trajectory to users' needs and expectations and providing clinicians with the opportunity for self-reflection and revisions. Unfortunately, many mHealth apps are designed without considering the needs of their users in terms of either patients or clinicians [4]. The literature lacks empirically validated guidelines or process models on how to design apps with stakeholders rather than for stakeholders. As a result, a recent overview of systematic studies by Byambasuren et al revealed that most mHealth apps are of low quality [40], which hinders their recommendations by clinicians and their use by patients. In cancer care only, Brouard et al evaluated 117 apps for oncological information and treatment monitoring [41] and found that the validation of those apps was generally poor (27.4%).

First of all, these results suggest that designers and developers need to recognize that a one-size-fits-all approach will not work when it comes to apps dealing with conditions such as cancer. Specific apps that account for differences in types of patients, variance in the stage of the disease, and the kind of care one is receiving, as well as the expertise one has with using mobile technologies, will have better chances for increasing adoption and regular usage.

Second, most health apps lack evidence of clinical effectiveness and do not undergo a formal validation and evaluation process [42]; the lack of evidence on whether and under what conditions mHealth delivers on its promise to improve patients' health outcomes and the efficiency of the health care process further contributes to restraining greater utilization [29,43-45]. More cancer apps need to be tested for their efficacy as, with few exceptions [31,46,47], the evidence base in support of mHealth technologies is still lacking [9]. However, given that the overall performance of mHealth apps is multidimensional, that is, it can be measured from different perspectives (eg, patients, caregivers, and clinicians), it is necessary to develop a methodological framework to include a wider array of benefits beside clinical outcomes. This is part of the objectives of Pushing the boundaries of Cost and Outcome Analysis of

Medical Technologies, a large, 3-year, European Union funded project whose recommendations on how to assess mHealth apps are expected in 2020 [48].

Also, the National Institute for Health and Care Excellence in the United Kingdom has recently started a project aimed at providing guidance on the assessment of mobile technologies (ie, *Behaviour change: digital and mobile health interventions*) expected to be delivered in 2020 [49].

Third, efficient regulation can help promote the adoption of mHealth apps. mHealth apps are classified as medical devices when they are used for diagnosis, prevention, treatment, monitoring, or alleviation of disease in human beings and for this reason must respond to high regulatory standards for demonstrating clinical benefit and safety [50]. Nevertheless, regulatory systems have rarely been able to catch up with the exponential launch of mHealth apps in the global market and have often been equivocal in establishing whether a software-based technology has to be treated as a medical device. This resulted in a very poor number of clinical trials that included digital health technologies, 860 worldwide in 2017 [51] compared with the number of mHealth apps for managing health conditions in the same period (126,000) [3], which means that the large majority of mHealth apps have entered the market without any clinical evidence in support [52]. This might have reduced clinicians' and ultimately patients' confidence in the reliability and efficacy of the apps. The US Food and Drug Administration (FDA) [53] and the Directorate for Health and Consumers of the European Commission [54] have long been trying to clarify the regulatory standards that digital health technologies need to meet and what evidential requirements need to be developed by app manufacturers. Furthermore, the FDA has led a working group within the International Medical Device Regulatory Forum aimed at harmonizing the regulatory framework for software-based technologies across different jurisdictions [55] and will include in the fiscal year 2019 budget a Center of Excellence for Digital Health that will aim at modernizing the regulatory approach to digital health [56]. These efforts are crucial as they would guarantee common rules to manufacturers that work in a global environment and would increase the level of trust in end users.

Fourth, data protection must be addressed to increase end users' confidence in using mHealth apps. Clinicians still do not feel the full reassurance about the reliability of the data collected and of the available apps [57]. Data should not only be reliable but their usability is also particularly critical: the vast amounts of data potentially available to patients and providers could easily overwhelm them if not put to best use. The FDA has published premarket and postmarket guidance that offers recommendations for the comprehensive management of medical device cybersecurity risks and continuous improvement throughout the total product life cycle as well as incentivized changing marketing and distributed medical devices to reduce risks [58]. More recently, the General Data Protection Regulation, a European Union law, aimed at regulating personal data in the digital world [59]. Although they are too new to be assessed, we think these efforts go in the right direction of increasing clinicians' and consequently patients' confidence in using mHealth apps.

Finally, innovative, multidisciplinary home-based models of care are now available for cancer patients who can be actively maintained with oral anticancer drugs and have shown preliminary success in optimally managing adherence during pilot testings [60-62]. Although the impact of personalized mHealth apps on adherence and other significant outcomes of patients on oral anticancer medications is yet to be assessed [63], appropriate economic incentives and related formulas are needed to spur their utilization of these devices [20].

In conclusion, like all other medical devices, mHealth uptake and diffusion largely depend on clinicians' conviction, but, differently from some other medical devices (eg, implantable devices), the effectiveness of mHealth heavily depends upon patients playing an active role and using it at the same pace as clinicians.

Study Strengths and Limitations

This is the first survey including a large, international sample size comparing 6 different countries in North America and Europe and, even more importantly, covering the two most important end users of mHealth: patients and clinicians. In fact, albeit scant, previous research has primarily addressed cancer patient needs and attitudes toward mobile technologies and not those of clinicians. Although past estimates exist for other specializations [64], to our knowledge, this represents the first evidence of mHealth utilization by clinicians in the cancer field. Second, this is the first study that combines the survey approach with a more qualitative method (workshop and roundtable with key stakeholders) to better interpret and complement the quantitative evidence emerging from the survey to ultimately provide concrete recommendations to decision makers.

However, this study suffers from some limitations that should be considered in subsequent studies. First, the study was based on a volunteer online access panel and, thus, is not entirely representative of the reference population as only individuals who possess some degree of digital competence could be reached and included. However, we believe that the online tool contributed to highlighting the smallest divide between clinicians and patients, which would likely be larger had we not used a digital tool. Moreover, the investigation in the user groups of the activities performed by patients and clinicians who use mHealth was self-reported and not based on actual records of their practice. Finally, this survey presents the limit of generalizability; thus, the divide and the models tested are valid in cancer and cancer supportive care only, and as much as the results are extremely significant, they might not hold true for other types of diseases.

Conclusions

The use of mobile apps in health and in cancer care is literally booming but poor knowledge exists on who is using mHealth, for what purposes, what kind of apps are used, and what is the likely future of mHealth in clinical practice. In this study, we contributed to filling these gaps: our findings highlight 2 types of digital divides in cancer care—one mediated by socioeconomic and educational inequalities among patients and the other by the rift between how doctors and patients are deploying these technologies. For mHealth to yield its full

benefits, it will have to integrate these two ends rather than foment the existing divide.

Acknowledgments

This study was resourced via an unrestricted grant by Helsinn Healthcare. The funders have had no influence on the content of the manuscript. The authors would like to acknowledge Kathi Apostolidis for providing great expertise during the survey draft and pilot and all anonymous respondents who completed the draft survey and helped us improve it with their valuable insights. Further appreciation goes to Francesco De Lorenzo and Alberto Costa for their comments and feedback. The authors also thank Pat Truman and Bets Davis for their valuable contributions to previous versions of the manuscript. The authors state that the manuscript is an honest and accurate account of the study reported here, and no aspects of the study have been omitted.

Authors' Contributions

RT and MC conceived the research; PA and FP analyzed the data; and RT, MC, KCD, and FP wrote the manuscript and contextualized the findings. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
2. Cleary M. How mHealth technology is revolutionizing clinical research. Value & Outcomes Spotlight 2018 Sep;4(5):20-23 [FREE Full text]
3. Research2guidance. 2017. mHealth App Developer Economics URL: https://research2guidance.com/mhealth-app-developer-economics/ [accessed 2019-01-17] [WebCite Cache ID 75UwsH4IK]
4. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med 2018 Jan 15;1(1). [doi: 10.1038/s41746-017-0002-4]
5. 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]
6. Kuckman R, Stern A. Harvard Business Review. 2018 Apr. Why apps for managing chronic disease haven't been widely used URL: https://hbr.org/2018/04/why-apps-for-managing-chronic-disease-havent-been-widely-used-and-how-to-fix-it [accessed 2019-01-19]
7. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of health information technology use between American adults with and without chronic health conditions: findings from the National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
8. World Health Organization. Cancer key facts URL: https://www.who.int/news-room/fact-sheets/detail/cancer [accessed 2019-01-07] [WebCite Cache ID 75FjOtXTw]
9. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
10. Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res 2015 Feb 12;17(2):e26 [FREE Full text] [doi: 10.2196/jmir.3757] [Medline: 25679446]
11. Esther Kim JE, Dodd MJ, Aouizerat BE, Jahan T, Miaskowski C. A review of the prevalence and impact of multiple symptoms in oncology patients. J Pain Symptom Manage 2009 Apr;37(4):715-736 [FREE Full text] [doi: 10.1016/j.jpainsymman.2008.04.018] [Medline: 19019626]
12. Mooney K, Berry DL, Whisenant M, Sjoberg D. Improving cancer care through the patient experience: how to use patient-reported outcomes in clinical practice. Am Soc Clin Oncol Educ Book 2017;37:695-704 [FREE Full text] [doi: 10.14694/EDBK_175418] [Medline: 28561689]
13. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
14. 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]

15. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
16. Im EO, Chee W, Liu Y, Lim HJ, Guevara E, Tsai H, et al. Characteristics of cancer patients in internet cancer support groups. Comput Inform Nurs 2007;25(6):334-343 [FREE Full text] [doi: 10.1097/01.NCN.0000299655.21401.9d] [Medline: 18000430]
17. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
18. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
19. Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs--same or different? Value Health 2009 Jun;12(4):402-404 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00476_1.x] [Medline: 19138306]
20. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 5;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
21. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
22. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an app? Oncology providers' views on mobile health apps for cancer care. JCO Clin Cancer Inform 2017 Dec;1:1-7 [FREE Full text] [doi: 10.1200/CCI.17.00107] [Medline: 30657404]
23. Organisation for Economic Co-operation and Development. OECD Broadband Statistics Portal URL: http://www.oecd.org/internet/oecdbroadbandportal.htm#map [accessed 2019-01-17]
24. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 1;136(5):E359-E386. [doi: 10.1002/ijc.29210] [Medline: 25220842]
25. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin L, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788 [FREE Full text] [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
26. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht UV. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
27. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
28. Ericcson. Mobility report: key figures URL: https://www.ericsson.com/en/mobility-report/reports/november-2017/key-figures [accessed 2019-01-17] [WebCite Cache ID 75Uy4jSjF]
29. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
30. Vollmer DD, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
31. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Jun 04. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
32. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jan 31. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
33. Gurtner S. Making the right decisions about new technologies: a perspective on criteria and preferences in hospitals. Health Care Manage Rev 2014;39(3):245-254. [doi: 10.1097/HMR.0b013e3182993b91] [Medline: 23727787]
34. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: a systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250. [doi: 10.1258/jhsrp.2010.009137] [Medline: 20592046]
35. Auster RD, Oaxaca RL. Identification of supplier induced demand in the health care sector. J Hum Resour 1981;16(3):327-342. [Medline: 7264294]
36. Richardson J. Supply and demand for medical care: or, is the health care market perverse? Aust Econ Rev 2001 Sep;34(3):336-352. [doi: 10.1111/1467-8462.00203]
37. Fitzgerald L, Ferlie E, Wood M, Hawkins C. Interlocking interactions, the diffusion of innovations in health care. Hum Relat 2002;55(12):1429-1449 [FREE Full text] [doi: 10.1177/0018726602128782213]
38. Kuehn BM. Is there an app to solve app overload? J Am Med Assoc 2015 Apr 14;313(14):1405-1407. [doi: 10.1001/jama.2015.2381] [Medline: 25871657]

39. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
40. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1(1):12 [FREE Full text]
41. Brouard B, Bardo P, Bonnet C, Mounier N, Vignot M, Vignot S. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? Ann Med 2016 Jun 27;48(7):509-515. [doi: 10.1080/07853890.2016.1195010] [Medline: 27348761]
42. 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]
43. Agarwal S, Labrique A. Newborn health on the line: the potential mHealth applications. J Am Med Assoc 2014 Jul 16;312(3):229-230. [doi: 10.1001/jama.2014.6371] [Medline: 24953141]
44. Broderick A, Haque F. Mobile health and patient engagement in the safety net: a survey of community health centers and clinics. Issue Brief (Commonw Fund) 2015 May;9:1-9. [Medline: 26040018]
45. 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]
46. Denis F, Yossi S, Septans A, Charron A, Voog E, Dupuis O, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. Am J Clin Oncol 2017 Oct;40(5):464-469. [doi: 10.1097/COC.0000000000000189] [Medline: 25811297]
47. Maguire R, Fox PA, McCann L, Miaskowski C, Kotronoulas G, Miller M, et al. The eSMART study protocol: a randomised controlled trial to evaluate electronic symptom management using the advanced symptom management system (ASyMS) remote technology for patients with cancer. BMJ Open 2017 Jun 06;7(5):e015016 [FREE Full text] [doi: 10.1136/bmjopen-2016-015016] [Medline: 28592577]
48. Comedh2020. WP3 developing methods to assess patient reported outcomes of mhealth URL: http://www.comedh2020.eu/wps/wcm/connect/site/comed/home/project/work+packages/wp3 [accessed 2019-04-18]
49. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/indevelopment/gid-ng10101/documents [accessed 2019-04-09]
50. Shuren J, Patel B, Gottlieb S. FDA Regulation of Mobile Medical Apps. J Am Med Assoc 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]
51. IQVIA 2017. The growing value of digital health URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/the-growing-value-of-digital-health.pdf?_=1547749163783 [accessed 2019-01-17]
52. 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 Dec 11;392(10146):521-530. [doi: 10.1016/S0140-6736(18)31270-4] [Medline: 30017550]
53. US Food and Drug Administration. 2018. Guidances with Digital Health Content URL: https://www.fda.gov/MedicalDevices/DigitalHealth/ucm562577.htm [WebCite Cache ID 75V09mvFx]
54. European Commission. Report of the Working Group on mHealth assessment guidelines URL: https://tinyurl.com/y5vtgtfs [accessed 2019-01-17] [WebCite Cache ID 75V0TSQb9]
55. US Food and Drug Administration. 2018. International Medical Device Regulators Forum (IMDRF) URL: https://www.fda.gov/MedicalDevices/InternationalPrograms/IMDRF/default.htm [WebCite Cache ID 75V0ZQKZI]
56. Lee WC, Plowman RS, Savage GM. Oncology value framework in the era of digital health technology: a patient-centric approach. Value & Outcomes Spotlight 2018;4(5):A [FREE Full text]
57. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
58. Food & Drug Administration. Cybersecurity URL: https://www.fda.gov/MedicalDevices/DigitalHealth/ucm373213.htm [accessed 2019-01-17] [WebCite Cache ID 75V1dAHff]
59. European Union Law. General Data Protection Regulation URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2019-01-17] [WebCite Cache ID 75V1t3skC]
60. Bordonaro S, Raiti F, Di Mari A, Lopiano C, Romano F, Pumo V, et al. Active home-based cancer treatment. J Multidiscip Healthc 2012;5:137-143 [FREE Full text] [doi: 10.2147/JMDH.S31494] [Medline: 22807631]
61. Khandelwal N, Duncan I, Ahmed T, Rubinstein E, Pegus C. Oral chemotherapy program improves adherence and reduces medication wastage and hospital admissions. J Natl Compr Canc Netw 2012 May;10(5):618-625. [Medline: 22570292]
62. Molassiotis A, Brearley S, Saunders M, Craven O, Wardley A, Farrell C, et al. Effectiveness of a home care nursing program in the symptom management of patients with colorectal and breast cancer receiving oral chemotherapy: a randomized, controlled trial. J Clin Oncol 2009 Dec 20;27(36):6191-6198. [doi: 10.1200/JCO.2008.20.6755] [Medline: 19917849]
63. Agboola S, Flanagan C, Searl M, Elfiky A, Kvedar J, Jethwani K. Improving outcomes in cancer patients on oral anti-cancer medications using a novel mobile phone-based intervention: study design of a randomized controlled trial. JMIR Res Protoc 2014 Dec 23;3(4):e79 [FREE Full text] [doi: 10.2196/resprot.4041] [Medline: 25537463]

64. Loeb S, Bayne CE, Frey C, Davies BJ, Averch TD, Woo HH, American Urological Association Social Media Work Group. Use of social media in urology: data from the American Urological Association (AUA). BJU Int 2014 Jun;113(6):993-998 [FREE Full text] [doi: 10.1111/bju.12586] [Medline: 24274744]

Abbreviations

FDA: US Food and Drug Administration
HPSM: Heckman probit selection model
IP: internet protocol
mHealth: mobile health
PSBRA: propensity score–based regression adjustment

Edited by G Eysenbach; submitted 02.02.19; peer-reviewed by P Innominato, F Levi, B Koontz, C Berkowitz, D Check, A Dinkel, Z Su; comments to author 02.03.19; revised version received 28.04.19; accepted 30.04.19; published 06.09.19.

Please cite as:
Tarricone R, Cucciniello M, Armeni P, Petracca F, Desouza KC, Hall LK, Keefe D
Mobile Health Divide Between Clinicians and Patients in Cancer Care: Results From a Cross-Sectional International Survey
JMIR Mhealth Uhealth 2019;7(9):e13584
URL: https://mhealth.jmir.org/2019/9/e13584/
doi:10.2196/13584
PMID:31493318

Original Paper

The Association Among Social Support, Self-Efficacy, Use of Mobile Apps, and Physical Activity: Structural Equation Models With Mediating Effects

Taotao Wang[1], MB; Mengyuan Ren[2], MB; Ying Shen[1], BS; Xiaorou Zhu[1], MM; Xing Zhang[1], MM; Min Gao[1], MS; Xueying Chen[1], MB; Ai Zhao[1], PhD; Yuhui Shi[1], PhD; Weizhong Chai[3], PhD; Xinchuan Liu[4*], PhD; Xinying Sun[1*], PhD

[1]Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

[2]Institute of Reproductive and Child Health, School of Public Health, Peking University, Beijing, China

[3]Department of Nutrition and Food Hygiene, School of Public Health, Peking University, Beijing, China

[4]School of Journalism and Communication, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xinying Sun, PhD
Department of Social Medicine and Health Education
School of Public Health
Peking University
No 38, Xueyuan Road
Haidian District
Beijing, 100191
China
Phone: 86 13691212050
Email: xysun@bjmu.edu.cn

Abstract

Background: Physical inactivity is a risk factor for chronic noncommunicable diseases. Insufficient physical activity has become an important public health problem worldwide. As mobile apps have rapidly developed, physical activity apps have the potential to improve the level of physical activity among populations.

Objective: This study aimed to evaluate the effect of physical activity apps on levels of physical activity among college students.

Methods: A Web-based questionnaire was used to survey college students in Beijing from December 27, 2017, to January 5, 2018. According to a previous survey, 43% of college students using physical activity apps and 36% of those who never used such apps achieved the physical activity recommendations. In this study, the sample size was calculated to be 500. The questionnaire consisted of 5 parts: the use of physical activity apps, sports habits, social support, self-efficacy, and social demographic information. Structural equation modeling was used to test the relationships between the use of physical activity apps, self-efficacy, social support, and level of physical activity.

Results: Of the 1245 participants, 384 college students (30.8%) used physical activity apps (in the past month). Of these 384 students, 191 (49.7%) gained new friends via the app. College students who were using physical activity apps had a higher level of physical activity and higher scores for social support and self-efficacy (P<.001) than those who did not use such apps. The use of physical activity apps significantly affected the mediating effect of physical activity level through social support (beta=.126; P<.001) and self-efficacy (beta=.294; P<.001). Gender played an important role in app use, self-efficacy, and physical activity in the mediation model: male users spent more time on physical activity and had higher self-efficacy scores (P<.001).

Conclusions: This study focused on college students in Beijing and found that the use of physical activity apps is associated with higher physical activity levels among these students. This effect is mainly through the mediation effect of social support and self-efficacy, rather than the direct effect of physical activity apps. The use of physical activity apps is associated with a higher social support level and higher self-efficacy score. Furthermore, a high social support level and high self-efficacy score are associated with higher physical activity levels.

(JMIR Mhealth Uhealth 2019;7(9):e12606) doi:10.2196/12606

KEYWORDS

mobile apps; physical activity; social support; self-efficacy; structural equation modeling

Introduction

Physical activity is an important foundation of general health. People's way of life has changed dramatically, including changes in diet, a decrease in physical activity, and increase in tobacco use [1]. Insufficient physical activity has become an important public health problem worldwide [2] and is associated with the progression of many chronic diseases [3]. A survey on disease risk factors conducted by the World Health Organization (WHO) indicated that physical inactivity has become one of the 4 leading causes of death [4]. The International Physical Activity Questionnaire (IPAQ) classifies physical activity as occupational, domestic, traffic, and leisure activities. In terms of the metabolic equivalent of various activities, the intensity of physical activity can be delineated as high, moderate, and low [5], which are distinguished according to the change in heartbeat, physiological sensation, and energy consumption. Moderate-intensity physical activities (MPA) are those that may cause slight sweating and make the heart beat slightly faster. Vigorous-intensity physical activities (VPA) lead to excessive sweating and make the heart beat significantly. The WHO recommends that adults aged 18 to 64 years allocate at least 150 min a week for MPA, or 75 min a week for VPA, or an equivalent combination of MPA and VPA [4].

Globally, 80.3% of adolescents aged 13 to 15 years do not achieve the current physical activity recommendations [6]. In China, only 18.7% of adults aged 20 to 69 years achieve the current physical activity recommendations [7]. Furthermore, physical activity tends to decline with age throughout adolescence [8,9] and tracks into adulthood [10]. In the case of college students, who are in the transition from adolescence to adulthood, developing good physical activity habits can help them maintain a good physical health. Thus, the college age is an important time for intervention. Many factors affect physical activity levels. For example, demographic characteristics such as age, gender, and education level are associated with physical activity level, environmental factors, individual physical health, and psychological factors [11-13]. Furthermore, numerous studies have shown that social support and self-efficacy are among the most important factors affecting the physical activity level [14-16]. As these 2 factors are so important and can be improved through interventions [17], we focused on them in this study.

With the development of the internet and mobile phones, the number of mobile phone apps for improving physical activity has been increasing in recent years [18]. In China, many mobile phone apps, such as KEEP, Gu Dong, and Yue Dongquan, can help people participate in physical activity. Physical activity apps can be roughly delineated as fitness and running apps. This review showed that physical activity apps lack sufficient inclusion in health behavior change theories and evidence-based content [19]. However, using physical activity apps provides more convenience in terms of use and more flexibility regarding time. As such, physical activity apps have become increasingly popular [20].

According to a statistical report of the International Telecommunication Union [21], the total number of global internet users has increased from 1.99 billion in 2010 to 3.38 billion in 2016. By the end of 2017, the number of cellular mobile subscriptions totaled around 700 million, with a penetration rate of 70% even in the less developed countries. In addition, the scale of Chinese internet users totaled 802 million, of which mobile phone users accounted for 98.3%, as of June 2018 [22]. All types of scales for mobile app users are rising. In addition, users of mobile physical activity apps accounted for 78% of all users of mobile health–related apps in 2014, up from 39% in the previous year [23].

The influence of physical activity apps has been examined at the level of physical activity. Most existing studies have confirmed the promotional effect of physical activity apps on physical activity. One study of undergraduates at the Southeast University, in Jiangsu Province, China, showed that only 27.85% of college students had never used physical activity apps [24], suggesting that physical activity apps may be extremely popular among Chinese college students. Another study [25] confirmed the role of physical activity apps in increasing the time spent on physical activity and suggested that such changes could further increase people's self-efficacy. Harries [26] demonstrated the significant effect of physical activity apps on young people who lack physical activity but noted no significant increase in social feedback. However, other studies have highlighted that although physical activity apps may play a role in increasing physical activity, the mechanism is not clear and needs to be improved to achieve better results [27,28].

Social support and self-efficacy may be the factors that affect physical activity. Social support is defined as the exchange of resources between at least 2 individuals perceived by the provider or recipient to be intended to enhance the well-being of the recipient [29]. Self-efficacy is defined as the belief that one can successfully execute the behavior required to produce the desired outcomes [30]. Shariff [14] indicated that social support from family, friends, and coaches can influence the behavior of teenagers with regard to sports and psychological development and social competence. Zhang [15] reported that social support can play such a role with certain conditions. As such, the actual effect of social support on physical activity is not clear [31]. With regard to the effect of physical activity apps on self-efficacy, many results suggest that self-efficacy is a powerful predictor of physical activity. Bezjak [16] found a significant correlation between self-efficacy and involvement of college students in physical activity. Smith [32] showed that self-efficacy, as a mediating variable between physical activity apps and physical activity, affects its effect strength. This finding elaborates the specific mechanism of physical activity apps in improving physical activity levels.

In general, research on the influence of physical activity apps on physical activity level is incomplete. The results of existing works are contradictory, which may be attributed to the subjects and the region of study. Most studies have confirmed that the use of physical activity apps can promote physical activity levels. However, the degree of influence and mechanism of action remain unclear. As a special group, college students are highly adaptable and need to cultivate good physical activity habits. As such, college students represent an ideal research sample when investigating interventions.

Therefore, the study aimed to elucidate the mediating role of social support and self-efficacy in physical activity apps and physical activity, and on the specific mechanism of how use of physical activity apps is associated with physical activity level. Our findings will contribute to the development of physical activity interventions and improvement of health at the national level.

These are hypotheses of this study:

- The use of physical activity apps is associated with higher physical activity levels among college students.
- Physical activity apps may be associated with college students' physical activity levels, mainly through the mediating effect of social support and self-efficacy.
- According to the conclusions of this study, a set of potential measures hypothesized to improve college students' physical activity level with the help of physical activity apps can be formulated.

Methods

Participants

We conducted a closed survey. An electronic questionnaire was used to survey college students in Beijing from December 27, 2017, to January 5, 2018. Convenience sampling was adopted to issue the questionnaire. According to a previous survey, 43% of college students who use physical activity apps and 36% of those who had never used them achieved the WHO's physical activity recommendations. Thus, the sample size was calculated to be 500. We contacted student union leaders and teachers at several colleges in Beijing and sent the electronic questionnaire through their WeChat (the Chinese version is Weixin) group. To obtain cooperation, an incentive (about 2 yuan per participant) was paid out with each questionnaire.

The questionnaire was distributed to a wide range of students through the WeChat group, not sent to individuals. Thus, it was difficult to calculate the exact number of questionnaires sent and the response rate. However, it was estimated that the response rate ranged from 70% to 95% in the different WeChat groups. Users first had to log in using their WeChat account to prevent multiple entries from the same individual. Only those who agreed to participate and completed the questionnaire could submit the completed questionnaire. Therefore, we could not calculate the exact view rate and participation rate.

We controlled the filling range of the data when designing the questionnaire to avoid the occurrence of invalid data. After collecting the electronic questionnaires, we manually checked the original data. Strict criteria were used to screen valid questionnaires. Only completed questionnaires with no missing data and no logical errors were considered valid. Otherwise, the questionnaire was excluded. For example, if the participant did not indicate a specific amount of time spent on physical activity, his/her questionnaire would be excluded. Finally, 1476 questionnaires were completed. After excluding invalid questionnaires, 1245 valid questionnaires were used in this analysis, with an efficiency rate of 84.35%.

Measures

The questionnaire consisted of 5 parts: use of physical activity apps, sports habits, social support, self-efficacy, and social demographic information. The electronic questionnaire consisted of 5 pages, and each page included 15 items. We used adaptive questioning (only conditionally displayed based on responses to other items) to reduce the number and complexity of questions. In addition, respondents were able to review and change their answers before submitting the questionnaire. Before fielding the questionnaire, we asked experts for their advice and tested the usability and technical functionality of the electronic questionnaire on a small scale.

In the research, those who used physical activity apps in the past month were defined as current users, those who had used physical activity apps before but not in the past month were defined as past users, and those who had never used physical activity apps were defined as nonusers.

The measurement of physical activity habits was based on the Chinese simplified IPAQ [33], which is mainly used to evaluate people's physical activity level against the recommended level. It can also be used to evaluate the results of a physical activity intervention. In 2004, Chinese scholars Qu and Li studied the reliability and validity of the Chinese version of IPAQ. Their results indicated that the retest reliability and validity are higher than or equal to the questionnaire for the same use [34]. Thus, the Chinese version of the questionnaire was used in the present research.

To measure social support, we consulted Chogahara's research on older adults [33] and Cavallo's study on college girls [35]. Social support was delineated as partnership, information, and respect support types. Then, 2 subevaluation indicators were selected from each aspect, which respondents graded according to the actual frequency of occurrence on a scale ranging from 1 (never) to 5 (often). In addition, a homogeneity reliability test and factor analysis were performed. The results of the sphericity test confirmed that Cronbach alpha was .917, KMO (Kaiser-Meyer-Olkin)=0.887, and the P value was $<.001$. The reliability and validity of this scale were considered good.

The evaluation of self-efficacy was based on Wang's study on adolescents with disabilities [36] and Ashraf's research on adolescent girls [37]. A 5-point Likert scale with 6 items was used. The participants rated every item of the scale, according to their confidence level regarding participating in physical activity in different situations. Each item was scored on a scale ranging from 1 (very uncertain) to 5 (very certain). In addition, a homogeneity reliability test and factor analysis were performed. The results of the sphericity test showed that the

Cronbach alpha was .914, KMO=0.906, and the *P* value was ＜.001. Again, the reliability and validity of the scale were considered good.

Ethics Statement

This study has been approved by the Peking University Institutional Review Board Office. Before participating in the study, each participant was informed of the purpose of the investigation and the duration of the survey and assured that the results would be used only for the purpose of this study, and that their privacy would be guaranteed. If participants did not want to participant in the survey, they could simply turn off the electronic questionnaire and drop out. If the questionnaire was completed and submitted, the participant was considered to have provided informed consent. Only those who voluntarily agreed to participate in the survey were included in the research.

Statistical Analyses

Data were analyzed using SPSS (IBM) version 22.0 and Mplus (Linda Muthén & Bengt Muthén) version 7.0. A descriptive analysis was conducted to analyze the social demographic information and use of physical activity apps. A chi-square test was carried out on the difference in rates between the groups. The Kruskal-Wallis rank test and Mann-Whitney *U* test were performed for the differences in ordinal level variables between groups. The correlation between variables was tested linearly. Structural equation modeling (SEM) was used to test the relationships between the use of physical activity apps, self-efficacy, social support, and level of physical activity. A *P* value <.05 was considered statistically significant. SEMs were employed for the mediation effect analysis.

First, we indicated self-efficacy as a mediator between physical activity app usage (app use) and physical activity to test whether physical activity app usage has a significant indirect effect on physical activity through self-efficacy and whether this indirect effect can completely explain the relation between app use and physical activity. App use was clarified as a 3-level categorical predictor (current users=3, past users=2, and nonusers=1), with self-efficacy as the mediator and physical activity as the outcome variable. The 6-item self-efficacy scale indicators were categorized in 3 parts (bad mood, support deficiency, and time deficiency), representing the manifest indicators by which self-efficacy was significantly explained (*P*<.001). The durations of VPA and MPA were also set as manifest indicators by which physical activity can also be significantly explained.

Social support was selected as the second mediator in the mediation model. The 5-item social support scale indicators were classified in 3 parts (partnership, information, and respect support), representing the manifest indicators by which social support was significantly explained (*P*<.001).

Finally, we added age and gender as covariables to the model. Gender was a 2-level categorical variable, where male=1 and female=2, whereas age was a continuous variable. The model was set to test for whether the direct and indirect effects of app use on physical activity changed with the combination of these covariables and how gender and age affected app use.

Results

Demographic Information

Of the 1245 participants, 466 were male (37.4%). Other demographic characteristics are shown in Table 1.

Usage of Physical Activity Apps

There was a difference between males and females in the distribution of the 3 types of users (χ^2=26.6, *df*=2, *P*<.001). Other information are shown in Figure 1.

Of 384 current users, 99 (25.8%) indicated having used physical activity apps more than 4 times a week, 39.1% (150/384) used them 2 to 3 times a week, and 35.1% (135/384) used apps less than once a week. The popular functions current users mostly used were *calculating the number of steps* (54.9%, 211/384), *recording movement* (52.9%, 203/384), the *training plan* (49.2%, 189/384), *recording calorie consumption* (40.6%, 156/384), and the *video coach* (40.6%, 156/384). Functions such as *ranking campaign* (20.8%, 80/384), *sharing exercise data on social media* (11.7%, 45/384), and *stimulating users* (7.3%, 28/384) were less used by the current users.

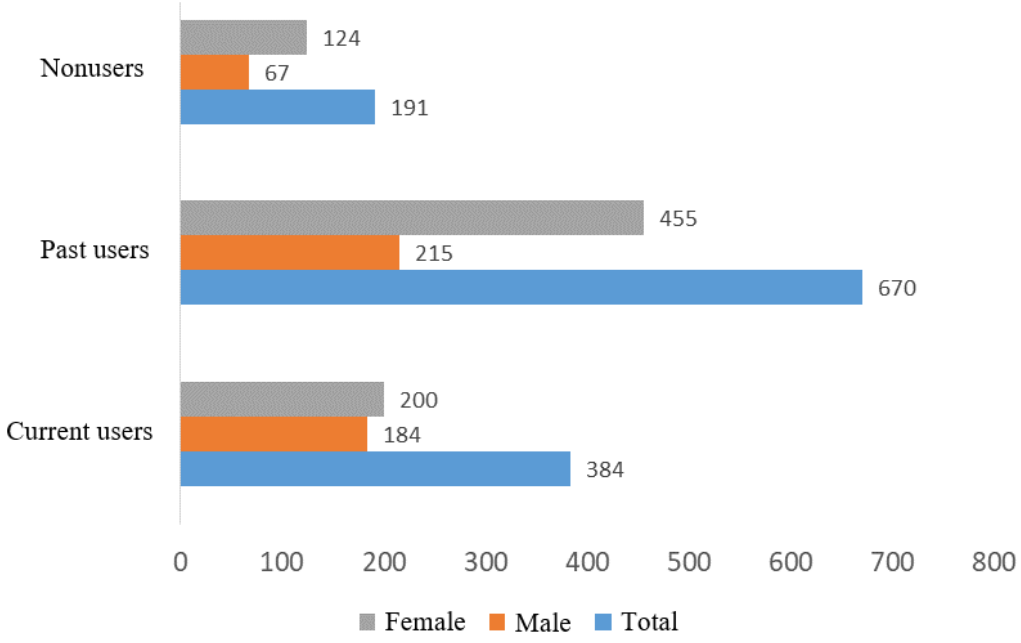
Table 1. Participants' demographic information.

Demographics	Total (N=1245)	Male (n=466)	Female (n=779)	Chi-square *t* test (*df*)[a]	*P* value[b]
Age (years), mean (SD)	20.5 (2.6)	20.8 (2.7)	20.3 (2.4)	2.99 (1244)	.003
Body mass index (kg/m^2), mean (SD)	22.3 (6.5)	23.6 (7.0)	21.5 (5.8)	5.55 (1244)	<.001
Grade, n (%)				13.0 (4)	.01
One	190 (15.3)	52 (11.2)	138 (17.7)		
Two	357 (28.7)	135 (29.0)	222 (28.5)		
Three	295 (23.7)	127 (27.3)	168 (21.6)		
Four and five	286 (23.0)	104 (22.2)	182 (23.3)		
Graduate student or above	117 (9.4)	48 (10.3)	69 (8.9)		
Birthplace, n (%)				6.6 (2)	.04
Urban area	858 (68.9)	305 (65.5)	533 (71.0)		
Rural area	360 (28.9)	146 (31.3)	214 (27.5)		
International students	27 (2.2)	15 (3.2)	12 (1.5)		
Ethnicity, n (%)				3.7 (1)	.06
Han	1079 (86.7)	415 (89.1)	664 (85.2)		
Others	166 (13.3)	51 (10.9)	115 (14.8)		
Average monthly expense (yuan), n (%)				6.1 (4)	.19
<500	36 (2.9)	18 (3.9)	18 (2.3)		
500 to 1000	191 (15.3)	80 (17.2)	111 (14.2)		
1000 to 1500	374 (30.0)	139 (29.8)	235 (30.2)		
1500 to 2000	319 (25.6)	107 (23.0)	212 (27.2)		
>2000	325 (26.1)	122 (26.2)	203 (26.1)		

[a]The age and the body mass index was tested using a *t* test. Others were tested using a Pearson chi-square test.

[b]The *P* value here refers to whether the difference in these demographic characteristics between participants of different genders was statistically significant.

Figure 1. Number of each type of app user (male and female).



Physical Activity

Normality tests were conducted for concerning variables, and the results of these did not support the null hypothesis. Thus, we used quartiles to describe the distribution. The median duration of VPA per week was 30 min and 40 min for MPA. Overall, only 48.3% (602/1245) of the participants attained the WHO's weekly standard amount of physical activity.

For all types of physical activity, we noted a significant difference between groups (Table 2): Current users reported the most VPA and MPA time consumption, and nonusers reported the least time spent on these physical activity types. For social support, current users had the highest score (median=21), followed by past users (median=18). Nonusers scored the lowest (median=17; $P<.001$). There was also a significant difference in the self-efficacy scores of these groups ($P<.001$).

The physical activity rates of participants who attained the WHO's recommended value in the 3 groups were calculated. A cross-table test was performed, revealing a significant difference between the 3 physical activity apps usage groups ($\chi^2=385.0$; $P<.001$). Current users had the highest rate (66.75%) and nonusers had the lowest (35.94%; Table 2).

Relationship Between Physical Activity App Usage, Physical Activity, Social Support, and Self-Efficacy

The zero-order correlations between measures are displayed in Table 3. Current users scored higher for social support and self-efficacy and reported more physical activity. Gender (male=1, female=0) was significantly correlated with all measurements, demonstrating that this factor might be an assignable confounding variable. Self-efficacy was significantly correlated with all measurements, demonstrating that it could be an intensive mediator.

Age was also significantly correlated with all measurements, except physical activity app usage, revealing that it could also be a confounding variable. BMI was significantly correlated with gender, age, and self-efficacy, but not with social support and physical activity. Social support was significantly correlated with all measurements, except BMI, and the same was observed for physical activity. Overall, BMI may not be an important factor in our research.

Table 2. Differences between current users, past users, and nonusers.

Items per week	Current users (n=384)			Past users (n=670)			Nonusers (n=191)			Kruskal-Wallis and Mann-Whitney *U* rank sum tests, chi-square (*df*)[b]	*P* value
	P25[a]	M[a]	P75[a]	P25[a]	M[a]	P75[a]	P25[a]	M[a]	P75[a]		
Time, VPA[c] (min)											
Total	23	60	126	0	30	80	0	2	73	104.36 (1242)	<.001
Male	17	60	120	2	30	90	0	20	90	43.87 (463)	<.001
Female	37	90	180	0	25	60	0	0	60	49.42 (776)	<.001
Time, MPA[d] (min)											
Total	20	60	150	1	30	84	0	20	90	59.11 (1242)	<.001
Male	20	60	120	2	40	100	0	30	120	26.31 (463)	<.001
Female	22	90	180	0	30	80	0	20	60	27.45 (776)	<.001
Walk time (min)											
Total	50	140	280	45	120	210	21	105	210	9.42 (1242)	.009
Male	60	140	280	50	125	250	40	140	315	0.27 (463)	.88
Female	45	140	247	45	120	210	11	90	120	14.60 (776)	.88
Sedentary behavior (min)											
Total	60	600	1698	120	840	2100	50	630	2100	6.53 (1242)	.04
Male	60	420	1440	60	720	2100	60	630	2100	0.46 (463)	.80
Female	90	760	2100	180	900	2100	42	647	1890	14.34 (776)	.80
Social support score											
Total	14	21	25	12	18	22	11	17	21	156.67 (1242)	<.001
Male	14	19	24	12	18	22	11	17	21	21.87 (463)	<.001
Female	14	21	25	12	18	22	12	16	19	20.03 (776)	<.001
Self-efficacy score											
Total	16	20	24	12	16	19	10	16	19	124.66 (1242)	<.001
Male	16	19	23	12	17	21	11	17	21	38.72 (463)	<.001
Female	17	21	25	12	16	18	9	15	18	75.93 (776)	<.001
Rate of reaching the WHO[e] standard (%)											
Total	66.7			42.1			35.9			385.0 (1242)	<.001
Male	77.4			49.8			46.3			23.3 (463)	<.001
Female	60			38.5			29.8			19.5 (776)	<.001

[a]P25 is the upper quartile, M is the median, and P75 is the lower quartile.

[b]The rate of reaching the WHO standard was tested using a Pearson chi-square test. Others were tested using Kruskal-Wallis and Mann-Whitney *U* rank sum tests.

[c]VPA: vigorous-intensity physical activity.

[d]MPA: moderate-intensity physical activity.

[e]WHO: World Health Organization.

Table 3. Zero-order correlations between measures.

Measures[a]	(1)	(2)	(3)	(4)	(5)	(6)	(7)
App use (1)	1.000	—[b]	—	—	—	—	—
Gender (2)[c]	–.112[d]	1.000	—	—	—	—	—
Age (3)	.050	.084[d]	1.000	—	—	—	—
Bpdy mass index (4)	–.006	.156[d]	–.230[d]	1.000	—	—	—
Social support (5)	.375[d]	.133[d]	–.130[d]	–.047	1.000	—	—
Self-efficacy (6)	.298[d]	.158[d]	–.119[d]	–.066[d]	–.808[d]	1.000	—
Physical activity (7)	.111[d]	.142[d]	–.060[d]	–.031	–.288[d]	–.273[d]	1.000

[a]Numbers in parentheses correspond to column numbers.

[b]To avoid duplication of data and to keep tables concise, there are empty cells in the table.

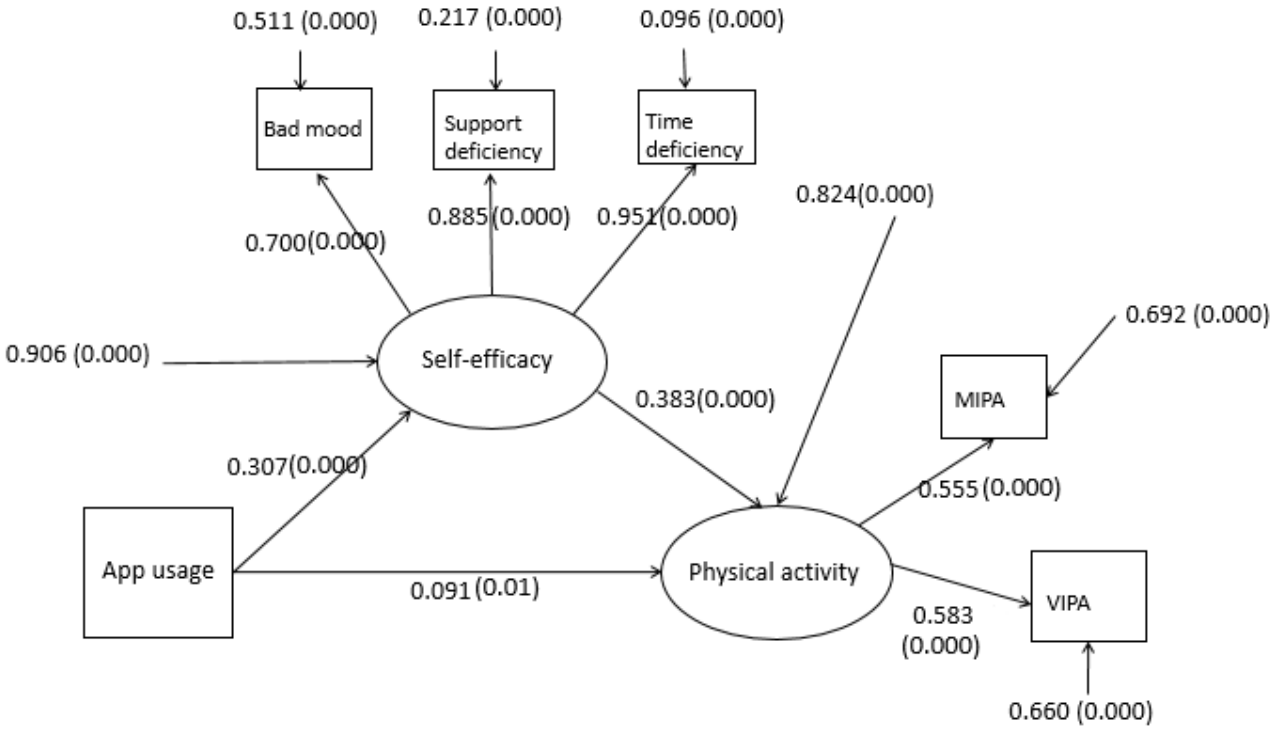
[c]Gender: 1=male, 0=female.

[d]Correlation is significant at the .001 level (2-tailed).

Simple Mediation Model

The standardized path coefficients of the simple mediation model are displayed in Figure 2. The model fit indices were as follows: χ^2=2.9, comparative fit index (CFI)=.995, standardized root mean square residual (SRMR)=.018, and root mean square error of approximation (RMSEA)=.039. The single mediation model indicated that the path coefficient of the indirect effect was .118, slightly higher than the coefficient of the direct effect (.091). Furthermore, it showed that the use of an app was weakly associated with a higher physical activity level and strongly associated with a higher physical activity level through self-efficacy.

Figure 2. Single mediation model, in which app-usage is a three-level categorical predictor (current users=3, past users=2, and non-users=1). Self-efficacy is described by bad mood, support deficiency, and time deficiency; VIPA is vigorous-intensity physical activity and MIPA is moderate-intensity physical activity. (0.000) represents significant path coefficients at the .001 level and (0.01), at the .05 level.



Multi-Serial Mediator Model

Self-efficacy was affected by social support directly, meaning that social support had a second role in this model, namely as a serial mediator between app use and self-efficacy, according to self-efficacy theory. Furthermore, social support also acted as a sole mediator between app use and physical activity. The standardized path coefficients are displayed in Figure 3. The model fit indices were as follows: χ^2=3.4, CFI=.990, SRMR=.018, RMSEA=.044. The multi-serial mediator model demonstrated that the use of an app was associated with higher physical activity through social support. In addition, the combination of social support in this model did not change the significant indirect effect of self-efficacy and direct effect of app use on physical activity. Moreover, the association between app use and higher physical activity through self-efficacy and social support was stronger than the association directly between app use and higher physical activity. Compared with the single mediation model, (1) the combination of social support did not change the lower direct effect compared with the total indirect effect; (2) the association between app use and higher physical activity level and between app use and higher physical activity level through self-efficacy weakened with the combination of social support and (3) the association between app use and higher physical activity level through social support was weaker than the association between app use and higher physical activity level through self-efficacy.

Multi-Serial Mediator Model Containing Covariables

The standardized path coefficients of the multi-serial mediator model containing covariables are displayed in Figure 4.

Figure 3. Multi serial mediator model, in which social support is described by three manifest indicators (partnership support, information support, and respect support). Self-efficacy is described by bad mood, support deficiency, and time deficiency; VIPA is vigorous-intensity physical activity and MIPA is moderate-intensity physical activity. (0.000) represents significant path coefficients at the .001 level and (0.01), at the .05 level.

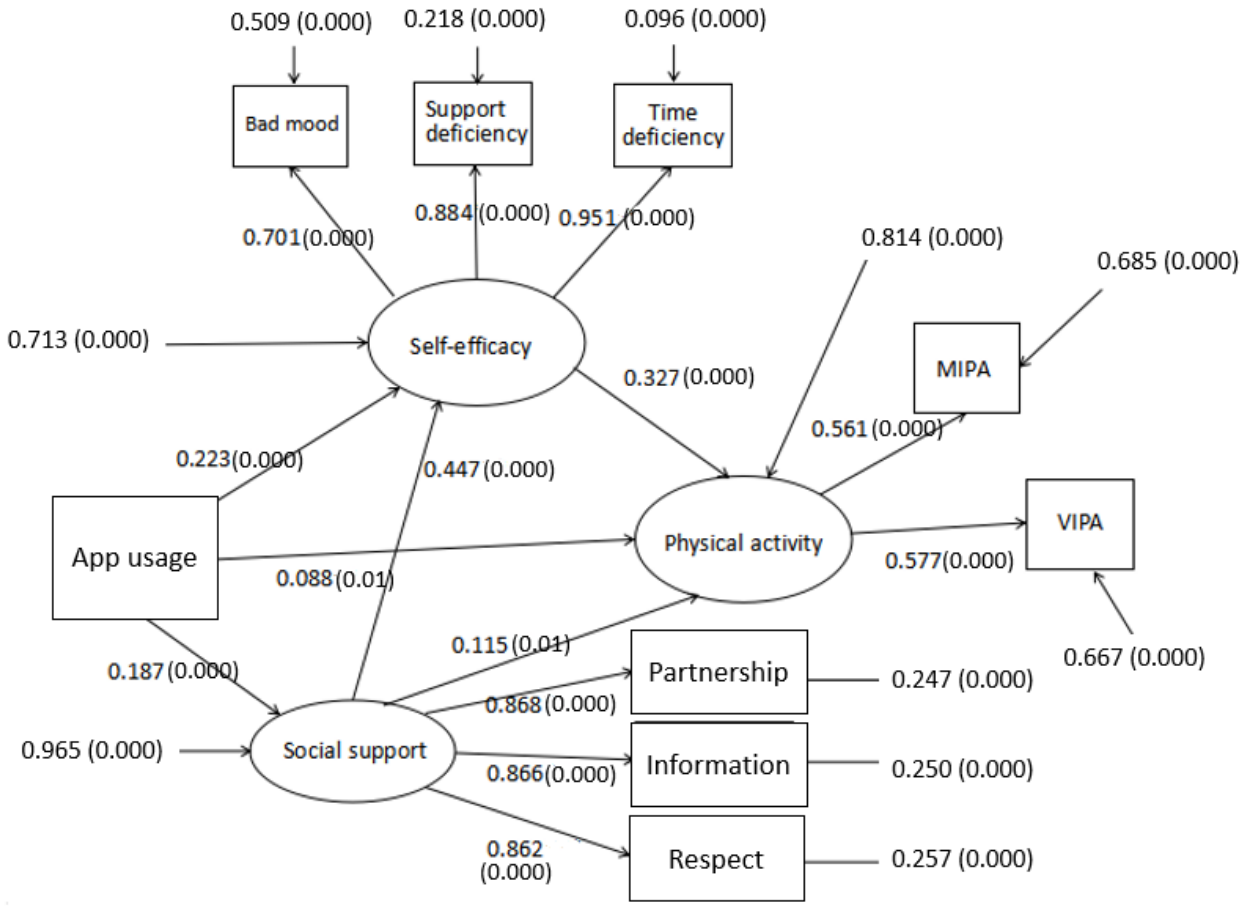
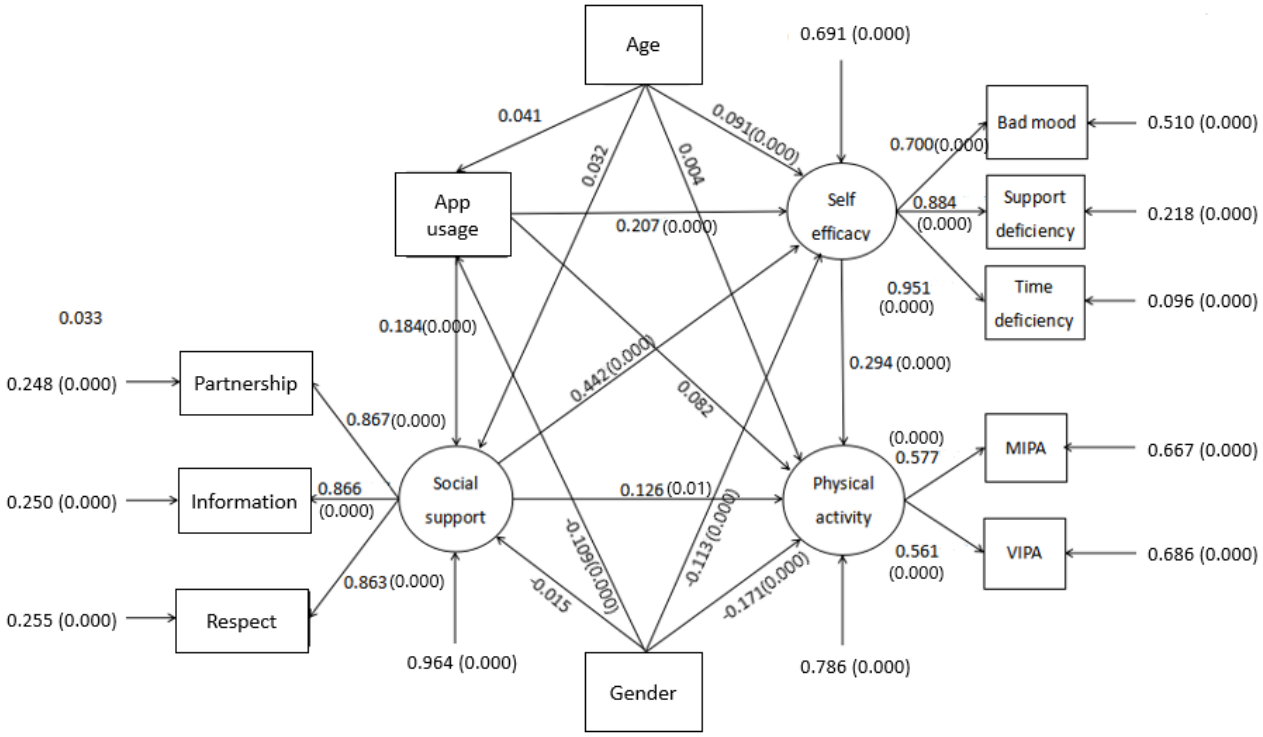


Figure 4. Multi serial mediation model containing age and sex. Age is a continuous co-variable; gender is described as a two-level categorical variable where male=1 and female=2. Social support is described by three manifest indicators (partnership support, information support, and respect support). Self-efficacy is described by bad mood, support deficiency, and time deficiency; VIPA is vigorous-intensity physical activity and MIPA is moderate-intensity physical activity. (0.000) represents significant path coefficients at the .001 level and (0.01), at the .05 level.



The model fit indices were as follows: χ^2=3.1, CFI=0.987, SRMR=0.018, RMSEA=0.041. The model showed that the path coefficients from gender to app use, self-efficacy, and physical activity were all significant at the .001 level, whereas the path coefficient was only significant from age to self-efficacy (*P*<.001). For gender, males had a higher app use rate and higher social support and self-efficacy scores. Furthermore, they reported more physical activity. Regarding age, older students tended to obtain higher scores for self-efficacy. Adding age and gender into this model did not change the association between app use and higher physical activity level through self-efficacy and social support. However, the model indicated that there was no longer a direct association between app use and higher physical activity level after including the covariant items. The standardized direct effect calculated in Mplus7.0 was 0.082 (*P*=.05) slightly lower than the total indirect effect of social support (0.084; *P*<.001), of which self-efficacy was 0.061 (*P*<.001) higher than social support (0.023; *P*<.05). Compared with the multi-serial mediation model, this model provided the following information: (1) the combination of age and gender did not change the situation of the weaker association directly between app use and higher physical activity, compared with the association through self-efficacy and social support; (2) with the combination of age and gender, the association through self-efficacy and social support became weaker; (3) the association through social support was still weaker than the association through self-efficacy.

Discussion

Principal Findings and Interpretation

This study confirmed the correlation between the use of physical activity apps and the level of physical activity. The use of physical activity apps is associated with a higher social support level and higher self-efficacy score. Furthermore, a high social support level and high self-efficacy score are associated with higher physical activity levels.

The results showed that 13.3% of college students had never used physical activity apps, less than the percentage of students who never used physical activity apps which found in a survey of college students in Jinan [38]. However, the difference may be attributed to regional and methodological differences. Furthermore, this research confirmed the association between use of physical activity apps and physical activity levels, coinciding with a previous research on undergraduates in Southeast University [23].

Research in the United States [39] has shown that 55.4% of adults aged 18 to 24 years met the aerobic guidelines of the US Department of Health and Human Services, whereas 26.9% were totally inactive. In Beijing, 27.3% of students exercised regularly (3 times a week for 30 minutes each time) [40]. In this study, 66.75% of college students who were using physical activity apps met the WHO's recommended level. This could be attributed to both population differences and the different methodologies used to measure activity. However, it can still

be concluded that college students who were using physical activity apps had a higher level of physical activity. Among college students who had never used physical activity apps, 35.94% met the WHO's recommended level. This is perhaps because of school policies and physical education requirements, and perhaps because college students have received health education on physical activity from various sources. As far as we know, some colleges in Beijing require that students spend a certain amount of time on physical activity to earn credits. Therefore, particular proportion of college students in Beijing have a high level of physical activity. This can be further explored in future investigations. Thus, while the physical activity level of college students in Beijing is not low, but it still needs improvement.

Our findings showed that 49.7% of college students who used physical activity apps had app friends. App friends are network friends who follow each other on physical activity apps and engage in interactive behaviors such as *liking* a post. This kind of network friend can be friend in life, or not. According to the statistical analysis of the social support scores of different sources and classifications, social support from friends is related to VPA. This might be because friends on apps are mostly network strangers, and thus, the correlation is weaker. Social support requires specific conditions. The role of social support from networks only comes into play after a change in personal attitudes toward physical activity [15]. This tendency may also lead to more social support from actual friends than app friends. This study found that different types of social support had different effects on physical activity. Respect support is associated with VPA and total duration of physical activity, whereas partner support is associated with MPA. This may be attributed to different social support mechanisms, which can be further be explored and refined in subsequent studies.

Self-efficacy was significantly associated with physical activity, as revealed by earlier findings [41]. Among the app use groups, current users scored the highest for self-efficacy. Thus, app usage may strengthen college students' self-efficacy in exercise, confirming earlier findings [27]. As previously reported, self-efficacy mediates the relationship between app use and exercise [25], indicating the effect of app use on self-efficacy. Furthermore, self-efficacy then influences physical exercise. Indeed, social support from family members, friends, and coaches is considered to affect adolescents' physical activity–related behavior [14]. Self-efficacy was also associated with social support, indicating that self-efficacy may be predicted by social support to some extent. In a study on junior high school students in Shanghai, social support reportedly explained self-efficacy [42], which is consistent with our results. In the same study, social support and self-efficacy were employed in a mediation model, where self-efficacy mediated the effect of social support on satisfaction with youth physical activity. This study assumed social support as the first mediator of physical activity and mobile app usage, which can also affect self-efficacy.

The results of the multi-mediation model showed that the 2 mediators of social support and self-efficacy were both significant, with social support having a significant effect on self-efficacy. These results were consistent with those of previous studies that indicated the underlying mechanism between exercise app usage and physical activity. Moreover, for the indirect effect, self-efficacy had a greater impact (standardized coefficient: 0.073 of 0.094), indicating its contribution to the larger mediation effect compared with social support.

To control potential confounding variables, we selected age and gender as covariables in the mediation model. Our study appears to be the first to consider gender and age as covariables in a mediation model. The direct effect was no longer significant after controlling for these 2 variables in the model, indicating that the significant coefficient value between app use and physical activity may have been erroneous before covariables were added into the model. However, the controlled multi-mediation model confirmed the unchanged significant association of mediators (social support and self-efficacy) between app use and physical activity. This controlled mediation model also indicated that gender and age may be important covariables that need to be considered in future studies. This might be because gender is associated with self-efficacy and social support, and the older students become, the more they need to do, meaning they have less time for physical activity. Our research only focused on college students in Beijing, meaning that the age range was narrow. In future studies, more participants with a greater age range should be surveyed to obtain more convincing results.

This study had a number of limitations. First, given the characteristics of the electronic questionnaire, the samples were obtained through convenience sampling and the survey limited to several colleges in Beijing. At the same time, we could not obtain information on the differences in demographic characteristics of participants taking part and those who did not. As such, the representativeness of the survey participants may not be high. Second, the proportion of female students (62.6%) was higher than that of male students, with a notable statistical difference. This trend may be because the response rate for the questionnaire was higher for females. Third, we set strict data inclusion criteria, resulting in the slightly low efficiency of the questionnaire (84.35%). Finally, the survey was conducted in winter, which may have affected the level of physical activity reflected in the results. In future research, these limitations need to be addressed.

Conclusions

This study focused on college students in Beijing and found that the use of physical activity apps is associated with higher physical activity levels among them. This effect is mainly through the mediation effect of social support and self-efficacy, rather than the direct effect of physical activity apps. The use of physical activity apps is associated with a higher social support level and the higher self-efficacy score. Finally, a high social support level and high self-efficacy score are associated with higher physical activity levels.

Authors' Contributions

The authors appreciate the efforts of all the authors. TW and MR designed the study, collected the data, performed the statistical analysis, interpreted the results, and wrote the manuscript. YS, X Zhu, X Zhang, MG, and XC helped to improve the language of this paper and provided helpful suggestions on the manuscript. AZ, YS, WC, XL, and XS were responsible for reviewing the study design, manuscript, and all study results.

Conflicts of Interest

None declared.

References

1. Waxman A, World Health Assembly. WHO global strategy on diet, physical activity and health. Food Nutr Bull 2004 Sep;25(3):292-302. [doi: 10.1177/156482650402500310] [Medline: 15460274]
2. Guilbert JJ. The world health report 2002 - reducing risks, promoting healthy life. Educ Health (Abingdon) 2003 Jul;16(2):230. [doi: 10.1080/1357628031000116808] [Medline: 14741909]
3. Marcus BH, Williams DM, Dubbert PM, Sallis JF, King AC, Yancey AK, American Heart Association Council on Nutrition, Physical Activity,Metabolism (Subcommittee on Physical Activity), American Heart Association Council on Cardiovascular Disease in the Young, Interdisciplinary Working Group on Quality of Care and Outcomes Research. Physical activity intervention studies: what we know and what we need to know: a scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity); Council on Cardiovascular Disease in the Young; and the Interdisciplinary Working Group on Quality of Care and Outcomes Research. Circulation 2006 Dec 12;114(24):2739-2752. [doi: 10.1161/CIRCULATIONAHA.106.179683] [Medline: 17145995]
4. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010. [Medline: 26180873]
5. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
6. 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]
7. Gu JF. Interpretation of report on nutrition and chronic diseases of Chinese residents (2015). Acta Nutr Sin 2016;38(6):525-529. [doi: 10.3969/j.issn.0512-7955.2016.06.005]
8. Dumith SC, Gigante DP, Domingues MR, Kohl HW. Physical activity change during adolescence: a systematic review and a pooled analysis. Int J Epidemiol 2011 Jun;40(3):685-698. [doi: 10.1093/ije/dyq272] [Medline: 21245072]
9. Corder K, Sharp SJ, Atkin AJ, Griffin SJ, Jones AP, Ekelund U, et al. Change in objectively measured physical activity during the transition to adolescence. Br J Sports Med 2015 Jun;49(11):730-736 [FREE Full text] [doi: 10.1136/bjsports-2013-093190] [Medline: 24273308]
10. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195 [FREE Full text] [doi: 10.1159/000222244] [Medline: 20054224]
11. Wu SY, Zhang XX, Yang SS, Sun KG, Jia WL, Shao CX, et al. [Physical activity level and its influence factors among residents in one suburb district of Beijing]. Beijing Da Xue Xue Bao Yi Xue Ban 2016 Jun 18;48(3):483-490 [FREE Full text] [Medline: 27318912]
12. Choi J, Lee M, Lee JK, Kang D, Choi JY. Correlates associated with participation in physical activity among adults: a systematic review of reviews and update. BMC Public Health 2017 Apr 24;17(1):356 [FREE Full text] [doi: 10.1186/s12889-017-4255-2] [Medline: 28438146]
13. Boulton ER, Horne M, Todd C. Multiple influences on participating in physical activity in older age: developing a social ecological approach. Health Expect 2018 Feb;21(1):239-248 [FREE Full text] [doi: 10.1111/hex.12608] [Medline: 28768065]
14. Shariff AR, Mustaph R. Social support mechanism and input factors on catharsis predictors in sport. Procedia Soc Behav Sci 2010;7:588-591. [doi: 10.1016/j.sbspro.2010.10.079]
15. Zhang N, Campo S, Yang J, Janz KF, Snetselaar LG, Eckler P. Effects of social support about physical activity on social networking sites: applying the theory of planned behavior. Health Commun 2015;30(12):1277-1285. [doi: 10.1080/10410236.2014.940669] [Medline: 26086237]
16. Bezjak JE, Lee JW. Relationship of self-efficacy and locus of control constructs in predicting college students' physical fitness behaviors. Percept Mot Ski 2016 Aug 31;71(2):499-508. [doi: 10.2466/pms.1990.71.2.499]
17. Ory MG, Lee S, Han G, Towne SD, Quinn C, Neher T, et al. Effectiveness of a lifestyle intervention on social support, self-efficacy, and physical activity among older adults: evaluation of texercise select. Int J Environ Res Public Health 2018 Jan 30;15(2):E234 [FREE Full text] [doi: 10.3390/ijerph15020234] [Medline: 29385779]

18. Middelweerd A, Mollee JS, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
19. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
20. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 1;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
21. International Telecommunication Union. The State of Broadband: Broadband Catalyzing Sustainable Development. Geneva, Switzerland: International Telecommunication Union; 2018. URL: http://unesdoc.unesco.org/images/0024/002459/245905e.pdf
22. China Internet Network Information Center. 2018. The 42nd China Statistical Report on Internet Development URL: https://cnnic.com.cn/IDR/ReportDownloads/201807/P020180711391069195909.pdf [accessed 2019-08-26]
23. Citrix Systems. 2015. Mobile Analytics Report URL: https://www.citrix.com/content/dam/citrix/en_us/documents/products-solutions/citrix-mobile-analytics-report-february-2015.pdf?accessmode=direct [accessed 2017-12-21]
24. Chen P, Li XC, Chen DL. [Research on the influences of sports app on undergraduates exercise behaviors-take Southeast University as an example]. Bull Sport Sci Tech 2016;24(7):109-111. [doi: 10.19379/j.cnki.issn.1005-0256.2016.07.044]
25. Litman L, Rosen Z, Spierer D, Weinberger-Litman SL, Goldschein A, Robinson J. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. J Med Internet Res 2015 Aug 14;17(8):e195 [FREE Full text] [doi: 10.2196/jmir.4142] [Medline: 26276227]
26. Harries T, Eslambolchilar P, Rettie R, Stride C, Walton S, van Woerden HC. Effectiveness of a smartphone app in increasing physical activity amongst male adults: a randomised controlled trial. BMC Public Health 2016 Sep 2;16:925 [FREE Full text] [doi: 10.1186/s12889-016-3593-9] [Medline: 27590255]
27. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for improving fitness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
28. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: perceived effectiveness by app users. JMIR Mhealth Uhealth 2016 Apr 7;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
29. Shumaker S, Brownell A. Toward a theory of social support: closing conceptual gaps. J Soc Issues 1984;40(4):11-36. [doi: 10.1111/j.1540-4560.1984.tb01105.x]
30. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Dec;84(2):191-215 [FREE Full text] [doi: 10.1037/0033-295X.84.2.191]
31. Laird Y, Fawkner S, Kelly P, McNamee L, Niven A. The role of social support on physical activity behaviour in adolescent girls: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2016 Jul 7;13:79 [FREE Full text] [doi: 10.1186/s12966-016-0405-7] [Medline: 27387328]
32. Anderson-Bill ES, Winett RA, Wojcik JR. Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation. J Med Internet Res 2011 Mar 17;13(1):e28 [FREE Full text] [doi: 10.2196/jmir.1551] [Medline: 21441100]
33. Qu NN, Li KJ. [Study on the reliability and validity of international physical activity questionnaire (Chinese vision, IPAQ)]. Zhonghua Liu Xing Bing Xue Za Zhi 2004 Mar;25(3):265-268. [doi: 10.3760/j.issn:0254-6450.2004.03.021] [Medline: 15200945]
34. Chogahara M. A multidimensional scale for assessing positive and negative social influences on physical activity in older adults. J Gerontol B Psychol Sci Soc Sci 1999 Nov;54(6):S356-S367. [doi: 10.1093/geronb/54b.6.s356] [Medline: 10625971]
35. Cavallo DN, Brown JD, Tate DF, DeVellis RF, Zimmer C, Ammerman AS. The role of companionship, esteem, and informational support in explaining physical activity among young women in an online social network intervention. J Behav Med 2014 Oct;37(5):955-966 [FREE Full text] [doi: 10.1007/s10865-013-9534-5] [Medline: 24081454]
36. Wang C, Gang HE, Jianzhong LI, Zheng Y. [Physical activity levels and exercise self-efficacy in adolescents with disability: a preliminary study using accelerometry]. J Capital Univ Phys Educ Sports 2016;28(4):380-384. [doi: 10.14036/j.cnki.cn11-4513.2016.04.020]
37. Pirasteh A, Hidarnia A, Asghari A, Faghihzadeh S, Ghofranipour F. Development and validation of psychosocial determinants measures of physical activity among Iranian adolescent girls. BMC Public Health 2008 May 7;8:150 [FREE Full text] [doi: 10.1186/1471-2458-8-150] [Medline: 18462488]
38. Guo W, Wang W, Yao JJ. [Survey on the usage of sport app among college students in Jinan]. Bull Sport Sci Tech 2018;3(2):123-4, 132. [doi: 10.19379/j.cnki.issn.1005-0256.2018.02.051]
39. Schoenborn CA, Adams PF, Peregoy JA. Health behaviors of adults: United States, 2008-2010. Vital Health Stat 10 2013 May(257):1-184 [FREE Full text] [Medline: 25116426]
40. Liao SB. [Investigation and analysis of after-school sports activities for university students in Beijing]. Shandong Sports Sci Tech 2012;7(1):85-87. [doi: 10.3969/j.issn.1009-9840.2012.01.023]

41. van der Horst K, Paw MJ, Twisk JW, van Mechelen W. A brief review on correlates of physical activity and sedentariness in youth. Med Sci Sports Exerc 2007 Aug;39(8):1241-1250. [doi: 10.1249/mss.0b013e318059bf35] [Medline: 17762356]
42. Yang S. [Relationship between social support, self-efficiency and satisfaction of youth physical activity]. J Wuhan Inst Phys Ed 2016;50(2):90-94. [doi: 10.3969/j.issn.1000-520X.2016.02.015]

Abbreviations

BMI: body mass index
CFI: comparative fit index
IPAQ: International Physical Activity Questionnaire
KMO: Kaiser-Meyer-Olkin
MPA/MIPA: moderate-intensity physical activity
PA: physical activity
SRMR: standardized root mean square residual
VPA/VIPA: vigorous-intensity physical activity
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.10.18; peer-reviewed by A Zeleke, K Corder; comments to author 21.03.19; revised version received 29.06.19; accepted 28.07.19; published 25.09.19.

Please cite as:
Wang T, Ren M, Shen Y, Zhu X, Zhang X, Gao M, Chen X, Zhao A, Shi Y, Chai W, Liu X, Sun X
The Association Among Social Support, Self-Efficacy, Use of Mobile Apps, and Physical Activity: Structural Equation Models With Mediating Effects
JMIR Mhealth Uhealth 2019;7(9):e12606
URL: http://mhealth.jmir.org/2019/9/e12606/
doi:10.2196/12606
PMID:31573936

Original Paper

User Experience of 7 Mobile Electroencephalography Devices: Comparative Study

Thea Radüntz[1], Dr rer nat; Beate Meffert[2], Dr-Ing

[1]Mental Health and Cognitive Capacity, Federal Institute for Occupational Safety and Health, Berlin, Germany

[2]Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany

Corresponding Author:
Thea Radüntz, Dr rer nat
Mental Health and Cognitive Capacity
Federal Institute for Occupational Safety and Health
Nöldnerstr 40-42
Berlin, 10317
Germany
Phone: 49 30 51548 4418
Email: raduentz.thea@baua.bund.de

Abstract

Background: Registration of brain activity has become increasingly popular and offers a way to identify the mental state of the user, prevent inappropriate workload, and control other devices by means of brain-computer interfaces. However, electroencephalography (EEG) is often related to user acceptance issues regarding the measuring technique. Meanwhile, emerging mobile EEG technology offers the possibility of gel-free signal acquisition and wireless signal transmission. Nonetheless, user experience research about the new devices is lacking.

Objective: This study aimed to evaluate user experience aspects of emerging mobile EEG devices and, in particular, to investigate wearing comfort and issues related to emotional design.

Methods: We considered 7 mobile EEG devices and compared them for their wearing comfort, type of electrodes, visual appearance, and subjects' preference for daily use. A total of 24 subjects participated in our study and tested every device independently of the others. The devices were selected in a randomized order and worn on consecutive day sessions of 60-min duration. At the end of each session, subjects rated the devices by means of questionnaires.

Results: Results indicated a highly significant change in maximal possible wearing duration among the EEG devices (χ^2_6=40.2, n=24; *P*<.001). Regarding the visual perception of devices' headset design, results indicated a significant change in the subjects' ratings (χ^2_6=78.7, n=24; *P*<.001). Results of the subjects' ratings regarding the practicability of the devices indicated highly significant differences among the EEG devices (χ^2_6=83.2, n=24; *P*<.001). Ranking order and posthoc tests offered more insight and indicated that pin electrodes had the lowest wearing comfort, in particular, when coupled with a rigid, heavy headset. Finally, multiple linear regression for each device separately revealed that users were not willing to accept less comfort for a more attractive headset design.

Conclusions: The study offers a differentiated look at emerging mobile and gel-free EEG technology and the relation between user experience aspects and device preference. Our research could be seen as a precondition for the development of usable applications with wearables and contributes to consumer health informatics and health-enabling technologies. Furthermore, our results provided guidance for the technological development direction of new EEG devices related to the aspects of emotional design.

(JMIR Mhealth Uhealth 2019;7(9):e14474) doi:10.2196/14474

KEYWORDS

wearable devices; user experience; electroencephalography; mobile applications; electrodes; dry electrodes

Introduction

User Experience Research of Emerging Electroencephalography Technology

In the previous years, registration of brain activity has become more and more popular not only in science but also in the home and gaming sector. Users look forward to identifying and quantifying their mental state directly there where human information processing takes place, and electroencephalography (EEG) offers a way to assess the levels of fatigue, stress, or emotions. The state feedback can then be used to prevent undesired situations, enhance wanted effects, or control devices. The increasing number of publications related to brain-computer interfaces [1-7] indicates an ever-growing interest in communication systems where encoded brain activity from the user is used as an alternative channel to send information to a computer. In addition, progress in sensor technology enables the production of low-cost, light-weighted, and marketable devices. However, extended use of the EEG is hampered by user experience challenges and user acceptance issues regarding the measuring technique.

Only a few years ago, one of the main issues was the limited mobility of the subjects because of the wired connections going from the electrode cap to an amplifier and computer. Meanwhile, wireless signal transmission helps to overcome this problem and allows subjects to move more freely. Further concerns are related to the application of gel electrodes and skin preparation for reducing the impedance. Emerging sensor technology uses gel-free sensors to enable a quick and easy application of the electrodes by the users themselves. For assuring an acceptable signal quality, impedance between electrodes and skin must be low, that is, electrodes need a good and permanent contact to the skin. This becomes particularly difficult to achieve for dry electrodes that work without the conductive gel. Given this, the question of wearing comfort and user experience becomes even more evident.

Finally, there are also user experience issues related to the unflattering visual appearance of the traditional EEG caps and thus linked to the research field of emotional design [8]. The core idea thereby is that products' design strives to elicit positive emotions and thus influence users' perception to provide a greater level of user experience. The 3-level model of emotional design includes the visceral, the behavioral, and the reflective level [8,9]. The visceral is the most basic, immediate level and addresses our first reactions to visual or sensory aspects (eg, aesthetics and quality) of the product. The behavioral level refers to usability aspects of the product, whereas the reflective level comprises conscious cognition. More general, the reflective level asks how well the product fits in with user's current self-image and addresses not only mental and emotional but also social aspects.

To recap, there is growing interest among users in brain state monitoring and increased efforts by developers for developing mobile EEG devices. However, serious user experience research in this field is rare, and it remains still unclear whether user acceptance of the new devices is improved compared with traditional EEG technology. In our study, we aimed to address this issue and advance the state of the art regarding user experience of emerging EEG devices. Thereby, we focused on the wearing comfort of the devices and aspects of emotional design, particularly the behavioral and reflective levels.

Related Work

During the previous years, the advances in sensor technology promoted the research regarding the usability of emerging EEG devices. Most of the published papers concentrated only on device functionality and signal quality comparison between the traditional gel-based electrodes and the new dry electrodes [7,10-12].

Only a small number of studies were concerned with devices' wearing comfort and design requirements. Nikulin et al [13] reported that for designing a new kind of electrodes, they considered not only signal quality but also electrodes' visual appearance and wearing comfort. They put effort to create extremely light and small electrodes that could be applied with some conductive gel directly on the head without any cap or headset. During the study, subjects reported that the electrodes were not noticeable and also not visually detectable by other people. Subjects felt less watched and thus better. Nikulin et al argued that this was particularly important when working outside the laboratory, and subjects were asked to behave naturally and free, in particular, during field experiments in real work environments. However, the main limitation was that the electrodes had to be applied with gel. This application procedure was time consuming and required specific knowledge about electrodes' precise positions on the head. Hence, it had to be done by an experienced investigator and could not be done by the subject itself. A further limitation was that the subjects did not have the opportunity to compare the new electrode device with another.

Similarly, Grozea et al [14] reported on their work on new electrodes with fine, flexible, and metal-coated polymer bristles. The bristles should allow for a good contact through the hair, and simultaneously, they should be comfortable during wearing. The researchers tested the electrodes on subjects (ie, colleagues) that had previous experience with other kinds of electrodes (eg, gel-based and pin electrodes). The subjects concluded that although the bristles electrodes were better than the pin electrodes, the bristles could have been softer and more flexible to increase comfort. Limitations of the study were the small number of subjects participating and the lack of direct comparison among the different kinds of electrodes instead of recalling the wearing comfort from previous experiences.

Comparison studies among different commercial EEG devices regarding user experience were rare. A study by Ekandem et al [15] dealt with the comparison between Emotiv's EPOC device and NeuroSky's MindWave device. Research questions concerned the wearing comfort, the preparation, and the application time. The latter was less than 5 min for both devices and thus clearly less compared with traditional EEG devices. After 15 min of wearing, subjects were asked to answer questions about the overall comfort of the worn device, the length of time they would be able to wear it, and the type of discomfort [15]. Thereby, the EPOC device was rated more comfortable compared with the MindWave device. A main

limitation of the study concerned the wearing time of 15 min because this could be insufficient for determining discomfort issues.

A study by Izdebski et al [16] was divided into 2 similar experiments that tested in total 7 devices. Of 7 devices, 4 devices (g.tec's g.SAHARA, Emotiv's EPOC, ANT Neuro's asalab, and Brain Products' [Brain Products GmbH] actiCAP) were tested by 4 subjects, and the remaining 3 devices (BioSemi's ActiveTwo, Cognionics' Dry System, and Cognionics' Wet System) were tested by 9 subjects. Duration of the sessions varied between 1 and 3 hours, and the usability was assessed at the end of each session by a questionnaire. Surveyed usability aspects were comfort, cap fit, mood, and movement restriction. Izdebski et al reported that the gel-based electrode headsets asalab and actiCAP induced general discomfort although participants did not report an unpleasant feeling under the cap nor a high pressure of the electrodes. Regarding cap fit, the ActiveTwo and systems without adjustment possibilities received negative ratings. The EPOC, g.SAHARA, and asalab devices yielded a more negative mood at the end of the session, whereas the wired systems asalab and actiCAP were rated as more movement restricting. A limitation of the study concerns the lack of a consistent within-subject design and the very different session durations.

Hairston et al [17] conducted a usability research experiment with a wearing time duration of 60 min. They compared 4 EEG devices: 3 wireless EEG systems (Emotiv's EPOC, Advanced Brain Monitoring's B-Alert X10, and QUASAR's HMS) and 1 wired, laboratory-grade device (Bio-Semi's ActiveTwo). The main user experience aspects they focused on, besides signal quality issues, were the adaptability of the devices to different head sizes, comfort, and subjects' device preference. They found that subjects preferred the B-Alert X10 device more than the other 2 wireless systems although it had gel-based electrodes. Subjects reported that the gel-infused pads of the B-Alert X10 device were more comfortable than the others. Finally, Hairston et al stated that future work was needed to systematically study usability factors and improve development efforts of new systems.

To compare the usability of a brain-computer interface for communication, Nijboer et al [18] tested 3 different EEG headsets (g.tec's g.SAHARA, Emotiv's EPOC, and BioSemi's ActiveTwo). Apart from signal quality, Nijboer et al also assessed the speed and ease of headset's setup, subjects' rating about their appearance with headset, comfort, and general device preference. Nijboer et al obtained the highest setup time for the gel-based ActiveTwo device, the best aesthetic ratings for the EPOC device, and the best comfort ratings for the gel-based ActiveTwo and pin-based g.SAHARA devices. Although the EPOC device yielded the worst ratings regarding comfort, it was the device of choice in the ranking of preference. Nijboer et al assumed that aesthetics and ease of use could be more important factors than comfort when it comes to preference ranking. They stated that more research was needed to understand which user experience aspects influence subjects' preference choice.

Table 1 summarizes the above-mentioned studies in a symmetric presentation style. To conclude, considering that duration of registration sessions and thus device wearing can take a long time, comfort requirements are particularly important. Existing studies regarding the usability of EEG headsets indicated that for assuring user acceptance, devices should be lightweight, comfortable, not painful to wear, and with an unobtrusive design. However, limitations of these studies were a limited number of participants, lack of comparisons among different devices, or a too short wearing duration of the EEG headsets. Most of the studies focused primarily on wearing comfort and neglected user experience aspects such as emotional design. In our study, we considered these things and systematically compared 7 different EEG devices.

Table 1. Literature review regarding user experience of emerging electroencephalography technology.

Reference	Devices tested	Electrode type and number	Set size	Wearing duration	User aspects and items	Results
Nikulin et al 2010 [13]	Proprietary development, traditional EEG[a] cap	Miniaturized C-electrodes with gel, 3; standard electrodes with gel, 3	4 subjects	40-60 min	Wearing comfort, tactile sensation, shame	No tactile sensations associated with C-electrode wearing, no negative emotional impact in the presence of others, and no discomfort
Grozea et al [14]	Proprietary development	Dry bristle electrodes; no information about number of electrodes	8 colleagues (2 of them excluded)	<1 hour	Comfort issues	Most subjects reported them to be more advanced than the previously known
Ekandem et al [15]	Emotiv's EPOC, NeuroSky's MindWave	Saline-based, 14; dry, 1	13 subjects (2 of them excluded)	15 min	Comfort and wearing duration	EPOC more comfortable; at least 20 min possible
Izdebski et al [16]	g.tec's g.SAHARA, Emotiv's EPOC, Cognionics' Dry System, ANT Neuro's asalab, Brain Products' actiCAP, BioSemi's ActiveTwo, and Cognionics' Wet System	Dry, 32; saline-based, 14; dry, 64; gel, 128; gel, 64; gel, 128; gel, 64	4 subjects (g.SAHARA, EPOC asalab, and actiCAP); 9 subjects (ActiveTwo, Cognionics' Dry System, and Cognionics' Wet System)	4 subjects (2-3 hours); 9 subjects (1-2 hours)	Comfort, cap fit, mood, and movement restriction	asalab and actiCAP induced general discomfort although participants did not report unpleasant feeling under cap nor high pressure of electrodes; ActiveTwo and systems without adjustment possibilities received negative ratings regarding cap fit; EPOC, g.SAHARA, and asalab yielded a more negative mood at the end of the session; the wired systems asalab and actiCAP were rated as more movement restricting
Hairston et al [17]	Emotiv's EPOC, Advanced Brain Monitoring's B-Alert X10, QUASAR's HMS, and BioSemi's ActiveTwo	Saline-based, 14; gel, 9; dry, 9; gel, 64	16 subjects (3-4 of them excluded)	60 min	Comfort, preference	Most preferred: B-Alert; comfortable to wear
Nijboer et al [18]	g.tec's g.SAHARA, Emotiv's EPOC, BioSemi's ActiveTwo	Dry, 8; saline-based, 14; gel, 32	13 subjects	~1 hour	Speed and ease of setup, appearance with headset, comfort, and general preference	Highest setup time for ActiveTwo; best aesthetic ratings for EPOC; best comfort ratings for ActiveTwo and g.SAHARA; in general, most preferred: EPOC

[a]EEG: electroencephalography.

Research Objectives

As the registration of brain activity outside the laboratory becomes more popular, aspects of user experience attract more attention when new devices are to be developed. Apart from improving wearing comfort that is crucial regarding user experience, developers also put more emphasis on the headset design of the EEG devices. This can lead to extraordinary designs that are not always flattering and easy to use for the user. In such cases, the visual appearance and behavior of the device can influence the well-being of a person [13].

Our first research objective was concerned with the test of the devices. First, we referred to the well-known issue of wearing comfort linked to the different electrode types and the question of how comfortable the different electrodes were after a longer wearing time. We assumed that maximal possible wearing duration would vary significantly among the devices depending on the type of electrode. Spring-loaded or rigid pin electrodes were expected to apply more pressure on the head and thus to have a smaller comfort and a low possible wearing duration. Gel-based electrodes were expected to assure a better comfort and could be worn for longer. Furthermore, we were interested in testing the devices in regard to the visceral and behavioral levels of emotional design. These comprised the design of the devices and the ease of use. To this end, we formulated the following research questions for the evaluation of the devices:

Research question 1a: Does maximal possible wearing duration differ among devices with different electrode types?

Research question 1b: Does the visual perception of devices' design differ among each other?

Research question 1c: Does practicability of the devices differ among each other?

Especially in cases where the EEG device is worn in public (eg, workplace), some users could prefer a more unobtrusive design. This can be linked to the reflective level of Norman's 3-level model of emotional design [8]. Thereby, information from the visceral and behavioral levels are combined with our knowledge and experiences, filtered, and cognitively processed. At this level, user's self-image plays a crucial role. Beyond the intended use of the product, user preferences are based on who will see it and how these viewers will judge the user with it.

Hence, we were interested to find out if users were willing to accept less comfort for a more attractive headset design. On the basis of this consideration, we formulated our second research objective:

Research question 2: Does visual appearance affect the overall rating of the devices more than wearing comfort?

In the Methods section of our study, we introduce the EEG devices, material used, sample set, and procedure for conducting the experiments. The gained results are presented in the Results section and discussed in the following section. Thereby, we mention potential limitations to the study. Finally, the Conclusions subsection aims to highlight the main points of our study and draw general conclusions from the investigation.

Methods

Electroencephalography Systems

The investigation focused on 7 currently available mobile EEG devices. Table 2 shows the devices and summarizes their characteristics that are briefly described in the following.

NeuroSky's MindCap device is a 1-channel EEG system. It comes with a frontal electrode and an ear clip reference electrode. The use of conductive gel is not necessary, and the signal is transmitted wirelessly through Bluetooth interface. The weight is 119 g. The device is recommended for neurofeedback training and gaming.

Emotiv's EPOC device comes with 14 saline-based wet felt sensors. These are mounted on quite flexible plastic branches. The signal is transmitted wirelessly through Bluetooth interface. The EPOC device has a weight of 116 g.

Mindo's 4S Jellyfish device is a wireless dry electrode EEG device. The 4 electrodes that are mounted on a headband can be applied at either frontal or parietal sites. In our case of frontal EEG, foam-based electrodes (Figure 1, left) are recommended. In case of parietal EEG, spring-loaded pin electrodes (Figure 1, right) are to be applied. The reference is an adhesive electrode at the mastoid. The device weighs 95 g.

Mindo's 32 Trilobite device comprises 32 EEG channels. The frontal 3 of them are foam-based electrodes (Figure 1, left). The remaining 29 are spring-loaded pin electrodes (Figure 1, right). Furthermore, the device includes a ground and a reference electrode, both applied with a clip on the ear lobes. Signal transmission occurs wirelessly through Bluetooth. Its weight is 524 g.

BRI's BR8+ device has got 8 dry electrodes. The frontal 2 of them are foam-based electrodes (Figure 1, left). The remaining 6 are spring-loaded pin electrodes (Figure 1, right). The device includes ground and reference ear clip electrodes and a wireless signal transmission through Bluetooth. The earpads of the device do not have any technical functionality. They are thought to reduce the headset pressure and help positioning the headset at the center of the head. The BR8+ weighs 269 g.

g.tec's g.SAHARA/g.Nautilus device comprises 16 pin electrodes (Figure 2) that are mounted on a traditional EEG cap. The cap size can vary among small, medium, and large. However, to reduce financial costs, we used only the medium-sized cap. Adhesive ground and reference electrodes are applied at the mastoids. The signal is transmitted wirelessly by means of g.Nautilus device that is attached at the back of the EEG cap. It has a weight of 233 g.

g.tec's g.LADYbird/g.Nautilus device is a traditional gel-based EEG system with 16 active electrodes. An ear clip electrode serves as reference. Similar to the g.SAHARA/g.Nautilus device, the cap size can vary. However, in our study, we used only the medium-sized cap. The g.Nautilus device at the back of the cap allows for wireless signal transmission. The total weight of the EEG headset amounts to 165 g. Unlike the other devices, the g.LADYbird/g.Nautilus device is not designed for home and biofeedback applications. It is primarily developed for research and medical use and the treatment of locked-in patients. We included it to our study as state-of-the-art reference for EEG regarding user experience issues.

Finally, all manufacturers of our EEG devices promote their EEG systems as highly comfortable and easy to use.

Table 2. Electroencephalography (EEG) devices used.

EEG device	Headset	Electrode type	Number of electrodes	Weight
MindCap (NeuroSky Inc, San Jose, CA, USA)		Dry	1	119 g
EPOC (Emotiv Inc, San Francisco, CA, USA)		Saline-based	14	116 g
Jellyfish (Mindo, Hsinchu, Taiwan)		Foam-based	4	95 g
Trilobite (Mindo, Hsinchu, Taiwan)		3 foam-based, 29 spring-loaded pins	32	524 g
BR8+ (BRI Inc, Hsinchu, Taiwan)		2 foam-based, 6 spring-loaded pins	8	269 g
g.SAHARA (g.tec GmbH, Graz, Austria)		Pin electrodes	16	233 g
g.LADYbird (g.tec GmbH, Graz, Austria)		Gel-based	16	165 g

Figure 1. Foam-based frontal electrodes (left) and spring-loaded pin electrodes (right).

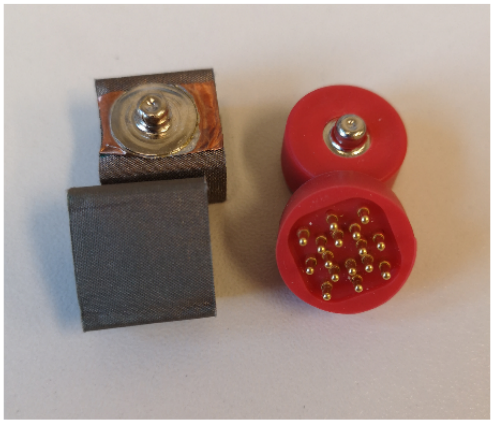


Figure 2. Pin electrodes of g.tec's g.SAHARA device.

Procedure and Subjects

Our study took place in a typical office setting. The 24 subjects participating (Table 3) completed over the course of 9 consecutive workdays a total of 9 sessions. The first session was aimed at familiarizing the subjects with the computer tasks and games they had to perform while wearing the EEG devices. In this session, we also assessed subjects' attitude toward technology by means of the 19 items of the TA-EG questionnaire (TA-EG: translated from the original German title: "Fragebogen zur Technikaffinität - Einstellung zu und Umgang mit elektronischen Geräten") [19-22]. The items are answered on a 5-point Likert scale (1=fully disagree and 5=fully agree) and address 4 dimensions: technology enthusiasm, competence in handling technology, positive attitude, and negative attitudes toward electronic devices. Subjects with calculated values below the median were assigned to the group of negative attitudes, whereas subjects with values over the median were assigned to the group of positive attitudes toward technology.

Table 3. Sample set used for analysis.

Age (years)	Male, n (%)	Female, n (%)	Total, N
26-34	2 (20)	8 (80)	10 (100)
35-49	3 (50)	3 (50)	6(100)
50-66	8 (100)	0 (0)	8(100)
Total	13	11	24

In the following 7 days, 1 device per day was selected in random order and tested independently of the others. Thereby, the subjects wore the device for 60 min and performed the same sequence of tasks and 1-min rest measurements with eyes closed and eyes opened. The devices were applied by an expert. At the end of each session, they were asked how long they would be able to wear the EEG headset. They indicated their answers on a 5-min steps scale between 0 and 120 min. They also answered questions regarding the device's design. Next, the subjects applied the device on their own. The expert inspected the signal quality of the EEG and gave instructions for improving it when needed. Moreover, 1-min rest measurements with eyes closed and eyes opened were performed, and thereafter, subjects rated the practicability of the device (Table 4). An exception was made for the g.LADYbird device that could not be taken off, reapplied, and properly used because of the smeared gel that builds conductive bridges. For the g.LADYbird device, we solely skipped the rest measurements.

During the last session, all EEG devices were rated. First, paired comparisons were conducted between every 21 pairs of 2 devices presented. Participants were asked to select the headset that they were willing to wear over a longer period of time or even daily. To avoid reliance on memory, subjects were instructed to reapply each of the 2 presented headsets and decide consciously. A mirror in front of them allowed them to include the visual appearance of the headset in their preference rating. Furthermore, we paid attention to the presentation order of the pairs and proceeded as recommended by Ross [23].

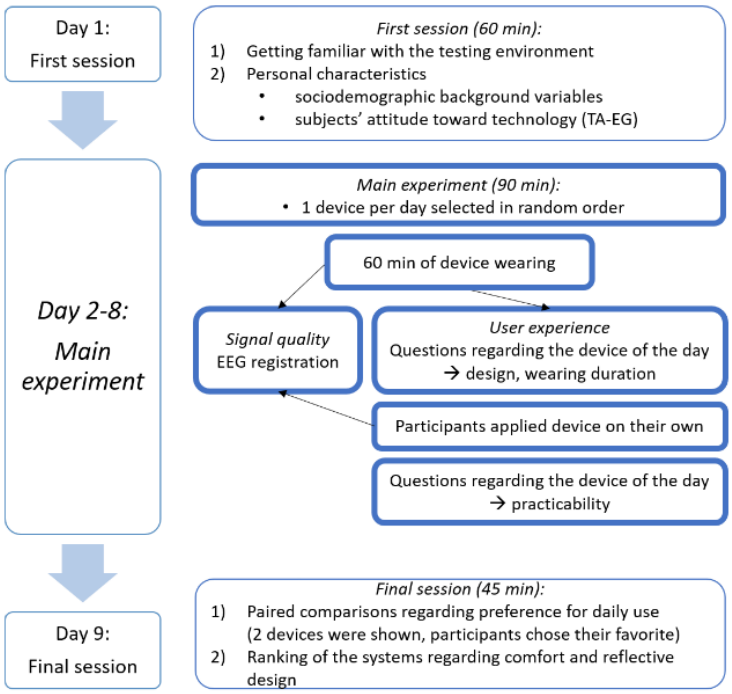
Finally, subjects completed a questionnaire where they had to rank the devices regarding wearing comfort and visual appearance separately (Table 4). Thereby, the item for visual appearance aimed to also integrate aspects from the reflective level of emotional design. Each of the headsets was set on a rank order between 1 (the most appropriate) and 7 (the least appropriate). Figure 3 outlines the experimental design of the study. All procedures were carried out with the adequate understanding and written consent of the subjects. The investigations acquired were approved by the local review board of our institution.

Table 4. User experience acquisition.

Aspects of emotional design	Item	Possible answers	Conducted	Research question
Visceral level	The headset has an attractive design	1: does not apply at all and 5: applies fully	After each session	1b
Behavioral level	I could apply and use the EEG[a] headset without aid	1: does not apply at all and 5: applies fully	After each session	1c
Behavioral level	How long are you able to wear EEG headset? Please mark the maximal-possible time duration in minutes on the scale below	Scale from 0 to 120 with 5 min steps	After each session	1a
Behavioral level	Wearing the device was comfortable	Ranking of the devices: 1: most appropriate and 7: least appropriate	Final session	2
Reflective level	It would not be a problem for me to be seen by my colleagues wearing the device	Ranking of the devices: 1: most appropriate and 7: least appropriate	Final session	2

[a]EEG: electroencephalography.

Figure 3. Experimental design of the study. EEG: electroencephalography.



Results

Comparisons Among Devices

The first research objective was concerned with the test of the devices regarding their wearing comfort after a longer period of time, visual appearance, and ease of use. For evaluation, we used subjects' answers conducted after each session (Table 4). Statistical analysis was conducted using nonparametric Friedman tests of differences among the repeated measures.

Maximal Possible Wearing Duration Differs Among Devices

Results indicated a highly significant change in maximal possible wearing duration among the EEG devices (χ^2_6=40.2, n=24; *P*<.001). Rankings are presented in Table 5. Dunn-Bonferroni posthoc tests were calculated for the examination of the differences among the devices (Table 5; see also Multimedia Appendix 1 for the exact values). Significant differences were obtained between the Trilobite device and all other devices except the BR8+. The Trilobite device was ranked lower regarding maximal wearing duration than the other devices.

Perception of Headset Design Differs Among Devices

Regarding the visual perception of devices' headset design, results indicated a significant change in subjects' ratings (χ^2_6=78.7, n=24; *P*<.001). Rankings are presented in Table 6. Dunn-Bonferroni posthoc tests were calculated for the examination of the differences among the devices (Table 6; Multimedia Appendix 2).

Table 5. Maximal possible wearing duration (min) for each device over all subjects.

EEG device	Mean (SD)	Median (min, max)
MindCap	92.29 (35.87)	112.5 (5, 120)
Jellyfish	86.66(31.78)	90.0 (30, 120)
BR8+	73.54 (30.16)	60.0 (30, 120)
EPOC	101.87 (25.10)	117.5 (30, 120)
g.Sahara	81.04 (33.45)	80.0 (10, 120)
Trilobite	48.75 (28.59)	50.0 (5, 120)
g.Ladybird	100.41 (23.99)	112.5 (45, 120)

Table 6. "The headset has an attractive design." (1: does not apply at all and 5: applies fully). Statistics calculated over all, male, and female subjects for each device.

EEG device	All		Male		Female	
	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)
MindCap	3.71 (0.95)	4.0 (1, 5)	4.15 (0.68)	4.0 (3, 5)	3.18 (0.98)	3.0 (1, 4)
Jellyfish	3.58 (0.97)	4.0 (2, 5)	3.85 (0.80)	4.0 (2, 5)	3.27 (1.10)	4.0 (2, 5)
BR8+	3.58 (0.97)	4.0 (1, 5)	3.92 (0.64)	4.0 (3, 5)	3.18 (1.16)	3.0 (1, 5)
EPOC	4.08 (0.77)	4.0 (2, 5)	4.23 (0.92)	4.0 (2, 5)	3.91 (0.53)	4.0 (3, 5)
g.Sahara	2.21 (1.10)	2.0 (1, 5)	2.62 (1.12)	3.0 (1, 5)	1.73 (0.90)	2.0 (1, 4)
Trilobite	2.58 (0.92)	2.5 (1, 5)	2.92 (0.86)	3.0 (2, 5)	2.18 (0.87)	2.0 (1, 4)
g.Ladybird	2.08 (0.83)	2.0 (1, 4)	2.46 (0.66)	2.0 (2, 4)	1.64 (0.80)	1.0 (1, 3)

Significant differences were obtained between the g.LADYbird device and all other devices except g.SAHARA and Trilobite. The g.SAHARA device showed significant differences to all devices except Trilobite and g.LADYbird. The Trilobite device showed significant differences to the EPOC, MindCap, and Jellyfish devices. At this point, we also looked at possible gender effects relating to the perception of headsets' design. We evaluated the ratings separately for male and female participants (Table 6) and found highly significant differences among devices for both groups (male: χ^2_6=41.9, n=13, P<.001; female: χ^2_6=38.3, n=11, P<.001). Dunn-Bonferroni posthoc tests for male participants' ratings indicated significant differences between the Trilobite and EPOC devices as well as between g.SAHARA and MindCap and g.SAHARA and EPOC (Table 6; Multimedia Appendix 3). Furthermore, there were significant differences between the g.LADYbird device and all other devices except g.SAHARA and Trilobite. Dunn-Bonferroni posthoc tests for female participants' ratings indicated significant differences between the Trilobite and EPOC devices, g.SAHARA and EPOC as well as between g.LADYbird and EPOC and g.LADYbird and Jellyfish (Table 6; Multimedia Appendix 4).

Practicability Differs Among Devices

Results of subjects' ratings regarding the practicability of the devices indicated highly significant differences among the EEG devices (χ^2_6=83.2, n=24; P<.001). Rankings are presented in Table 7. Dunn-Bonferroni posthoc tests were calculated for the examination of the differences among the devices (Table 7; Multimedia Appendix 5).

Significant differences were obtained between the g.LADYbird device and all remaining devices. To evaluate possible differences among subjects related to their attitude toward technology, we used the results from the TA-EG questionnaire and clustered our subjects in 2 groups. Subjects with a value below the overall median of 69.5 (range between 41 and 81) were assigned to the group with a negative attitude toward technology (mean age of cluster: 41 years, 5 females, and 7 males) and subjects with a value over the median to the group with a positive attitude (mean age of cluster: 44 years, 6 females, and 6 males). We evaluated the practicability ratings separately and found highly significant differences among devices for both groups (negative attitude: χ^2_6=48.5, n=12, P<.001; positive attitude: χ^2_6=40.6, n=12, P<.001).

Dunn-Bonferroni posthoc tests for the ratings of subjects with a negative attitude toward technology indicated significant differences between the g.LADYbird and all remaining devices (Table 7; Multimedia Appendix 6). Dunn-Bonferroni posthoc tests for the ratings of subjects with a positive attitude toward technology indicated significant differences between the g.LADYbird and all other devices except the Trilobite and g.SAHARA (Table 7; Multimedia Appendix 7).

The critical reader could argue that for evaluating the practicability, the signal quality of the device had to be taken into account after self-fitting the device. For the sake of completeness, we compared the signal quality of the rest measurements from self-fitting versus expert fitting of the system. The evaluation of the electroencephalogram was done in the time domain manually. A medical technical assistant with specialization in EEG and years of experience visually inspected the electroencephalograms and manually marked artifact segments. We computed the percentage of denoted artifacts compared with the entire recording time for each channel. We calculated the means over the channels for each subject and device. For comparison between the signal qualities from self-fitting versus expert fitting, we conducted a Wilcoxon paired difference test for each EEG system. The results are presented in Table 8. Rest measurements with closed eyes did not show significant differences between the fittings for none of the devices. Rest measurements with eyes opened indicated significant differences between the fittings for the BR8+ and the g.SAHARA devices (BR8+: z=−3.886, P<.001, r=0.56; g.SAHARA: z=4:086, P<.001, r=0.59).

For readers more interested in the signal quality evaluation of the devices, we would like to draw their attention on our paper on that topic [24].

Table 7. "I could apply and use the EEG headset without aid." (1: does not apply at all and 5: applies fully). Statistics calculated over all subjects, subjects with positive attitude, and subjects with negative attitude toward technology for each device. EEG: electroencephalography.

EEG device	All		Positive attitude		Negative attitude	
	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)
MindCap	4.63 (0.64)	5.0 (3, 5)	4.42 (0.79)	5.0 (3, 5)	4.42 (0.79)	5.0 (3, 5)
Jellyfish	4.67 (0.56)	5.0 (3, 5)	4.50 (0.67)	5.0 (3, 5)	4.50 (0.67)	5.0 (3, 5)
BR8+	4.21 (1.10)	5.0 (2, 5)	3.83 (1.19)	4.0 (2, 5)	3.83 (1.19)	4.0 (2, 5)
EPOC	4.54 (0.58)	5.0 (3, 5)	4.50 (0.52)	4.5 (4, 5)	4.50 (0.52)	4.5 (4, 5)
g.Sahara	4.04 (1.04)	4.0 (2, 5)	3.58 (1.24)	4.0 (2, 5)	3.58 (1.24)	4.0 (2, 5)
Trilobite	3.54 (1.31)	4.0 (1, 5)	3.00 (1.27)	3.0 (1, 5)	3.00 (1.27)	3.0 (1, 5)
g.Ladybird	1.75 (0.89)	1.5 (1, 5)	1.75 (0.86)	1.5 (1, 3)	1.75 (0.86)	1.5 (1, 3)

Table 8. Artifact proportions (%) of rest measurements with eyes open and closed from self-fitting and expert fitting of the system averaged over channels and subjects and considered for each device separately.

EEG device	Eyes closed				Eyes open			
	Expert fitting		Self-fitting		Expert fitting		Self-fitting	
	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)	Mean (SD)	Median (min, max)
MindCap	15.37 (33.54)	0.0 (0.0, 99.9)	17.60 (35.92)	0.0 (0.0, 99.9)	16.75 (33.38)	0.0 (0.0, 99.9)	10.77 (23.68)	0.0 (0.0, 99.9)
Jellyfish	23.26 (27.56)	13.4 (.0, 99.7)	14.64 (18.16)	6.9 (0.0, 61.9)	24.15 (25.98)	14.7 (0.0, 87.3)	20.94 (22.88)	11.8 (0.0, 80.0)
BR8+	48.38 (21.55)	49.9 (3.7, 87.5)	59.51 (22.44)	63.23 (12.5, 99.9)	45.12 (17.36)	47.7 (14.3, 80.0)	75.62 (20.89)	78.4 (26.5, 100)
EPOC	23.25 (37.07)	3.6 (0.0, 99.9)	37.82 (45.16)	11.4 (0.0, 99.9)	22.18 (36.61)	5.0 (0.0, 99.9)	37.60 (42.69)	13.3 (0.0, 99.9)
g.Sahara	32.05 (11.47)	34.1 (5.7, 55.5)	32.54 (13.46)	33.0 (0.0, 65.1)	9.79 (12.74)	4.0 (0.0, 41.8)	21.33 (16.85)	18.5 (3.5, 74.5)
Trilobite	29.14 (26.33)	18.6 (3.1, 106.25)	22.10 (26.19)	14.4 (0.0, 106.1)	33.83 (25.32)	23.9 (0.0, 91.4)	23.69 (20.56)	16.6 (0.0, 83.5)

Wearing Comfort and Visual Appearance

Our research question 2 asked if visual appearance affects the overall rating of the devices more than their wearing comfort. For the evaluation, we used multiple linear regression analysis. Ranking values of the items for visual appearance and wearing comfort (Table 4) served as independent variables. The criterion was the devices' ranking order regarding preference for daily use. This was calculated from the conducted paired comparisons.

For the sake of completeness, we have to mention that results from paired comparisons were not transitive for 6 subjects. In these cases, some devices have been selected with the same frequency, and thus, subjects' preference could not be mapped on an ordinal scale. Analysis of these subjects' decisions regarding the less rejected devices did not yield to a result, either. Hence, the 6 subjects with inconsistent answers were disclosed from further analysis.

We computed a multiple linear regression for each device separately. The results are presented in Table 9. Wearing comfort and visual appearance of the devices were able to statistically significant predict subjects' preference for daily use, except for the g.LADYbird device ($F_{2,15}$=0.752; P=.49). Wearing comfort had a large impact on device preference for almost all devices, whereas visual appearance was a poor predictor. An exception was the EPOC device. Hereby, visual appearance had a large impact on the preference, whereas wearing comfort had none. For the BR8+ device, both predictors were important. However, the wearing comfort was more influential.

At this point, we also looked at possible gender effects relating to the utilitarian versus hedonic aspects of the experience. For the male participants, wearing comfort and visual appearance were able to statistically significant predict subjects' preference for daily use, except for the g.LADYbird device ($F_{2,7}$=0.147; P=.87). Wearing comfort had a large impact on device preference for all devices except for the EPOC device where visual appearance was a better predictor. For the female participants, a significant regression equation with significant predictors was found for the Jellyfish ($F_{2,5}$=29.837; P=.002) and EPOC ($F_{2,5}$=25.571, P=.002) devices. For Jellyfish, wearing comfort significantly predicted subjects' preference, whereas for EPOC, visual appearance had a greater impact on subjects' preference ratings. Overall, it can be said that in cases where the regression models became significant, we were not able to identify opposing effects between female and male participants (Table 9).

Table 9. Results of multiple linear regression analysis for each device.

EEG[a] device and gender	R^2	Model		Wearing comfort		Visual appearance	
		F test (*df*)	*P* value	Coefficient	*P* value	Coefficient	*P* value
MindCap							
Both	0.938	112.518 (2,15)	<.001	1.112	<.001	−0.245	.21
Male	0.989	305.051 (2,7)	<.001	0.985	<.001	−0.024	.90
Female	0.707	6.018 (2,5)	.05	0.472	.16	0.067	.79
Jellyfish							
Both	0.825	35.319 (2,15)	<.001	0.797	<.001	0.069	.77
Male	0.764	11.357 (2,7)	.006	0.751	.01	0.210	.60
Female	0.923	29.837 (2,5)	.002	0.731	.008	0.235	.59
BR8+							
Both	0.846	41.182 (2,15)	<.001	0.701	<.001	0.327	.04
Male	0.952	70.150 (2,7)	<.001	0.802	.001	0.312	.07
Female	0.498	2.479 (2,5)	.18	0.540	.09	−0.010	.98
EPOC							
Both	0.849	42.080 (2,15)	<.001	0.149	.14	0.656	<.001
Male	0.823	16.286 (2,7)	.002	0.191	.53	0.627	.02
Female	0.911	25.571 (2,5)	.002	0.153	.14	0.655	.009
g.SAHARA							
Both	0.742	21.603 (2,15)	<.001	0.740	<.001	−0.040	.76
Male	0.939	54.275 (2,7)	<.001	0.777	<.001	0.011	.89
Female	0.633	4.312 (2,5)	.08	0.677	.03	0.000	>.99
Trilobite							
Both	0.737	21.026 (2,15)	<.001	0.943	<.001	0.139	.39
Male	0.770	11.706 (2,7)	.006	1.043	.002	0.109	.63
Female	0.485	2.354 (2,5)	.19	0.620	.12	0.260	.41
g.LADYbird							
Both	0.091	0.752 (2,15)	.49	0.018	.93	0.243	.25
Male	0.040	0.147 (2,7)	.87	0.063	.68	0.038	.75
Female	0.335	1.261 (2,5)	.36	−0.400	.42	2.300	.20

[a]EEG: electroencephalography.

Discussion

Comparisons Among Devices

In our first research objective, we were concerned to test the devices regarding 3 user experience aspects: wearing comfort, visual appearance, and ease of use.

Pin Electrodes Had the Lowest Wearing Comfort

Evaluation of the maximal possible wearing time as an indicator of devices' wearing comfort revealed the Trilobite device to be significantly less pleasant to wear than the remaining. The reason could be the uncomfortable pin electrodes. Overall means of maximal possible wearing duration indicated devices without pin electrodes such as the EPOC, MindCap, and g.LADYbird as the most favorable for a longer wearing time and with significant differences to the Trilobite. The finding that pin electrodes were less preferred was similar to findings by Grozea et al [14] but inconsistent to the results by Nijboer et al [18] and Izdebski et al [16]. However, Hairston et al [17] also emphasized the importance of the headset's ability to adjust to the different heads to assure comfort. In their work, they highlighted the need of flexible headsets to assure comfort during wearing. This aspect was also prominent in the work of Izdebski et al [16] who found that cap fit was rated as poor for headsets with rigid headsets. In our study, Trilobite's headset was the most rigid one. Furthermore, the Trilobite device was much heavier than the other devices. These 2 facts could have multiplied the impact of the pin electrodes on wearing comfort. The BR8+ device had pin electrodes, a rather rigid headset but less weight. Similar to the Trilobite, it yielded small values

regarding the maximal possible wearing duration. The g.SAHARA with pin electrodes but flexible headset and less weight had small wearing duration ratings, but these were higher than those of the Trilobite and BR8+ devices. We concluded that pin electrodes had the lowest wearing comfort, in particular when coupled with a rigid, heavy headset.

An Unobtrusive Design Coped Better With Individual Preferences

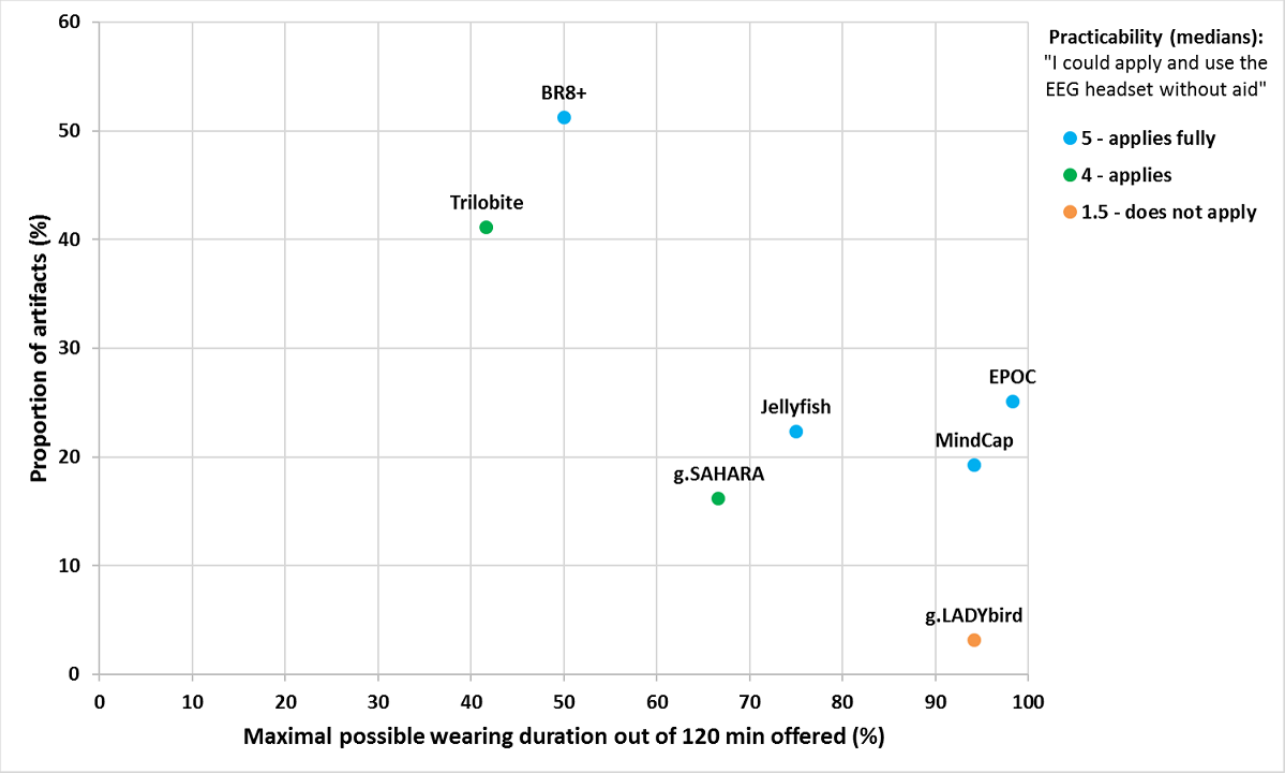
Headset design is not only responsible for the wearing comfort but also primarily responsible for device's visual appearance. Overall ratings of headset design indicated that the devices with a traditional EEG cap (ie, g.LADYbird and g.SAHARA) were significantly less preferred than all others, except the Trilobite device. The latter was also significantly less preferred than the MindCap, Jellyfish, and EPOC devices. Females' ratings indicated more variability than males' ratings leading to less significant differences among the devices. However, both genders perceived the design of g.LADYbird's and g.SAHARA's traditional caps and Trilobite's helmet as less attractive. Both groups primarily preferred the headsets of EPOC and Jellyfish with EPOC, indicating more significant differences to the other devices, in particular, by female subjects. This result was consistent with the results by Nijboer et al [18] where participants rated their appearance with the EPOC as best. Nijboer et al stated that reasons for the refusal of caps were that the whole head and part of the face were covered, and hair was flattened and invisible. In our study, the g.LADYbird, g.SAHARA, Trilobite, and MindCap devices covered subjects' whole head. However, ratings of the MindCap were significantly better compared with the other 3 devices. This was particularly true among the male subjects. We assumed that rating of the design was related to aspects of aesthetics, fashion style, and individual preference. These aspects might be strongly connected to the reflective level of emotional design. An unobtrusive headset design could have more potential to cope with different individual preferences because it is not eye-catching.

Practicability Was Closely Linked to Gel Electrodes and Attitude Toward Technology

Finally, we asked the subjects to rate the ease of use of the devices. Results indicated significant differences between the gel-based g.LADYbird and all remaining devices. This was reasonable, especially when considering that a second person was needed for applying the gel. Furthermore, subjects had to wash their hair after they took off the cap. We concluded that the effort for use was definitely high. The g.SAHARA and Trilobite devices were also rated as less easy to use. We supposed that this might be because of their larger number of electrodes but have to be aware that g.SAHARA had only 2 electrodes more than the EPOC device. Subjects with a negative attitude toward technology showed similar results regarding the practicability of the devices. However, subjects with a positive attitude toward technology did not indicate significant differences between the gel-based g.LADYbird and pin-based g.SAHARA neither between the g.LADYbird and Trilobite devices. Although these findings were surprising, we supposed that technical affine subjects were more critical during their ratings, and this could lead to more variability in their ratings. Taken the results of the signal quality comparison (Figure 4) into account, we noted similar tendencies between practicability ratings from subjects with a positive attitude toward technology and increased proportion of artifacts by self-fitting the devices. This was particularly true during the rest measurements with eyes opened, as subjects might have behaved more actively than with eyes closed. Thereby, the BR8+, g.SAHARA, and, to a lesser extent, the EPOC devices yielded more artifacts when compared with the fittings by an expert and revealed less practicability when rated by technical affine subjects. Nevertheless, the g.LADYbird device had the worst practicability ratings across subjects although a limitation of our study might be that we did not give the opportunity to the subjects to apply the device and the gel on their own. We believe that self-fitting of the gel-based electrodes would not have altered the ratings but must admit that future user experience research should consider this issue. Finally, we argue that subjects with a positive attitude toward technology were more accurate in their rating of device practicability.

In conclusion, although the practicability of the devices was closely linked to gel or dry electrodes, wearing comfort and design of the devices seemed to be more expressive. Thereby, we observed that devices that could be worn for a longer period of time did not always have an attractive design.

Figure 4. Relation between signal quality and comfort and trade-off against practicability. EEG: electroencephalography.



Wearing Comfort and Visual Appearance

In our second research question, we were interested to find out if wearing comfort was more important to the user than the visual appearance of the device. Thus, we asked subjects to rank all devices regarding both aspects separately. Furthermore, paired comparisons of the devices led us to a rank order regarding preference for daily use.

Results of a multiple linear regression analysis for each device indicated that, in general, wearing comfort was the better predictor for users' device preference. Exceptions were the EPOC and the g.LADYbird devices. Although for the g.LADYbird, none of the 2 aspects seemed to have any impact on device's preference ranking; the results for the EPOC device revealed an opposite tendency, that is, EPOC's visual appearance influenced subject's decision more than its wearing comfort. A reason for this could be that EPOC's wearing comfort was unobtrusive although its design was futuristic and professional. We assumed that this attracted the subjects and gave more weight to the visual appearance when it came to a preference for daily use. Interestingly, the design of the BR8+ was also one of the most modern and futuristic ones. The fact that BR8+'s visual appearance was a supplementary predictor to its comfort seemed to confirm our assumption.

Regarding the results of the g.LADYbird device, we had to speculate. The device was assumed to not cause any head pressure; hence, wearing comfort should be unobtrusive and a weak predictor for the preference for daily use. Its visual appearance was indeed not very attractive for daily wearing. However, this fact did not have a large influence on the preference either, similar to the g.SAHARA device that had the same cap. The main difference to all other devices was the application of gel and the necessity to wash the hair after each use of the device. Although comfortable to wear, the gel-based electrodes were undoubtedly inconvenient for daily use outside the laboratory. Hence, the ease of use could have affected the preference more than the examined factors.

Male and female participants did not show opposing results related to the predictors of daily use preference. Although almost all models (except g.LADYbird) became significant for the male participants, for the female participants, only 2 models reached the significance level (Jellyfish and EPOC). An explanation could be that females' ratings were not as consistent as males' ratings among each other. However, we have to be also aware of the small number of participants (8 females vs 10 males) that could have led to this result. To explore gender differences related to utilitarian versus hedonic aspects of experience, more research with larger subsets is needed. We have to draw attention to our sample's structure (Table 3) consisting of young female and older male participants. Disentangle the gender and age factors at these numbers seemed not possible. We assumed that regarding emotional design, the gender factor is more influential than the age, but the reader should note that the latter could have an effect, too. Further research should emphasize on this issue.

In general, the results of both genders emphasized that visual appearance was a better predictor only for the EPOC device. By taking into account the reflective level of emotional design, we add new insight about how the factors of comfort and visual appearance translate to user preference. Our results broaden the assumption by Nijboer et al [18] who postulated that the preference of EPOC was an evidence for the fact that aesthetics might be more important than comfort.

Conclusions

In our study, we investigated the user experience of mobile EEG devices. We compared 7 different EEG devices and offered a differentiated look at emerging mobile and gel-free EEG technology. The results yielded are summarized in Table 10. For the sake of convenience, we report only the artifact proportion differences between self-fitting and expert fitting from the eyes-closed measurement.

In addition, we gave insight into the relation between user experience aspects and device preference. The wearing comfort given by a device was the main factor for its daily use. The visual appearance of the device was certainly an important point. However, it only became influential when comfort was assured. Users were not willing to accept less comfort for a more attractive headset design. The reflective level of emotional design became important only if the behavioral level of the product was satisfactory.

To provide practical information to users of EEG devices, we combined the signal quality results from the study by Radüntz [24] with the current user experience results and concluded which system could be used under which condition. The EPOC device achieved the best results regarding user experience, but it suffered from a large proportion of artifacts. Although the EPOC device can be used in public because of its attractive design and the feeling of ease of use, potential users should be aware of the issues regarding signal quality, in particular, if the device is self-applied by a layman. Outstanding performances regarding maximal possible wearing duration and signal quality were obtained for the traditional gel-based but mobile g.LADYbird device. This device can be recommended for neuroscience research where precise and prolonged measurements are required without any deductions in comfort. However, devices wearing in public and self-application are not recommended. The MindCap device reviled good user experience results and satisfying signal quality. Users must consider that scientifically valid assertions could be hampered because of only 1 electrode available. The Jellyfish and g.SAHARA devices yielded similar results regarding comfort but differences regarding design (ie, better results for Jellyfish) and signal quality (ie, better results for g.SAHARA). We believe that g.SAHARA is a good solution for field experiments, where subjects are not exposed to the general public, and signal quality is important. Nevertheless, researchers should be aware of potential comfort issues that could arise in the course of time because of the pin electrodes. Potential applications for the Jellyfish device might be better suited for the gaming or biofeedback sector. The BR8+ and Trilobite devices did not meet our requirement for user experience, in particular, because of comfort issues. Furthermore, signal quality was lacking. Figure 4 illustrates the trade-offs between signal quality and user experience so that readers might be able to see if there are any devices of sufficient quality that might also be acceptable for daily use. The x-axis depicts devices' comfort rankings, calculated as a percentage of the maximal possible wearing duration in minutes out of 120 min offered. The y-axis represents the proportion of artifacts taken from the study by Radüntz [24].

Finally, we have to admit that there might be further factors that could have contributed to the preference decision. Our research could be seen as a precondition for the use of emerging EEG technology under realistic conditions in field experiments with longer duration. It paves the way for the development of usable applications with wearables and contributes to consumer health informatics and health-enabling technologies. Furthermore, our results provided guidance for the technological development direction of new EEG devices related to aspects of emotional design.

It has to be mentioned that the EEG equipment market shows rapid development. During this study, new devices appeared on the market that could not be tested, for example, the actiCAP Xpress Twist/LiveAmp device by Brain Products or the highly innovative approach using in-ear EEG technology [25,26]. However, our study design could easily be used in subsequent studies of new devices and benchmark the evaluation of further emerging EEG technology. Integration of test results from new devices into the findings already in existence would make it possible to compare the user experience of emerging EEG technology.

Table 10. User experience results of tested electroencephalography devices (medians over all subjects).

EEG[a] device	Comfort: maximal wearing duration (min)	Design (higher values indicate a more attractive design)	Practicability (higher values indicate greater practicability)	Artifact proportions (eyes closed: self-fitting-expert fitting [%]; higher values indicate more artifacts when self-fitted)
MindCap	113	4	5	2.2
Jellyfish	90	4	5	−8.6
BR8+	60	4	5	11.1
EPOC	118	4	5	14.6
g.SAHARA	80	2	4	0.5
Trilobite	50	2.5	4	−7
g.LADYbird	113	2	1.5	Not applicable

[a]EEG: electroencephalography.

Acknowledgments

The authors would like to thank Gerd Menzel and Ralph Blüthner for their technical and overall support, our student assistant Friederice Schröder for conducting the experiments, Marion Freyer for the visual inspection of the data and manual artifact marking, and the student assistants Yuexin Cao and Ilona Pritschke for computational support and graphic editing. Furthermore, the authors would like to thank Gabriele Freude and Uwe Rose for their general project support. Finally, they want to thank the reviewers of the manuscript for their insightful comments, which they deeply appreciate. The study was funded by the Federal Institute for Occupational Safety and Health (project number: F 2402).

Authors' Contributions

TR initiated the project and was responsible for the overall conception of the investigation. Data analysis was performed by TR. Data interpretation was performed by TR and BM. The manuscript was written by TR. Final critical editing was performed by BM.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Maximal possible wearing duration for each device over all subjects.
[PDF File (Adobe PDF File), 299KB - mhealth_v7i9e14474_app1.pdf]

Multimedia Appendix 2
Attractive design ratings for each device over all subjects.
[PDF File (Adobe PDF File), 434KB - mhealth_v7i9e14474_app2.pdf]

Multimedia Appendix 3
Attractive design ratings for each device over the male subjects.
[PDF File (Adobe PDF File), 436KB - mhealth_v7i9e14474_app3.pdf]

Multimedia Appendix 4
Attractive design ratings for each device over the female subjects.
[PDF File (Adobe PDF File), 433KB - mhealth_v7i9e14474_app4.pdf]

Multimedia Appendix 5
Practicability ratings for each device over all subjects.
[PDF File (Adobe PDF File), 436KB - mhealth_v7i9e14474_app5.pdf]

Multimedia Appendix 6
Practicability ratings for each device over subjects with negative attitude toward technology.
[PDF File (Adobe PDF File), 436KB - mhealth_v7i9e14474_app6.pdf]

Multimedia Appendix 7
Practicability ratings for each device over subjects with positive attitude toward technology.
[PDF File (Adobe PDF File), 436KB - mhealth_v7i9e14474_app7.pdf]

References

1. Rezeika A, Benda M, Stawicki P, Gembler F, Saboor A, Volosyak I. Brain-computer interface spellers: a review. Brain Sci 2018 Mar 30;8(4):pii: E57 [FREE Full text] [doi: 10.3390/brainsci8040057] [Medline: 29601538]
2. Ramadan RA, Vasilakos AV. Brain computer interface: control signals review. Neurocomputing 2017 Feb;223:26-44. [doi: 10.1016/j.neucom.2016.10.024]
3. Minguillon J, Lopez-Gordo MA, Pelayo F. Trends in EEG-BCI for daily-life: requirements for artifact removal. Biomed Signal Process Control 2017 Jan;31:407-418. [doi: 10.1016/j.bspc.2016.09.005]
4. Vaid S, Singh P, Kaur C. EEG Signal Analysis for BCI Interface: A Review. In: Proceedings of the Fifth International Conference on Advanced Computing & Communication Technologies. 2015 Presented at: ICACCS'15; February 21-22, 2015; Haryana, India p. 143-147. [doi: 10.1109/ACCT.2015.72]

5. Prashant P, Joshi A, Gandhi V. Brain Computer Interface: A Review. In: Proceedings of the 5th Nirma University International Conference on Engineering. 2016 Presented at: NUiCONE'16; November 26-28, 2016; Ahmedabad, India p. 1-6. [doi: 10.1109/NUICONE.2015.7449615]
6. Abdulkader SN, Atia A, Mostafa MS. Brain computer interfacing: applications and challenges. Egypt Inform J 2015 Jul;16(2):213-230. [doi: 10.1016/j.eij.2015.06.002]
7. Lee S, Shin Y, Woo S, Kim K, Lee HN. Review of wireless brain-computer interface systems. In: Fazel-Rezai R, editor. Brain-Computer Interface Systems: Recent Progress and Future Prospects. Rijeka, Croatia: IntechOpen; 2013.
8. Norman DA. Emotional Design: Why We Love (or Hate) Everyday Things. New York, US: Basic Books; 2004.
9. Ortony A, Norman DA, Revelle W. The role of affect and proto-affect in effective functioning. In: Fellous JM, Arbib MA, editors. Who Needs Emotions?: The Brain Meets the Robot. New York, US: Oxford University Press; 2005:173-202.
10. Ries AJ, Touryan J, Vettel J, McDowell K, Hairston WD. A comparison of electroencephalography signals acquired from conventional and mobile systems. J Neurosci Neuroeng 2014 Feb 1;3(1):10-20. [doi: 10.1166/jnsne.2014.1092]
11. Grummett TS, Leibbrandt RE, Lewis TW, DeLosAngeles D, Powers DM, Willoughby JO, et al. Measurement of neural signals from inexpensive, wireless and dry EEG systems. Physiol Meas 2015 Jul;36(7):1469-1484. [doi: 10.1088/0967-3334/36/7/1469] [Medline: 26020164]
12. Debener S, Minow F, Emkes R, Gandras K, de Vos M. How about taking a low-cost, small, and wireless EEG for a walk? Psychophysiology 2012 Nov;49(11):1617-1621. [doi: 10.1111/j.1469-8986.2012.01471.x] [Medline: 23013047]
13. Nikulin VV, Kegeles J, Curio G. Miniaturized electroencephalographic scalp electrode for optimal wearing comfort. Clin Neurophysiol 2010 Jul;121(7):1007-1014. [doi: 10.1016/j.clinph.2010.02.008] [Medline: 20227914]
14. Grozea C, Voinescu CD, Fazli S. Bristle-sensors--low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. J Neural Eng 2011 Apr;8(2):025008. [doi: 10.1088/1741-2560/8/2/025008] [Medline: 21436526]
15. Ekandem JI, Davis TA, Alvarez I, James MT, Gilbert JE. Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics 2012;55(5):592-598. [doi: 10.1080/00140139.2012.662527] [Medline: 22506831]
16. Izdebski K, Oliveira AS, Schlink BR, Legkov P, Kärcher S, Hairston WD, et al. Usability of EEG Systems: User Experience Study. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2016 Presented at: PETRA'16; June 29-July 1, 2016; Corfu, Island, Greece URL: https://dl.acm.org/citation.cfm?id=2910714 [doi: 10.1145/2910674.2910714]
17. Hairston WD, Whitaker KW, Ries AJ, Vettel JM, Bradford JC, Kerick SE, et al. Usability of four commercially-oriented EEG systems. J Neural Eng 2014 Aug;11(4):046018. [doi: 10.1088/1741-2560/11/4/046018] [Medline: 24980915]
18. Nijboer F, van de Laar B, Gerritsen S, Nijholt A, Poel M. Usability of three electroencephalogram headsets for brain–computer interfaces: a within subject comparison. Interact Comput 2015 Jul 15;27(5):500-511. [doi: 10.1093/iwc/iwv023]
19. Karrer K, Glaser C, Clemens C, Bruder C. Technikaffinität erfassen – der fragebogen TA-EG. In: Lichtenstein A, Stößel C, Clemens C, editors. Der Mensch im Mittelpunkt Technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme. Düsseldorf, Germany: VDI Verlag GmbH; 2009:196-201.
20. Schüssel F, Honold F, Weber M. Influencing factors on multimodal interaction during selection tasks. J Multimodal User In 2012 Nov 25;7(4):299-310. [doi: 10.1007/s12193-012-0117-5]
21. Mertens A, Wille M, Theis S, Rasche P, Finken L, Schlick CM. Attitudes of Elderly People Towards Assistive System: Influence of Amortization Barriers on the Adherence in Technically Assisted Rehabilitation and the Diffusion of Health Technologies. In: Proceedings of the Triennial Congress of the International Ergonomics Association. 2015 Presented at: IEA'15; August 9-14, 2015; Melbourne, Australia URL: https://tinyurl.com/y3rn6eqm
22. Nitsch V, Glassen T. Investigating the Effects of Robot Behavior and Attitude Towards Technology on Social Human-Robot Interactions. In: Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication. 2015 Presented at: RO-MAN'15; August 31-September 4, 2015; Kobe, Japan p. 535-540. [doi: 10.1109/ROMAN.2015.7333560]
23. Ross RT. Optimum orders for the presentation of pairs in the method of paired comparisons. J Educ Psychol 1934;25(5):375-382. [doi: 10.1037/h0070754]
24. Radüntz T. Signal quality evaluation of emerging EEG devices. Front Physiol 2018;9:98 [FREE Full text] [doi: 10.3389/fphys.2018.00098] [Medline: 29491841]
25. Goverdovsky V, von Rosenberg W, Nakamura T, Looney D, Sharp DJ, Papavassiliou C, et al. Hearables: multimodal physiological in-ear sensing. Sci Rep 2017 Jul 31;7(1):6948 [FREE Full text] [doi: 10.1038/s41598-017-06925-2] [Medline: 28761162]
26. Looney D, Kidmose P, Park C, Ungstrup M, Rank ML, Rosenkranz K, et al. The in-the-ear recording concept: user-centered and wearable brain monitoring. IEEE Pulse 2012;3(6):32-42. [doi: 10.1109/MPUL.2012.2216717] [Medline: 23247157]

Abbreviation

EEG: Electroencephalography

Edited by G Eysenbach; submitted 24.04.19; peer-reviewed by J Cañas, R Agrawal, A Mahnke, J Salisbury, T Steffert, J Cahill, T Gaber, T Mühlhausen; comments to author 19.06.19; revised version received 28.06.19; accepted 06.07.19; published 03.09.19.

Please cite as:
Radüntz T, Meffert B
User Experience of 7 Mobile Electroencephalography Devices: Comparative Study
JMIR Mhealth Uhealth 2019;7(9):e14474
URL: https://mhealth.jmir.org/2019/9/e14474/
doi:10.2196/14474
PMID:31482852

Original Paper

Reliability of a Smartphone Compared With an Inertial Sensor to Measure Shoulder Mobility: Cross-Sectional Study

Cristina Roldán-Jiménez[1,2*], PhD; Jaime Martin-Martin[1,3*], PhD; Antonio I Cuesta-Vargas[1,2,4*], PhD

[1]Clinimetric Group F-14 Biomedical Research Institute of Malaga, Malaga, Spain

[2]Department of Physiotherapy, Faculty of Health Sciences, University of Malaga, Malaga, Spain

[3]Legal Medicine Area, Department of Human Anatomy, Legal Medicine and History of Science, Faculty of Medicine, University of Malaga, Malaga, Spain

[4]Institute of Health & Biomedical Innovation, Faculty of Health, Queensland University Technology, Queensland, Australia

[*]all authors contributed equally

Corresponding Author:
Jaime Martin-Martin, PhD
Legal Medicine Area, Department of Human Anatomy, Legal Medicine and History of Science
Faculty of Medicine
University of Malaga
Bulevar Louis Pasteur, 32
Malaga, 29010
Spain
Phone: 34 951953282
Email: jaimemartinmartin@gmail.com

Abstract

Background: The shoulder is one of the joints with the greatest mobility within the human body and its evaluation is complex. An assessment can be conducted using questionnaires or functional tests, and goniometry can complement the information obtained in this assessment. However, there are now validated devices that can provide more information on the realization of movement, such as inertial sensors. The cost of these devices is usually high and they are not available to all clinicians, but there are also inertial sensors that are implemented in mobile phones which are cheaper and widely available. Results from the inertial sensors integrated into mobile devices can have the same reliability as those from dedicated sensors.

Objective: This study aimed to validate the use of the Nexus 4 smartphone as a measuring tool for the mobility of the humerus during shoulder movement compared with a dedicated InertiaCube3 (Intersense) sensor.

Methods: A total of 43 subjects, 27 affected by shoulder pathologies and 16 asymptomatic, participated in the study. Shoulder flexion, abduction, and scaption were measured using an InertiaCube3 and a Nexus 4 smartphone, which were attached to the participants to record the results simultaneously. The interclass correlation coefficient (ICC) was calculated based on the 3 movements performed.

Results: The smartphone reliably recorded the velocity values and simultaneously recorded them alongside the inertial sensor. The ICCs of the 3 gestures and for each of the axes of movement were analyzed with a 95% CI. In the abduction movement, the devices demonstrated excellent interclass reliability for the abduction humeral movement axis (Cronbach alpha=.98). The axis of abduction of the humeral showed excellent reliability for the movements of flexion (Cronbach alpha=.93) and scaption (Cronbach alpha=.98).

Conclusions: Compared with the InertiaCube3, the Nexus 4 smartphone is a reliable and valid tool for recording the velocity produced in the shoulder.

(JMIR Mhealth Uhealth 2019;7(9):e13640) doi:10.2196/13640

KEYWORDS

shoulder; kinematics; smartphone; mobile phone

Introduction

The shoulder is one of the joints with the widest range of pathological variations, with tendonitis, bursitis, frozen shoulder, or rotator cuff involvement being some of the most common ones [1]. These pathologies cause functional alterations in the structure, which influence certain specific evaluations. Questionnaires have been developed that assess the sensitive function, pain, neuromuscular alteration, movement of structures, functionality, and mobility [2]. However, the use of these questionnaires and their validation has produced conflict [3].

Of the different questionnaires used, there is no single tool that can evaluate all the clinical aspects involved. There are several questionnaires that evaluate the performance of tasks, such as Disabilities of the Arm, Shoulder, and Hand (DASH), Western Ontario Shoulder Instability Index, and QuickDash, of which DASH is the most widely used [2]. The DASH scale also has excellent psychometric properties, with a test-retest reliability of 0.94 [4]. However, although functional tests indicate whether or not the patient is able to perform an activity, in many cases, they do not assess the range of motion directly or do not evaluate the patient's dysfunction [5].

These questionnaires are often complemented with the use of goniometry on the assumption that this adds value to existing tests. The reliability of range-of-motion evaluations is determined by the measurement protocol used, as in the case of internal shoulder rotation [6,7]. In this respect, in addition to classical goniometry, there are systems based on digital goniometry that can be used to evaluate the range of motion with greater precision and to eliminate possible protocol deficits. However, they require specialized equipment, which is not available in many clinical situations [8]. These instruments are in many cases equipped with accelerometer-type inertial sensors [9].

Nowadays, because of the development of new technologies, the concept of telerehabilitation has emerged as an attractive method for rehabilitation at a distance, improving the quality of rehabilitation health care [10,11]. The diagnosis and assessment of musculoskeletal shoulder disorders through telerehabilitation have already been studied [12]. In this field, smartphones are well-known devices for therapeutic purposes [13], and mobile apps have transformed them into devices for clinicians [14].

Mobile phones, inertial sensors, or HALO digital goniometer have proven to be reliable instruments for the assessment of a range of motion [15,16]. These devices can be used in the clinical environment, offering savings in the costs of the assessments [17]. Devices such as the iPhone 4 have already proved to be a valid tool for measuring goniometry based on photographic measurements [18]. Smartphones are equipped with acceleration sensors that can measure variables such as speed, angulation, or acceleration [19]. These integrated sensors are accurate enough to provide angular measurements such as biofeedback in real time, if the measurement is short and the movement rate is within the effective frequency of the sensor [20]. The use of these devices in a clinical setting provides an increase in the accuracy of the record compared with visual assessments, and the same level of reliability is observed as in conventional goniometry [21].

An inertial measurement unit (IMU) sensor is one of the most reliable tools for measuring speed, acceleration, and displacement. However, there are few studies that have compared the reliability of a smartphone with that of an inertial sensor in the evaluation of patient movement [22]. The use of a smartphone's own internal sensor has high reliability when compared with the classic goniometer in static situations [22]. To the authors' knowledge, no additional studies have been conducted in which the reliability of a mobile device is linked to an IMU. However, smartphone devices have been used in the assessment of range of motion [17], and they are valid instruments for measuring different movements, except for the mobility of the hand. The authors of this study also mention the need to conduct validation studies in a dynamic environment.

Therefore, the aim of this study was to compare the intrasensor reliability of the measurements made by a dedicated inertial sensor with that of the measurements made by the inertial sensor integrated into a smartphone for the movements of shoulder flexion, extension, and abduction.

Methods

Design and Participants

This was a cross-sectional study that involved 43 subjects, of whom 16 were healthy and 27 were suffering from a shoulder injury. Asymptomatic subjects were recruited through advertisements. Patients were recruited from a specialized orthopedics clinic where they had been previously diagnosed by magnetic resonance imaging. Subjects were included if they were older than 18 years and had a body mass index (BMI) of between 18 kg/m^2 and 42 kg/m^2. Subjects were excluded if they refused to participate in the study. All participants were clinically examined by a physiotherapist and were interested in taking part in the project; none of them were found to meet any exclusion criteria.

Written informed consent was obtained from each individual. The study was approved by the ethics committee of the Faculty of Health Sciences at the University of Málaga, Spain.

Data Collection and Procedures

Descriptive and anthropometric independent variables related to age, gender, weight, size, and BMI were included. The Spanish version of DASH [23] and the Upper Limb Functional Index (ULFI) [24] questionnaires were used to obtain information about shoulder disability in pathological subjects.

A physical property was included corresponding to the dependent variable of velocity (degrees per second, [°/s]). This physical property was obtained through 2 different devices.

As the criterion standard, we used an IMU with 1 inertial sensor (InertiaCube3 Intersense Inc) with dimensions 26.2 × 39.2 × 14.8 mm and weight 17 g. It contained an inertial sensor with a 3-degree-of-freedom orientation tracking system: yaw, pitch, and roll, with accuracies of 1°, 0.25°, and 25°, respectively. It

also had an angular range of 360º and was able to detect an angular rate of between 0º/s and 1200º/s, with a sampling frequency of 1000 Hz. Activity values were recorded using kinematic Intersense Server Software.

The mobility angle was also measured along 3 orthogonal axes using the Nexus 4 (LG Electronics Inc) gyroscope (Invensense MPU-6050 Six-Axis [Gyro + Accelerometer]), which was attached to the posterior part of the humerus using an armband. The app used to obtain kinematic data was *Sensor Kinetics Pro* (Innoventions, Inc), which is available from Google Play. The *Nexus 4* has a storage capacity of 16 MB, and the data for each trial were transmitted by email for analysis and postprocessing. The data sampling rate was set to 14 Hz, allowing the device to record during all of the analytical tasks. Data from the smartphone and inertial sensors were subsequently passed to a Microsoft Excel 2007 database.

The inertial sensor was placed on the side of the body of each subject on which the shoulder presented pathology and was located on the middle third of the humerus, slightly to the posterior. Its surface was cleaned with alcohol to allow the sensor to adhere to the skin. To ensure fixation of the sensor to the patient's skin and to prevent slippage, a double-sided adhesive tape and an 8-cm-wide elastic cohesive (Rapidex) were used.

The smartphone was placed slightly below the inertial sensor and was snugly secured by a neoprene fixation belt over the humerus. The orientation and movement of the sensors are presented as roll, pitch, and yaw Euler angles. The equivalence of the axes on both devices and their anatomical interpretations are shown in Table 1.

After participants were recruited for the study, they were asked to attend the Human Movement Laboratory, Faculty of Health Sciences (University of Málaga). The tasks were explained concisely and clearly so that each participant understood the action to be performed. The beginning and end were determined by a verbal order by the researcher. Participants were instructed to stand; starting from a neutral position, they were asked to perform the following analytical tasks:

1. Shoulder abduction, with the elbow extended, wrist in a neutral position, and the palmar area of the hand toward the midline at the beginning and end of the movement (4 repetitions)
2. Shoulder flexion, with the elbow extended, wrist in a neutral position, and the palmar area of the hand toward the midline at the beginning and end of the movement (4 repetitions)
3. Shoulder scaption, with the elbow extended, wrist in a neutral position, and the palmar area of the hand toward the midline at the beginning and end of the movement (4 repetitions).

Data Processing

A computerized automatic analysis was conducted to filter the inertial sensor data. This analysis, which was designed to systematically obtain kinematic data for further statistical analysis, was performed using the basic software package R. As the sampling frequency of the Nexus 4 was 14 Hz and the frequency of the InertiaCube3 was 1000 Hz, the data were resampled to equalize the sampling frequencies at 100 Hz. Likewise, a common time 0 was established for all measurements made based on the time units obtained by the sensors. The automatic analysis was guided to obtain kinematic information from the accelerometer and gyroscope independently for each subject. The means and SDs of velocity in the 3 axes of movements (X, Y, and Z) were obtained from the accelerometer. The sign of the measured values of the accelerometer velocity along the X, Y, and Z axes is shown in Figure 1.

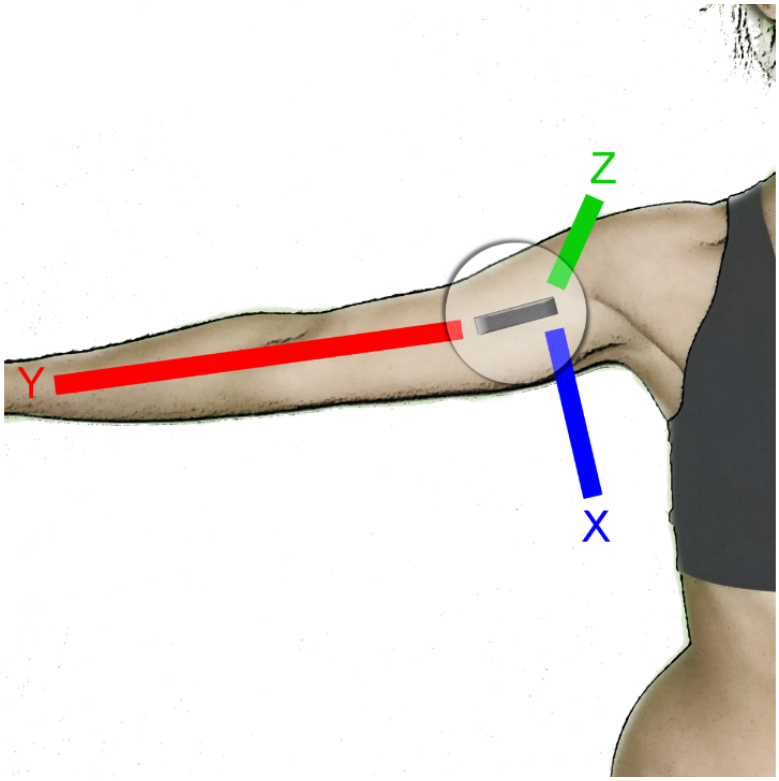
Statistical Analysis

A descriptive statistical analysis was conducted based on the means and SDs of the values characterizing the sample (age, weight, height, BMI, ULFI 100, and DASH 100). An interclass correlation analysis was performed between the variables recorded by the smartphone and the IMU. The results were analyzed for the 3 different gestures made by the participants (abduction, flexion, and scaption), and for each of these movements, the yaw, pitch, and roll axes were compared. The reliabilities of the mobile device and the IMU were estimated by means of the interclass correlation coefficient (ICC) and their 95% CIs based on a 2-way mixed model and absolute type using the SPSS statistical package version 22 [25]. The ICC value for a single measure is an index of the reliability of the ratings for 1 typical single rater (intrarater reliability). However, the average measure of the ICC is an index for the reliability of different rates averaged together (interrater reliability). This ICC is always higher than the single measurements of the ICC [26]. Values less than 0.5 are indicative of poor reliability, whereas values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability [27].

Table 1. Equivalence between the inertial sensor and smartphone placed on the humerus and the anatomical interpretation for shoulder movements.

Euler angles	Sensor axes	Smartphone axes	Anatomical axes	Anatomical planes	Shoulder movements
Yaw	Yaw	Z	Anteroposterior (dorsoventral)	Coronal (frontal)	Abduction
Pitch	Pitch	Y	Craniocaudal	Transverse (horizontal)	Rotation
Roll	Roll	X	Left-right	Midsagittal (median)	Flexoextension

Figure 1. X, Y, and Z motion axes.



Ethical Aspects

This study meets the criteria established by the Helsinki Declaration, and all participants were informed and signed an informed consent statement. The data obtained were treated anonymously. Ethical endorsement was obtained from the research committee of the Faculty of Health Sciences of the University of Málaga.

Results

The sample included 16 asymptomatic (controls) and 27 pathological subjects (patient) (7 with subacromial syndromes, 6 with supraspinatus tendon rupture, 9 with rotator cuff tear, 3 with supraspinatus tendinopathy, 1 with shoulder instability, and 1 with a Superior Labrum Anterior to Posterior lesion). The characteristics of the population and the results for shoulder functionality are shown in Table 2.

Shoulder Abduction

The ICC for the abduction shoulder movement was excellent for the abduction humeral movement (Table 3). This axis is related to the movement in the coronal plane, which directly determines the abduction movements (Table 1). The rotation and flexoextension movements of the humerus were moderate.

Shoulder Flexion

The ICC values for the flexion shoulder movement were excellent in terms of the abduction humeral movement between both the devices, and this was true for both the individual measurements and the averages of the measurements. In the same way, for the flexoextension humeral movement (left-right axes), the ICC was very good, both for the mean measures and single measures (Table 4).

Shoulder Scaption

For the scaption shoulder movement, excellent results were observed for the ICC for the abduction humeral movement (Table 5), and a very good ICC was recorded for the movements of flexion and humeral extension.

Table 2. Characteristics of participants.

Descriptive variable	Patients (n=27)	Controls (n=16)
Age (years), mean (SD)	52.8 (9.8)	55.6 (8.9)
Weight (kg), mean (SD)	77.1 (18.3)	73.7 (14.1)
Height (m), mean (SD)	1.6 (0.1)	1.6 (0.1)
Body mass index (kg/m^2), mean (SD)	28.4 (6.7)	26.9 (3.6)
Upper Limb Functional Index, % mean (SD)	70.1 (24.5)	0 (0)
Disabilities of the Arm, Shoulder and Hand, % mean (SD)	63.2 (20.4)	0 (0)

Table 3. Interclass correlation coefficient interdevices for abduction movement.

Humerus movement and measure	Interclass correlation coefficient (95% CI)	Cronbach alpha
Abduction		.980
Single	0.947 (0.841-0.978)	
Mean	0.973 (0.913-0.989)	
Rotation		.610
Single	0.435 (0.130-0.666)	
Mean	0.606 (0.230-0.800)	
Flexoextension		.656
Single	0.479 (0.186-0.696)	
Mean	0.648 (0.314-0.821)	

Table 4. Interclass correlation coefficient interdevices for flexion movement.

Humerus movement and measure	Interclass correlation coefficient (95% CI)	Cronbach alpha
Abduction		.925
Single	0.855 (0.733-0.924)	
Mean	0.922 (0.846-0.960)	
Rotation		.616
Single	0.444 (0.143-0.670)	
Mean	0.615 (0.251-0.803)	
Flexoextension		.876
Single	0.770 (0.593-0.876)	
Mean	0.870 (0.745-0.934)	

Table 5. Interclass correlation coefficient interdevices for scaption movement.

Humerus movement and measure	Interclass correlation coefficient (95% CI)	Cronbach alpha
Abduction		.948
Single	0.896 (0.795-0.949)	
Mean	0.945 (0.886-0.974)	
Rotation		
Single	0.427 (0.101-0.673)	.606
Mean	0.598 (0.184-0.804)	
Flexoextension		.824
Single	0.673 (0.413-0.830)	
Mean	0.805 (0.584-0.907)	

Discussion

Principal Findings

This study validated the use of a smartphone as an instrument to assess arm velocity during shoulder flexion, abduction, and scaption. An IMU was used as a reference system, and it proved to be an excellent, reliable, cheap, and easy-to-use tool for the measurement of humeral velocity. This approach would allow kinematic information to be transferred to a clinical context.

For the abduction movement (Table 3), the devices demonstrated excellent interclass reliability on the abduction humeral movement axis (alpha=.98); however, these values were not replicated in the rotation and flexoextension humeral movements. The same results were observed when both mean values and individual measurements were analyzed.

The intersensor reliability was excellent for the movements of flexion and scaption in abduction humeral movement, with ICC=0.925 and ICC=0.984, respectively (Tables 4 and 5). However, for the rotation and flexoextension humeral movement, the results ranged between moderate and good. The

reliability of the rotational humeral movement was moderate for the 3 movements analyzed (Tables 3-5). The agreement between the measurements made by the devices was good for the movements of flexion and scaption on the flexoextension humeral movement (Tables 4 and 5).

The lower reliability for the different axes of movement can be related to the compensation for the movement that is being made; this implies a modification of the speed of movement, which prevents an adequate record being made by the Nexus 4. Mourcou et al [22] observed that the Nexus 4 smartphone did not have the same reliability when analyzing the movement on the roll axis compared with the pitch axis. When the smartphone is in a static position, it has good roll and pitch reliability with an error of 0.2°, without significant differences; however, when there is an increase in velocity, the smartphone's detection algorithms have more difficulty in identifying and following the movement [22]. Likewise, the filter applied to the data also influences the variation in the recorded data, and the filter itself produces the most variation in the measurements.

Other studies have analyzed the reliability of mobile devices as clinical tools compared with an inertial sensor, in a similar way [22,28]. Pichonnaz et al validated the use of a smartphone (iPod, Apple) with an ICC of alpha=.97 compared with a Physilog reference system (Gait Up) using the B-B score test (hand to the Back and hand upwards as to change a Bulb) [29]. However, their study did not analyze the reliability of the isolated smartphone for each of the movements as in this study.

The inertial sensors included in smartphones can be taken as valid biofeedback if the variations in the acceleration are in the range ±8 g with an angular velocity of ±2000º /s and a sampling frequency of 100 Hz [30]. To these factors, we must add the smartphone's own error in the accelerometer record of ±40 mg and the error produced by the gyroscope of ±1°/s, which cannot be compensated for in the original data [15]. The lowest levels of reliability were observed in the rotational humeral movement relative to the corresponding craniocaudal axis (rotation abduction ICC=0.610, rotation flexion ICC=0.616, and rotation scaption ICC=0.606, as shown in Tables 3-5, respectively). These levels may be related to the internal limitations of the inertial sensor and to possible compensations made by the patients during the movements that are linked to the mobility of the scapula and that have a direct impact on the glenohumeral joint [31].

There are some inherent limitations of the method used in this study. First, the smartphone used was a Nexus 4, which is a relatively old model in terms of its inertial sensor; other, more current inertial sensors included in mobile devices may be more reliable. No additional filters were applied to the recorded data to keep the record intact in the same way as a clinician would, and this may imply a limitation on its reliability. In this study, only the humerus was evaluated, without assessing the possible implications that the scapula could have in terms of providing additional reliability or correlation. However, the application of our protocol was positive in both patients and healthy subjects. In the same way, the implementation environment was close to the clinical reality, and the use of a common mobile device fixed to the patient does not involve excessive equipment. There are few studies in which the reliability of a mobile device is evaluated in comparison with a reference system.

Conclusions

The objective of this study was to validate the Nexus 4 as a tool for the analysis of shoulder movement. Compared with an IMU, the inertial sensor included in the Nexus 4 smartphone proved to be a tool with excellent reliability for intersensor mediation in the velocities produced in the humerus during shoulder flexion, abduction, and scaption movements in the yaw axis. Its reliability is good when measurements are made on the pitch axis, which is linked to the left-right axis. To increase the reliability of the device, both the velocity of movement and the possible deviations or compensations that may appear must be controlled.

Conflicts of Interest

None declared.

References

1. Struyf F, Nijs J, Baeyens JP, Mottram S, Meeusen R. Scapular positioning and movement in unimpaired shoulders, shoulder impingement syndrome, and glenohumeral instability. Scand J Med Sci Sports 2011 Jun;21(3):352-358. [doi: 10.1111/j.1600-0838.2010.01274.x] [Medline: 21385219]
2. Roe Y, Soberg HL, Bautz-Holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the international classification of functioning, disability and health (ICF). BMC Musculoskelet Disord 2013 Feb 28;14:73 [FREE Full text] [doi: 10.1186/1471-2474-14-73] [Medline: 23445557]
3. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC, Roy J. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disabil Rehabil 2016;38(2):103-122. [doi: 10.3109/09638288.2015.1027004] [Medline: 25801922]
4. DASH: Disabilities of the Arm, Shoulder and Hand. URL: http://www.dash.iwh.on.ca/ [accessed 2018-09-26]
5. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. Am J Sports Med 2015 Oct;43(10):2572-2582. [doi: 10.1177/0363546514565096] [Medline: 25622986]

6. Cools AM, de Wilde L, van Tongel A, Ceyssens C, Ryckewaert R, Cambier DC. Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. J Shoulder Elbow Surg 2014 Oct;23(10):1454-1461. [doi: 10.1016/j.jse.2014.01.006] [Medline: 24726484]
7. Muir SW, Corea CL, Beaupre L. Evaluating change in clinical status: reliability and measures of agreement for the assessment of glenohumeral range of motion. N Am J Sports Phys Ther 2010 Sep;5(3):98-110 [FREE Full text] [Medline: 21589666]
8. Finley M, Jelinek JA, Misamore G. Three-dimensional analysis versus goniometric measurement of total active elevation in normal subjects. J Shoulder Elbow Surg 2015 Sep;24(9):1391-1396. [doi: 10.1016/j.jse.2015.02.005] [Medline: 25825139]
9. Picerno P, Viero V, Donati M, Triossi T, Tancredi V, Melchiorri G. Ambulatory assessment of shoulder abduction strength curve using a single wearable inertial sensor. J Rehabil Res Dev 2015;52(2):171-180 [FREE Full text] [doi: 10.1682/JRRD.2014.06.0146] [Medline: 26230401]
10. Russell TG. Telerehabilitation: a coming of age. Aust J Physiother 2009;55(1):5-6 [FREE Full text] [Medline: 19226236]
11. Theodoros D, Russell T. Telerehabilitation: current perspectives. Stud Health Technol Inform 2008;131:191-209. [Medline: 18431862]
12. Steele L, Lade H, McKenzie S, Russell TG. Assessment and diagnosis of musculoskeletal shoulder disorders over the internet. Int J Telemed Appl 2012;2012:945745 [FREE Full text] [doi: 10.1155/2012/945745] [Medline: 23193395]
13. Russell TG, Jones AF. Implications of regulatory requirements for smartphones, gaming consoles and other devices. J Physiother 2011;57(1):5-7 [FREE Full text] [doi: 10.1016/S1836-9553(11)70001-7] [Medline: 21402324]
14. Terry M. Medical apps for smartphones. Telemed J E Health 2010;16(1):17-22. [doi: 10.1089/tmj.2010.9999] [Medline: 20070172]
15. Cuesta-Vargas AI, Galán-Mercant A, Williams JM. The use of inertial sensors system for human motion analysis. Phys Ther Rev 2010 Dec;15(6):462-473 [FREE Full text] [doi: 10.1179/1743288X11Y.0000000006] [Medline: 23565045]
16. Correll S, Field J, Hutchinson H, Mickevicius G, Fitzsimmons A, Smoot B. Reliability and validity of the halo digital goniometer for shoulder range of motion in healthy subjects. Int J Sports Phys Ther 2018 Aug;13(4):707-714 [FREE Full text] [doi: 10.26603/ijspt20180707] [Medline: 30140564]
17. Milani P, Coccetta CA, Rabini A, Sciarra T, Massazza G, Ferriero G. Mobile smartphone applications for body position measurement in rehabilitation: a review of goniometric tools. PM R 2014 Nov;6(11):1038-1043. [doi: 10.1016/j.pmrj.2014.05.003] [Medline: 24844445]
18. Mitchell K, Gutierrez SB, Sutton S, Morton S, Morgenthaler A. Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation. Physiother Theory Pract 2014 Oct;30(7):521-525. [doi: 10.3109/09593985.2014.900593] [Medline: 24654927]
19. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove CT, Hart JM, et al. Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. J Shoulder Elbow Surg 2014 Nov;23(11):e275-e282. [doi: 10.1016/j.jse.2014.02.030] [Medline: 24925699]
20. Kos A, Tomažič S, Umek A. Suitability of smartphone inertial sensors for real-time biofeedback applications. Sensors (Basel) 2016 Feb 27;16(3):301 [FREE Full text] [doi: 10.3390/s16030301] [Medline: 26927125]
21. Shin SH, Ro DH, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther 2012 Aug;17(4):298-304. [doi: 10.1016/j.math.2012.02.010] [Medline: 22421186]
22. Mourcou Q, Fleury A, Franco C, Klopcic F, Vuillerme N. Performance evaluation of smartphone inertial sensors measurement for range of motion. Sensors (Basel) 2015 Sep 15;15(9):23168-23187 [FREE Full text] [doi: 10.3390/s150923168] [Medline: 26389900]
23. Hervás MT, Collado MJ, Peiró S, Pérez JL, Matéu PL, Tello IM. [Spanish version of the DASH questionnaire. Cross-cultural adaptation, reliability, validity and responsiveness]. Med Clin (Barc) 2006 Sep 30;127(12):441-447. [Medline: 17040628]
24. Cuesta-Vargas AI, Gabel PC. Cross-cultural adaptation, reliability and validity of the Spanish version of the upper limb functional index. Health Qual Life Outcomes 2013 Jul 26;11:126 [FREE Full text] [doi: 10.1186/1477-7525-11-126] [Medline: 23889883]
25. IBM. IBM SPSS Statistics URL: https://www.ibm.com/products/spss-statistics?mhsrc=ibmsearch_a&mhq=spss [accessed 2018-10-25]
26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420] [Medline: 18839484]
27. 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]
28. Pichonnaz C, Aminian K, Ancey C, Jaccard H, Lécureux E, Duc C, et al. Heightened clinical utility of smartphone versus body-worn inertial system for shoulder function B-B score. PLoS One 2017;12(3):e0174365 [FREE Full text] [doi: 10.1371/journal.pone.0174365] [Medline: 28319141]
29. Pichonnaz C, Lécureux E, Bassin JP, Duc C, Farron A, Aminian K, et al. Enhancing clinically-relevant shoulder function assessment using only essential movements. Physiol Meas 2015 Mar;36(3):547-560. [doi: 10.1088/0967-3334/36/3/547] [Medline: 25690269]

30. Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans Inf Technol Biomed 2006 Jan;10(1):156-167. [doi: 10.1109/TITB.2005.856864] [Medline: 16445260]
31. Lefèvre-Colau MM, Nguyen C, Palazzo C, Srour F, Paris G, Vuillemin V, et al. Kinematic patterns in normal and degenerative shoulders. Part II: review of 3-D scapular kinematic patterns in patients with shoulder pain, and clinical implications. Ann Phys Rehabil Med 2018 Jan;61(1):46-53 [FREE Full text] [doi: 10.1016/j.rehab.2017.09.002] [Medline: 28987866]

Abbreviations

BMI: body mass index
DASH: Disabilities of the Arm, Shoulder, and Hand
ICC: interclass correlation coefficient
IMU: inertial measurement unit
ULFI: Upper Limb Functional Index

Edited by G Eysenbach; submitted 06.02.19; peer-reviewed by S Sharif, N Chigateri; comments to author 28.04.19; revised version received 09.05.19; accepted 29.06.19; published 06.09.19.

Please cite as:
Roldán-Jiménez C, Martin-Martin J, Cuesta-Vargas AI
Reliability of a Smartphone Compared With an Inertial Sensor to Measure Shoulder Mobility: Cross-Sectional Study
JMIR Mhealth Uhealth 2019;7(9):e13640
URL: http://mhealth.jmir.org/2019/9/e13640/
doi:10.2196/13640
PMID:31493320

Review

Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions

Catherine Dinh-Le[1*], MPH; Rachel Chuang[2*], BA; Sara Chokshi[1*], DrPH; Devin Mann[1*], MS, MD

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

[2]Accenture, New York, NY, United States

[*]all authors contributed equally

Corresponding Author:
Devin Mann, MS, MD
Department of Population Health
New York University School of Medicine
227 E 30th Street
New York, NY, 10016
United States
Phone: 1 6465012503
Email: devin.mann@nyulangone.org

Abstract

Background: Due to the adoption of electronic health records (EHRs) and legislation on meaningful use in recent decades, health systems are increasingly interdependent on EHR capabilities, offerings, and innovations to better capture patient data. A novel capability offered by health systems encompasses the integration between EHRs and wearable health technology. Although wearables have the potential to transform patient care, issues such as concerns with patient privacy, system interoperability, and patient data overload pose a challenge to the adoption of wearables by providers.

Objective: This study aimed to review the landscape of wearable health technology and data integration to provider EHRs, specifically Epic, because of its prevalence among health systems. The objectives of the study were to (1) identify the current innovations and new directions in the field across start-ups, health systems, and insurance companies and (2) understand the associated challenges to inform future wearable health technology projects at other health organizations.

Methods: We used a scoping process to survey existing efforts through Epic's Web-based hub and discussion forum, UserWeb, and on the general Web, PubMed, and Google Scholar. We contacted Epic, because of their position as the largest commercial EHR system, for information on published client work in the integration of patient-collected data. Results from our searches had to meet criteria such as publication date and matching relevant search terms.

Results: Numerous health institutions have started to integrate device data into patient portals. We identified the following 10 start-up organizations that have developed, or are in the process of developing, technology to enhance wearable health technology and enable EHR integration for health systems: Overlap, Royal Philips, Vivify Health, Validic, Doximity Dialer, Xealth, Redox, Conversa, Human API, and Glooko. We reported sample start-up partnerships with a total of 16 health systems in addressing challenges of the meaningful use of device data and streamlining provider workflows. We also found 4 insurance companies that encourage the growth and uptake of wearables through health tracking and incentive programs: Oscar Health, United Healthcare, Humana, and John Hancock.

Conclusions: The future design and development of digital technology in this space will rely on continued analysis of best practices, pain points, and potential solutions to mitigate existing challenges. Although this study does not provide a full comprehensive catalog of all wearable health technology initiatives, it is representative of trends and implications for the integration of patient data into the EHR. Our work serves as an initial foundation to provide resources on implementation and workflows around wearable health technology for organizations across the health care industry.

(JMIR Mhealth Uhealth 2019;7(9):e12861) doi:10.2196/12861

KEYWORDS

wearable electronic devices; electronic health records; data collection; mobile health; patient monitoring

Introduction

Electronic Health Record Adoption and Expanded Access to Patient-Collected Data

Although electronic health records (EHRs) date back to the 1960s, widespread adoption was stagnant until the more recent passage of the Health Information Technology for Economic and Clinical Health Act in 2009 [1-5]. Between 2001 and 2011, the number of physicians using EHR systems increased from 18% to 57% [6]. Policies such as Meaningful Use (which prioritized quality, care coordination, and security of personal health information) incentivized the continued adoption of EHRs. By 2015, nearly 9 in 10 (87%) of office-based physicians adopted an EHR system [7-9]. Of all EHR vendors, Epic, Cerner, and Meditech are the most prevalent among health care systems [10].

In addition to driving EHR adoption among providers and health systems, legislation supporting meaningful use also paved the way for continued development of EHR capabilities to enhance the patient experience. Health systems are increasingly interdependent on EHR capabilities, offerings, and innovations to better capture patient data [11]. Features include secure messaging with patients and features to view, download, and transmit their EHR. Such capabilities are becoming more prevalent to facilitate streamlined patient data exchanges with their provider [7].

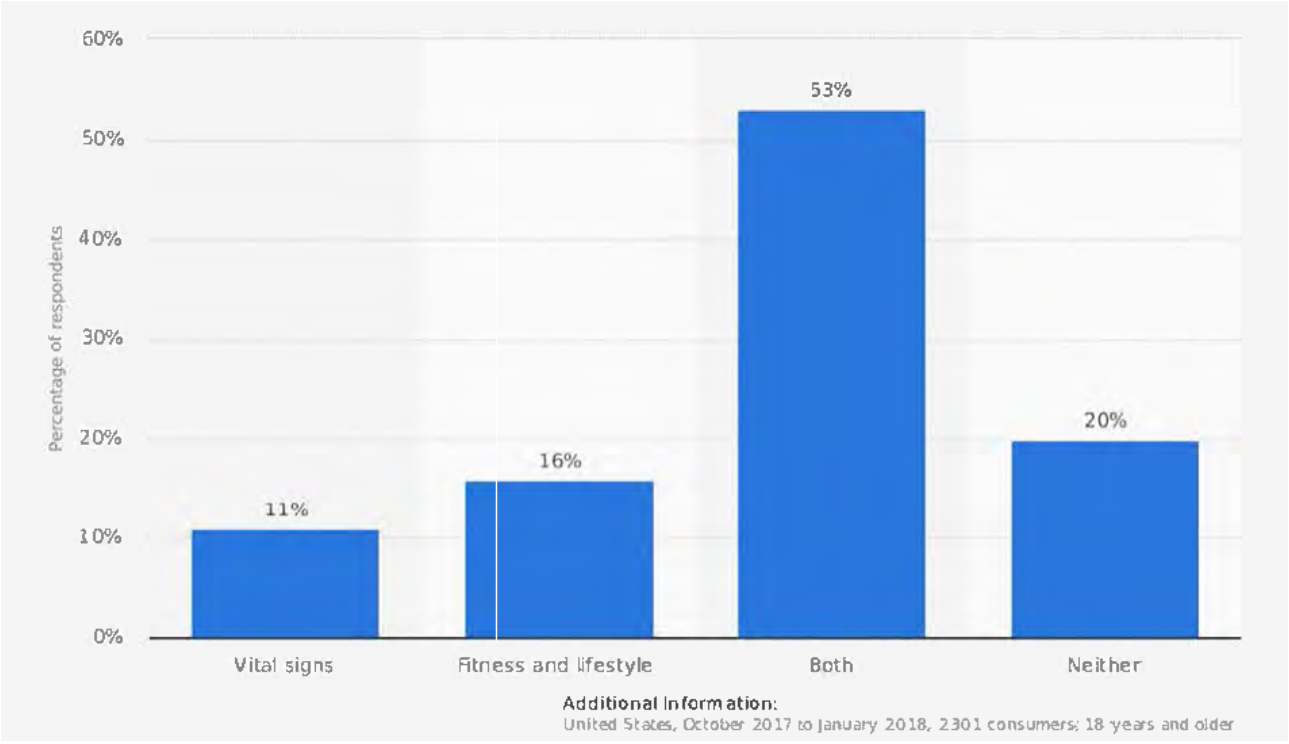
A novel capability offered by health systems encompasses the integration between EHRs and medical devices, including wearable health and fitness tracking devices. Although early device integration involved tracking a set of simple vital signs, the scope of patient data has expanded rapidly as health systems strive to meet new standards, new care models, as well as leverage innovation in digital technologies [12,13]. The primary focus of this review was to capture a sample of the rapidly changing field of patient data integration into the EHR [14]. Specifically, we review several health systems and organizations that are using patient data gathered through consumer-grade wearable devices to track and improve patient outcomes.

Availability and Adoption of Wearable Devices

Wearable devices include wristbands, smartwatches, wearable mobile sensors, and other mobile *hub* medical devices that collect a large range of data from blood sugar and exercise routines to sleep and mood. Patient data are collected either through consumer reporting or passively through sensors in apps that communicate with devices through application programming interfaces (APIs); these data are then shared through data aggregators such as Apple's HealthKit that pools data from multiple health apps [15].

According to a recent consumer survey on digital health by Accenture, a significant percentage of US adults were willing to wear technology that tracks their health statistics (see Figure 1) [16]. Due to mobile integration platforms such as Google Fit and Apple HealthKit, we can expect to see an increase in the number of health-wearable users over the next few years [17,18]. The upward trend in device usage to monitor health-related data additionally suggests there will be a correlated rise in patient data available for health management [19]. Large health systems are likely to trend toward larger rollouts of wearable technology in the next few years, potentially incorporating wearables as part of their preventative care strategy by monitoring heart rate, blood pressure, and other information [20,21]. There are currently more than 400 EHR-compatible devices on the market, a number that is expected to rise exponentially in the coming years [22].

Figure 1. Percentage of US adults who were willing to wear technology that tracks select health statistics as of 2018. Screenshot from www.statista.com [16].



Clinical Impact of Wearable Devices

Currently, these devices have the potential to help patients and providers manage chronic conditions such as diabetes, heart conditions, and chronic pain [23-25]. According to the Pew Research Center, 60% of US adults reported tracking their weight, diet, or exercise routine; 33% of US adults track health symptoms or indicators such as blood pressure, blood sugar, or sleep patterns; and 8% of adults specifically use medical devices, such as glucose meters [26]. Studies on the clinical impact of wearables on patient health outcomes offer varied results. Although some conditions such as physical activity and sleep did not show significant or conclusive change from wearable technology use and require further evaluation, other studies have reported improved subjective outcomes on patient health [27,28].

Recent literature reviews on the clinical impact of wearable devices and behavior change have shown promising effectiveness for digital technology [29]. However, much of the literature calls for more complete data analyses from commercially available tools and their impact on patients [30,31]. Further studies are necessary to assess clearer clinical outcomes on patient health by wearable health technology.

The purpose of this paper was to conduct a scoping review of the wearable health technology field to provide an overview of current wearable innovations in the EHR. Similar to a number of existing scoping reviews, we used internet search engines in addition to our database searches to capture the rapid updates in the area of health system integration of remotely collected patient data [32]. We used these sources to generate a targeted list of organizations that are leaders in the overall field of wearable health technology, along with their partnerships.

This paper provides an overview of (1) our process in determining the current landscape of wearable health technology and (2) descriptions of some leading innovations and partnerships by start-ups, providers, and insurance companies. By sharing our results, we hope to create a process to identify relevant organizations in this field and provide resources for organizations that are interested in joining or learning more about implementation and workflows around wearable health technology and patient data integration to EHRs. This study is specific to integration into the Epic portal and is not a comprehensive search; however, results are representative of the field because of Epic's prominence in the US acute care hospital market (25.8%) [10].

Methods

Search Process

To better understand the scope of wearables and other health tracking devices and the resulting impact on EHRs, we used a scoping process to survey existing efforts on the Web. Although not directly relevant to a scoping review, we reviewed Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines to enhance the quality of our search. In our search, to identify the leaders in the field, we contacted the largest commercial EHR system (Epic) for information on client work in the integration of patient-collected data. We used this information to further inform our search terms in Epic's UserWeb portal, the primary platform for Epic users to share and discuss topics such as innovative idea generation and event postings, and the general Web, PubMed, and Google Scholar database searches from July 2018 to January 2019.

We recognize the risk of bias in this study, as our search process was limited to Epic clients and the information publicly available on the World Wide Web.

Inclusion Criteria

We used a set of inclusion criteria in UserWeb to ensure that postings were accurate and up to date. Results had to meet the following standards: (1) be posted after June 2017 and (2) have responses to topic threads. Key search terms included Apple HealthKit, Patient remote data integration, Fitbit integration, Withings integration, and Wearables.

Similarly, our findings on wearable technology companies and initiatives from the general Web, PubMed, and Google Scholar database searches had to meet the following criteria: (1) be posted after June 2017 and (2) match search terms including but not limited to Device integration, EHR data integration, Epic MyChart integration, Patient MyChart integration, Patient remote data integration, Patient data access, Wearables, Provider wearables, Hospital wearables, Hospitals AND Apple HealthKit device integration, Apple HealthKit device integration AND Epic, Start-ups AND EHR integration, Insurance companies AND device integration, APIs AND device integration.

Results

Challenges of Wearable Device Integration

Although wearable health technology has the potential to transform patient care, issues such as concerns with patient privacy, system interoperability, and the immense amount of patient data pose a challenge to the adoption of wearables by providers [33,34]. Such challenges are critical to consider for future wearable use to deliver safe and quality care for patients. Although there are potential solutions for these implementation issues, more innovative work is required for wide-scale adoption of wearable health technology.

Protecting the Confidentiality and Privacy of Patients

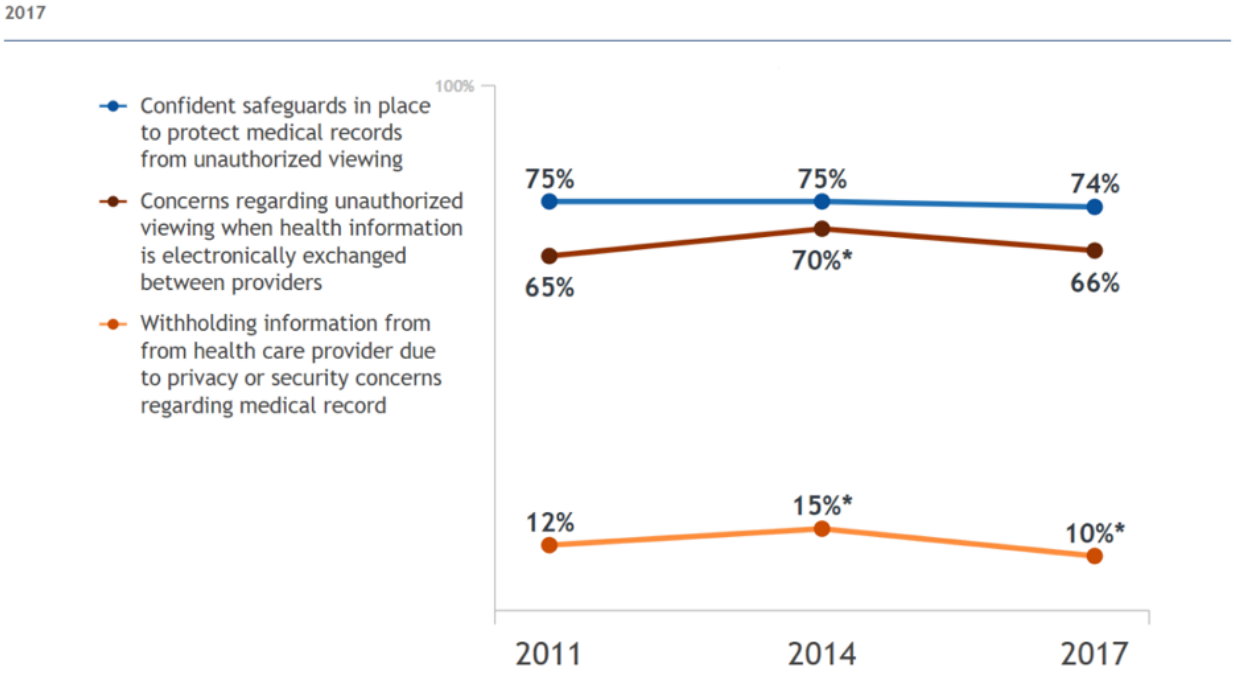
Wearable health technology requires critical checkpoints along the workflow to protect the confidentiality and privacy of patients [35]. Currently, there is limited empirical evidence in the literature on the appropriate implementation of security in wearable devices [36,37]. Key considerations include Health Insurance Portability and Accountability Act of 1996 (HIPAA) compliance and informed consent by wearable users.

The HIPAA is a US legislation that protects the privacy of individuals' medical records and applies to health providers and plans [38]. With the continuous stream of data from personal devices, data privacy and security for health information must be addressed as to meet HIPAA standards and not impede patients' willingness to share their data [39]. Figure 2 demonstrates that patients have some concerns about the electronic exchange of data between providers; the percentage of individuals expressing these concerns has remained relatively

the same since 2011 [7]. To protect against potential cybersecurity attacks and missing or stolen patient records through the implementation of wearable health technologies, hospitals must ensure that devices are connected to a secure network and monitor the hospital data network continuously [40]. To prioritize data privacy, health systems are likely to be required to set up another secure network for wearable devices, separate from the main network [41].

The complexities of wearables continue to grow as patient datasets from wearable devices are compiled and transferred [42]. Obtaining patient consent is also critical, as patients are likely to find constant physiological surveillance to be intrusive [43]. Misuse of personal health information by third parties could lead to discrimination, changes in insurance coverage, or even identity theft [15]. As a result, consent notices must provide enough detail regarding what and how often personal information is collected and specify the third parties that can access patient data, ensuring that informed consent by the patient occurs [42,44]. Additional policies and standards are necessary for the future of wearable health technology and patient data integration to the EHR to ensure the confidentiality and privacy of patients.

Figure 2. Individuals' perceptions of the privacy and security of medical records and health information exchange in 2017. Screenshot from https://dashboard.healthit.gov/quickstats/quickstats.php [7].



Lack of System Interoperability and Connectivity

As the integration of patient data through wearable devices is a relatively new area of health technology, health systems are lacking the necessary platforms to pull continuous streams of data from different patient devices for integration into the EHR [45]. Currently, device and EHR vendors use a range of methods that include distinct, proprietary, and closed communication methods [46,47]. These differences in methods make it difficult for various devices and EHR systems to communicate and transfer data streams, leading to the lack of system interoperability.

As a result, this barrier has created subsets of data collected from patients that become secondary in value because they cannot be easily integrated into patient historical data [48,49]. Researchers have recently looked to achieve *plug-and-play* interoperability to standardize platforms and integrate these information islands, a standard that already exists in the world of consumer electronics as consumers demand simple and seamless functionality [47]. *Plug-and-play* standards require ease of use, device compatibility, and streamlined scalability and reconfigurability between different vendors; systems must be able to detect new devices, negotiate communication, and allow devices to synchronize and work with each other [50].

As the need for system interoperability grows, third-party applications aimed to address interoperability issues have become more prominent [45]. Increased partnerships and opportunities between makers of these applications and health systems are necessary to reach high interoperability and streamlined communication between EHR platforms, patient devices, and providers. Improving these relationships can improve health care efficiency, provider safer transitions of care, and help lower health care costs [51].

Patient Information and Data Overload

Wearable health technology that is integrated into the EHR produces an enormous amount of data that require compilation and interpretation before becoming useful for patients and providers [43,52]. Storing daily patient data streams can be a barrier to health systems that are not prepared to host a database that is constantly growing [53]. Decisions around the life cycle of such data and how it can best fit into provider workflows

pose a unique challenge to using remotely collected data for patient care [52]. For example, the Apple Health and PulseOn Android apps provide heart rate data at 60-second long and 3-second long intervals, respectively; transmission of such large volumes of data will require backend analysis to be processed into a simpler and more usable form [54].

Due to the sheer volume of these data, extracting and presenting providers with necessary patient data has been a main discussion point among hospitals implementing wearable technology. Overall, many providers experience alert fatigue in their daily clinical decision support systems [55]. Although machine learning and artificial intelligence (AI) algorithms are potential solutions to this issue, current algorithms are often tested in fixed conditions that are not likely to hold up in live scenarios [35]. Successful solutions to patient data integration should be able to sift through the immense amount of data and automatically deliver meaningful and actionable items to providers [56].

In addition, a strong user interface (UI) for providers is important for provider buy-in and engagement during implementation. As a result, there has been an increasing trend within health care organizations to incorporate user experience and UI designers into a cross-functional information technology (IT) team to address this need [57]. The multidisciplinary skills of such teams can offer improved UIs combined with IT expertise and enhance the ability to comprehend wearable patient data. These improvements in provider engagement and workflow could improve overall time efficiency for providers and quality of care for patients.

Innovations in Wearable Health Technology

In response to these challenges, a number of health systems and organizations have begun to use a user-centered design approach to adapt workflows and collaborate with third-party applications to improve their integration of remote patient data [58,59]. Numerous health care providers have piloted and/or implemented wearable-EHR integration projects with Apple Health, Google Fit, Fitbit, Nokia, and Withings [60]. A number of devices on the market have the capability to connect directly to EHRs through HealthKit and Google Fit; simple data such as steps and weight are currently collected and displayed, with more devices and data types being brought on the Web over time [58,60]. In addition, as of October 2018, Epic customers representing at least 565 hospitals and 14,427 clinics support connecting data from Fitbit, HealthKit, or Withings today. Epic customers representing at least 1152 hospitals and 24,496 clinics support connecting other devices through Health Level-7 or manual entry of patient data through MyChart. Note that this is not a comprehensive list of all customers, as select organizations opted out of the data collected by Epic (data provided by Epic, October 2018).

However, EHRs still cannot connect to many other devices and require the development of new solutions to address challenges such as interoperability and visualization for the information they are currently collecting [61]. The wearable health technology space features numerous start-up partnerships with health care providers and insurance company innovations that are working to address these key challenges and promote growth in wearable usage and EHR integration capabilities.

The overall themes that we used to describe the different focus areas of each partnership included personalized patient experience, rewards program, data analytics, remote monitoring, access to patient records, and AI technology. A summary of key organizations working in wearable health technology compiled from the general Web search and Epic's UserWeb portal (as of May 2018) is presented in Tables 1 and 2, respectively.

Start-Up Partnerships

As listed in Table 1 below, we identified the following 10 start-up organizations that have developed or are in the process of developing technology to improve wearable health technology and/or patient data integration to EHRs: Overlap, Royal Philips, Vivify Health, Validic, Doximity Dialer, Xealth, Redox, Conversa, Human API, and Glooko. We reported sample start-up partnerships with a total of 16 health systems in addressing challenges of meaningful use of device data and streamlining provider workflows. The partnerships between these start-ups and health systems serve to improve the data collection process, synthesize actionable information for providers to review, and create a more personalized experience between patients and providers. Due to the rapidly moving field of wearables, our research represents a snapshot in time of wearable health technologies and is not meant to be a fully exhaustive list.

Table 1. Wearable health technology start-up partnerships.

Start-up organizations	Select hospital partnership(s)	Theme(s)	Technology overview
Overlap 2019 [62]	Columbia University Medical Center and UC Davis Health	Data analytics and remote monitoring	Collects patient data through a customizable Overlap app that integrates with EHRs[a] and various wearable devices
Royal Philips 2019 [63]	New York Presbyterian	Data analytics and remote monitoring	Helps physicians monitor patient health remotely and connect with 2-way video using a telehealth platform
Vivify Health 2018 [64]	Children's Health in Dallas and Ascension Health	Remote monitoring	Integrates patient mobile devices with EHRs through a remote care platform
Validic 2018 [65]	Kaiser Permanente and Mayo Clinic	Data analytics and remote monitoring	Simplifies collected health data from wearables and wellness applications and delivers comprehensive patient profiles to providers
Doximity Dialer 2018 [66]	Johns Hopkins Hospital	Access to patient records and personalized patient experience	Allows providers to access their patients' records and make patient calls on the go from their personal cell phones, using the office as the caller ID[b] while on personal phones
Xealth 2018 [67]	Providence Health & Services and University of Pittsburgh Medical Center	Personalized patient experience	Allows doctors to prescribe apps and digital tools to their patients. Doctors can also track patient's use of these tools from the EHR
Redox 2018 [68]	Brigham and Women's Hospital	Data analytics	Links hospitals' EHR systems to outside applications regardless of software vendor (Epic, and Allscripts)
Conversa 2018 [69]	Northwell Health and Ochsner Health System	Artificial intelligence technology and personalized patient experience	Allows providers to monitor patient status between visits through automated, personalized patient-provider conversation experiences. Patient also can send information through Conversa into their EHRs
Human API 2018 [70]	Mount Sinai and Cedars-Sinai	Data analytics	Pulls health data in real time and processes and normalizes actionable health data, regardless of source or original format
Glooko 2019 [71]	Mayo Clinic and Novant Health	Data analytics, personalized patient experience, and remote monitoring	Provides daily insights to people with diabetes through a mobile app; clinicians are able to access data and identify high-risk patients

[a]EHR: electronic health data.

[b]ID: identification.

Table 2. Insurance companies.

Organization	Theme(s)	Technology overview
Oscar Health 2018 [72]	Rewards program	Uses an app that synchronizes with Apple Health for its step-tracking program. More than three-fourths (80%) of Oscar members who download the app use step tracking
United Healthcare 2018 [73]	Rewards program	Offers UnitedHealthcare Motion program where members can earn money toward out-of-pocket medical expenses by walking. The United Healthcare Motion app syncs with wearables using Qualcomm Life's 2net Platform to track steps
Humana 2018 [74]	Personalized patient experience and rewards program	Launched Go365, a wellness and rewards program for members in 2017. The program operates on a points system and incentivizes healthier behavior with personalized health assessments and rewards, such as fitness gear and electronic devices
John Hancock 2018 [75]	Rewards program	Offers Vitality Points for physical activity and health screenings, which can be used for gift cards and travel. Policyholders can save up to 15% on their life insurance by using internet-connected Fitbits

Insurance Companies

In addition, we compiled a number of insurance companies that encourage the growth and uptake of wearable health technology through incentive programs: Oscar Health, United Healthcare, Humana, and John Hancock (see Table 2). These companies all offer health tracking through devices and promote the use of remote patient data to improve patient engagement and health. Key focus areas included patient data tracking and rewards programs for customers who use devices to track their health and achieve milestones. These rewards programs gamify health goals into point systems and offer incentives for customers, including gift cards, electronic devices, and travel. These initiatives by insurance companies support the uptake of wearable health technologies and expand the use of patient-collected data to improve patient health.

In addition to the tables above, we identified a number of health systems and organizations that engaged or stated interest in wearable health technology initiatives, such as NYU Langone Health, Penn Medicine, Duke Health, Novant Health, and Icahn School of Medicine at Mount Sinai. Such work includes integration of Fitbit and HealthKit data into patient health portals [60,76,77]. However, these groups have not yet published results from their work or current status of innovation. Limited reported information is likely to be because of the early stages of implementation and require follow-up in a future review of current innovations.

Data Analysis

On the basis of the information collected from our survey sample of 10 start-up organizations and 4 insurance companies, the most common themes included a personalized patient experience based on health goals and past medical history, gamification through a rewards program, and data analytics capabilities (see Table 3).

We also recorded several key observations based on analyzed data:

1. Current rewards programs are strongly linked with wearable devices. Of the identified organizations, all rewards programs relied on the use of wearable devices to track data that could be used for patient incentives. The most common data point was step tracking; a patient could earn money or points to be traded in for prizes when they walked a certain number of steps each day.
2. AI capabilities are still limited. AI has yet to become fully established in the field of wearable health technology. A limited number of organizations are leveraging these digital capabilities to collect, analyze, and integrate patient data and monitoring and creating ongoing dialog about patient health activities.
3. There are varied approaches for personalization of patient information. Personalization of a patient's experience was a prevalent theme across several of the surveyed organizations. The personalized experience was created through various approaches, including recommending health apps, facilitating ongoing conversations with a doctor or AI bot, or providing assessments so that a patient could better understand their health.
4. There are challenges and risks to all aspects of wearable health technology. Addressing system interoperability, patient privacy, and data overload risks will be critical to the use of wearable health technology. We mapped out the previously discussed challenges for each of the 6 themes in Table 4.

Table 3. Prevalence of wearable health technology themes across surveyed start-ups and insurance companies.

Theme	Number of surveyed organizations addressing themes
Personalized patient experience	4
Rewards program	4
Data analytics	3
Remote monitoring	2
Access to patient records	1
AI[a] technology	1

[a]AI: artificial intelligence.

Table 4. Challenges and risks associated with wearable health technology.

Theme	Challenges		
	System interoperability	Patient privacy	Data overload
Personalized patient experience	—[a]	X[b]	—
Rewards program	—	X	—
Data analytics	—	—	X
Remote monitoring	X	X	X
Access to patient records	X	X	X
AI[c] technology	—	X	—

[a]No expected challenge or risk associated with wearable technology theme.

[b]X: challenge or risk associated with wearable technology theme.

[c]AI: artificial intelligence.

Discussion

Principal Findings

This scoping study reviewed current innovations of wearable health technology and EHRs across health care systems, start-ups, and insurance companies and documented key innovation trends, partnerships, and incentives, along with challenges of wearables. Our findings reflect the movement toward the adoption of mobile health devices through the availability of digital tools and gamification of health data collection. However, numerous barriers to the efficient implementation of wearable health technology exist and are likely to hinder widespread adoption across health systems. Our report presents several current approaches to addressing wearable health technology and EHR integration barriers; these findings highlight the direction of wearable health innovation and serve to identify potential partnerships for future wearable adoption.

The development of technologies by start-ups outside of EHR systems highlights the interest in solving challenges in wearable health technology, such as information overload and system interoperability. Companies such as Redox are addressing interoperability issues by creating the technology to link hospitals' EHR systems to outside applications regardless of software vendor. Others, such as Validic and Human API, are working to improve the workload for providers by simplifying the data collection from devices and outputting processed and easily understandable results.

Across the field of wearable health technology, maintaining patient privacy with the expanding use of wearables, rewards programs, remote monitoring, and AI continues to pose the greatest challenge to the growth of wearable health technology. Obtaining informed patient consent will be critical to provide clarity regarding what data are collected and which third parties can access patient data; this will continue to be a key discussion topic, as organizations seek to create a personalized patient experience based on patient-collected data. For example, companies such as Conversa allow for automated and personalized virtual care using conversational AI technology and patient remote data.

The implementation of health tracking rewards programs by insurance companies additionally signifies the interest and direction in which wearable health technology is moving to improve consumer health. These health insurance companies' decisions to engage in wearables through rewards programs can offer increased opportunities in data collection and need for the above start-up technologies to provide a seamless experience for both providers and consumers. As wearable health technology becomes linked to gamification and rewards program initiatives for insurance companies, patient data integration across other platforms is also likely to become more commonplace.

This report serves as a starting point for those interested in wearable innovations rather than a comprehensive summary because of the rapidly changing nature of wearable health technology. As more institutions share their work in this area,

address challenges, and create more efficient workflows/processes, the ability to transform patient care and streamline the integration of mobile health devices will improve the health outcomes and quality of care for patients.

Limitations

Although we developed a detailed process to search and document the current state of wearables through our study, several challenges exist to create a comprehensive list. The nature of this report is Epic centric, as we were not able to access other internal EHR portals. There were a limited number of health systems actively publicizing or publishing their work on new integration methods through the general Web. Those that did also used different names (ie, remote data integration and device integration) that may not have been included in our search terms. Furthermore, based on the growing adoption of wearable health technology in health systems over the past few years, we anticipate that new names would have been added to this list since our search.

Conclusions

Wearable health technology will play a critical role in greater transparency between patients and providers and chronic condition management. Devices and technologies that enable the streamlined movement of data from patients to providers are key to improving a patient's care journey and empowering them to manage their own health. The future design and development of digital technology in this space will rely on continued analysis of best practices, pain points, and potential solutions to mitigate existing challenges.

By sharing our results, we have presented key challenges and emerging solutions to this rapidly evolving field. Our work serves as an initial foundation to the creation of a streamlined process to identify relevant entities in this field and provide resources on the implementation of and workflows around wearable health technology and EHR integration for organizations across the health care industry. As much of this work is still ongoing, we anticipate that these findings will serve as the foundation for future studies on wearable health technology.

Conflicts of Interest

None declared.

References

1. Tanner C, Gans D, White J, Nath R, Pohl J. Electronic health records and patient safety: co-occurrence of early EHR implementation with patient safety practices in primary care settings. Appl Clin Inform 2015;6(1):136-147 [FREE Full text] [doi: 10.4338/ACI-2014-11-RA-0099] [Medline: 25848419]
2. Atherton J. Development of the electronic health record. Virtual Mentor 2011 Mar 1;13(3):186-189 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.3.mhst1-1103] [Medline: 23127323]
3. Tripathi M. HIM Body of Knowledge - Ahima. 2012. EHR Evolution: Policy and Legislation Forces Changing the EHR URL: http://library.ahima.org/doc?oid=105689#.W74na2hKg2y [accessed 2019-08-08]
4. The University of Scranton. 2018. EMR: The Progress to 100% Electronic Medical Records URL: https://elearning.scranton.edu/resource/health-human-services/emr_the-progress-to-100-percent-electronic-medical-records [accessed 2019-08-08] [WebCite Cache ID 734HVy2po]
5. Cohen MF. Impact of the HITECH financial incentives on EHR adoption in small, physician-owned practices. Int J Med Inform 2016 Oct;94:143-154. [doi: 10.1016/j.ijmedinf.2016.06.017] [Medline: 27573322]
6. HealthIT. 2017. Electronic Health Records Infographic URL: https://www.healthit.gov/infographic/electronic-health-records-infographic [accessed 2018-10-10] [WebCite Cache ID 734I2kXHy]
7. Health IT Dashboard. 2018. Quick Stats URL: https://dashboard.healthit.gov/quickstats/quickstats.php [accessed 2019-08-08] [WebCite Cache ID 734FqslSF]
8. Centers for Disease Control and Prevention. 2017. Electronic Medical Records/Electronic Health Records (EMRs/EHRs) URL: https://www.cdc.gov/nchs/fastats/electronic-medical-records.htm [accessed 2019-08-08] [WebCite Cache ID 734FiMEmY]
9. Multi Care. 2018. Epic (Electronic Health Records) URL: https://www.multicare.org/care-connect/epic/ [accessed 2018-10-10] [WebCite Cache ID 734EVc2Ql]
10. Vaidya A. Becker's Hospital Review. 2017. Epic, Cerner Hold 50% of Hospital EHR Market Share: 8 Things to Know URL: https://www.beckershospitalreview.com/healthcare-information-technology/epic-cerner-hold-50-of-hospital-ehr-market-share-8-things-to-know.html [accessed 2019-05-27] [WebCite Cache ID 78fihENG5]
11. Bender AJ, Mecklenburg RS. Harvard Business Review. 2017. How the EMR Is Increasing Innovation and Creativity in Health Care URL: https://hbr.org/2017/10/how-the-emr-is-increasing-innovation-and-creativity-in-health-care [accessed 2019-05-27] [WebCite Cache ID 78filh3p6]
12. Fratt L. AI in Healthcare. 2013. Medical Device Integration: A Look Past the EHR URL: https://www.clinical-innovation.com/topics/technology-management/medical-device-integration-look-past-ehr [accessed 2018-10-10] [WebCite Cache ID 734KmoUEY]

13. McGeath J. Today's Medical Developments. 2014. Medical Device Integration With EHRs: The Landscape and Outlook for the Future URL: http://www.todaysmedicaldevelopments.com/article/tmd0614-medical-device-integration-ehr/ [accessed 2018-10-10] [WebCite Cache ID 734JtQubP]
14. Conn J. Health Data Management. 2018. Apple Health App Effort to Engage Consumers by Sharing EHR Info URL: https://www.healthdatamanagement.com/news/apple-health-app-effort-to-engage-consumers-by-sharing-ehr-info [accessed 2018-10-10] [WebCite Cache ID 7348HLDHZ]
15. Grundy Q, Held FP, Bero LA. Tracing the potential flow of consumer data: a network analysis of prominent health and fitness apps. J Med Internet Res 2017 Jun 28;19(6):e233 [FREE Full text] [doi: 10.2196/jmir.7347] [Medline: 28659254]
16. Statista. 2018. Percentage of US Adults That Were Willing to Wear Technology That Tracks Select Health Statistics as of 2018 URL: https://www.statista.com/statistics/829479/willingness-to-wear-health-tracking-technology-us-adults/ [accessed 2019-05-26] [WebCite Cache ID 78fMKO3oX]
17. Gruessner V. mHealth Intelligence. 2015. Rise of Mobile Health Technology Brings More Data Analysis URL: https://mhealthintelligence.com/news/rise-of-mobile-health-technology-brings-more-data-analysis [accessed 2018-10-10] [WebCite Cache ID 7349OTsYz]
18. Steakley L. Scope Blog - Stanford University. 2015. Harnessing Mobile Health Technologies to Transform Human Health URL: http://scopeblog.stanford.edu/2015/03/16/harnessing-mobile-health-technologies-to-transform-human-health/ [accessed 2018-10-10] [WebCite Cache ID 7349EXSF6]
19. Al-Siddiq W. Hitconsultant. 2018. Medical wearables: how next generation devices will change healthcare URL: https://hitconsultant.net/2018/01/24/medical-wearables-devices/ [accessed 2019-09-05] [WebCite Cache ID 7348besoc]
20. Pettey C. Gartner Inc. 2018. Wearables Hold the Key to Connected Health Monitoring URL: https://www.gartner.com/smarterwithgartner/wearables-hold-the-key-to-connected-health-monitoring/ [accessed 2018-10-10] [WebCite Cache ID 73494i0PV]
21. Joszt L. American Journal of Managed Care. 2018. 5 Things About Wearable Technology in Healthcare URL: https://www.ajmc.com/newsroom/5-things-about-wearable-technology-in-healthcare [accessed 2018-10-08] [WebCite Cache ID 731RznqVv]
22. Vandrico. 2018. Wearable Technology Database URL: https://vandrico.com/wearables/ [accessed 2018-10-10] [WebCite Cache ID 7349F35oO]
23. 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 [FREE Full text] [doi: 10.1007/s13347-016-0215-5]
24. Pai A. Mobi Health News. 2015. Global Health Tracking Technology Market to Reach $18.8B in 2019 URL: https://www.mobihealthnews.com/44873/global-health-tracking-technology-market-to-reach-18-8b-in-2019 [accessed 2018-10-10] [WebCite Cache ID 7349WxSVr]
25. BCC Research. 2015. Mobile Devices Driving Unprecedented Growth in Self-Monitoring Technologies Markets, According to BCC Research URL: https://www.bccresearch.com/pressroom/hlc/mobile-devices-driving-unprecedented-growth-in-self-monitoring-technologies-markets [accessed 2019-08-08] [WebCite Cache ID 7349XUSpA]
26. Fox S. Pew Research Center. 2013. The Self-Tracking Data Explosion URL: http://www.pewinternet.org/2013/06/04/the-self-tracking-data-explosion/ [accessed 2018-10-10] [WebCite Cache ID 7349NvePP]
27. Crowley O, Pugliese L, Kachnowski S. The impact of wearable device enabled health initiative on physical activity and sleep. Cureus 2016 Oct 11;8(10):e825 [FREE Full text] [doi: 10.7759/cureus.825] [Medline: 27882272]
28. Yurkiewicz IR, Simon P, Liedtke M, Dahl G, Dunn T. Effect of Fitbit and iPad wearable technology in health-related quality of life in adolescent and young adult cancer patients. J Adolesc Young Adult Oncol 2018 Oct;7(5):579-583. [doi: 10.1089/jayao.2018.0022] [Medline: 29924668]
29. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Human Behav 2016 Apr;57(3):61-74. [doi: 10.1016/j.chb.2015.12.023]
30. Rehman H, Kamal AK, Sayani S, Morris PB, Merchant AT, Virani SS. Using mobile health (mhealth) technology in the management of diabetes mellitus, physical inactivity, and smoking. Curr Atheroscler Rep 2017 Apr;19(4):16. [doi: 10.1007/s11883-017-0650-5] [Medline: 28243807]
31. Cooper C, Gross A, Brinkman C, Pope R, Allen K, Hastings S, et al. The impact of wearable motion sensing technology on physical activity in older adults. Exp Gerontol 2018 Oct 2;112:9-19. [doi: 10.1016/j.exger.2018.08.002] [Medline: 30103025]
32. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
33. Eramo L. Future Health Index. 2017. 5 Challenges of Integrating Wearable Data Into Electronic Health Records URL: https://www.futurehealthindex.com/2017/10/23/challenges-wearable-data-health-records/ [accessed 2019-08-08] [WebCite Cache ID 7348UxFm9]

34. Comstock J. Mobi Health News. 2017. Five Barriers to Wider Clinical Wearable Adoption URL: https://www.mobihealthnews.com/content/five-barriers-wider-clinical-wearable-adoption [accessed 2018-10-10] [WebCite Cache ID 7348OsZ2b]
35. 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]
36. Meingast M, Roosta T, Sastry S. Security and privacy issues with health care information technology. Conf Proc IEEE Eng Med Biol Soc 2006;1:5453-5458. [doi: 10.1109/IEMBS.2006.260060] [Medline: 17946702]
37. Ondiege B, Clarke M, Mapp G. Exploring a new security framework for remote patient monitoring devices. Comput 2017 Feb 24;6(1):11. [doi: 10.3390/computers6010011]
38. The US Department of Health and Human Services. 2015. The HIPAA Privacy Rule URL: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html [accessed 2019-08-08] [WebCite Cache ID 77WzcBHQb]
39. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
40. Blau M. The Boston Globe. 2017. Health Wearables Raise New Privacy Concerns URL: https://www.bostonglobe.com/business/2017/08/09/health-wearables-raise-new-privacy-concerns/0W2X9iNuLttoa7dUTFTzLJ/story.html [accessed 2019-08-08] [WebCite Cache ID 731RtFFc9]
41. Greenstone P. The Hill. 2018. HIPAA Guidelines Should Evolve With Wearable Technology URL: https://thehill.com/opinion/healthcare/378450-hipaa-guidelines-should-evolve-with-wearable-technology [accessed 2018-10-08] [WebCite Cache ID 731RQuras]
42. Banerjee S, Hemphill T, Longstreet P. Wearable devices and healthcare: data sharing and privacy. Inf Soc 2017 Dec 27;34(1):49-57. [doi: 10.1080/01972243.2017.1391912]
43. Nangalia V, Prytherch DR, Smith GB. Health technology assessment review: remote monitoring of vital signs--current status and future challenges. Crit Care 2010;14(5):233 [FREE Full text] [doi: 10.1186/cc9208] [Medline: 20875149]
44. Terry N, Wiley LF. SSRN. 2016. Liability for Mobile Health and Wearable Technologies URL: https://ssrn.com/abstract=2725450 [accessed 2019-08-08]
45. Wicklund E. mHealth Intelligence. 2018. Telehealth, mHealth Put The Spotlight on Interoperability at HIMSS18 URL: https://mhealthintelligence.com/news/telehealth-mhealth-put-the-spotlight-on-interoperability-at-himss18 [accessed 2018-10-12] [WebCite Cache ID 737OTq2MZ]
46. Clarke M, de Folter J, Verma V, Gokalp H. Interoperable end-to-end remote patient monitoring platform based on IEEE 11073 PHD and ZigBee health care profile. IEEE Trans Biomed Eng 2018 May;65(5):1014-1025. [doi: 10.1109/TBME.2017.2732501] [Medline: 28796600]
47. West Health. 2013. The Value of Medical Device Interoperability: Improving Patient Care With More Than $30 Billion In Annual Health Care Savings URL: https://www.westhealth.org/wp-content/uploads/2015/02/The-Value-of-Medical-Device-Interoperability.pdf [accessed 2018-10-10] [WebCite Cache ID 734JqOPzS]
48. Gay V, Leijdekkers P. Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability. J Med Internet Res 2015 Nov 18;17(11):e260 [FREE Full text] [doi: 10.2196/jmir.5094] [Medline: 26581920]
49. Mandl KD, Mandel JC, Kohane IS. Driving innovation in health systems through an apps-based information economy. Cell Syst 2015 Jul;1(1):8-13 [FREE Full text] [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]
50. Yao J, Schmitz R, Warren S. A wearable point-of-care system for home use that incorporates plug-and-play and wireless standards. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):363-371. [doi: 10.1109/TITB.2005.854507] [Medline: 16167690]
51. O'Connor S. Advanced Data Systems Corp. 2017. What Is Interoperability, and Why Is it Important? URL: https://www.adsc.com/blog/what-is-interoperability-and-why-is-it-important [accessed 2019-08-08]
52. Niksch A, Rothman B. HIMSS - Healthcare Information and Management Systems Society. 2014. The Value of Remote Patient Monitoring (RPM) Physicians' Perspectives URL: https://www.himss.org/value-remote-patient-monitoring-rpm-physicians-perspectives [accessed 2018-10-11] [WebCite Cache ID 736BtSQWB]
53. Kalid N, Zaidan AA, Zaidan BB, Salman OH, Hashim M, Muzammil H. Based real time remote health monitoring systems: a review on patients prioritization and related 'big data' using body sensors information and communication technology. J Med Syst 2017 Dec 29;42(2):30. [doi: 10.1007/s10916-017-0883-4] [Medline: 29288419]
54. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):pii: E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
55. Ramirez M, Maranon R, Fu J, Chon JS, Chen K, Mangione CM, et al. Primary care provider adherence to an alert for intensification of diabetes blood pressure medications before and after the addition of a 'chart closure' hard stop. J Am Med Inform Assoc 2018 Sep 1;25(9):1167-1174 [FREE Full text] [doi: 10.1093/jamia/ocy073] [Medline: 30060013]
56. Jacobson S, Wang T. Rock Health. 2015. The Shifting Center of Care URL: https://rockhealth.com/shifting-center-care [accessed 2018-10-10] [WebCite Cache ID 734KjUY4E]

57. Flach JM, Schanely P, Kuenneke L, Chidoro B, Mubaslat J, Howard B. Electronic health records and evidence-based practice: solving the little-data problem. Proc Int Symp Hum Factor Ergonom Health Care 2018 Jun 29;7(1):30-35. [doi: 10.1177/2327857918071002]
58. Apple. 2018. Apple Announces Effortless Solution Bringing Health Records to iPhone URL: https://www.apple.com/newsroom/2018/01/apple-announces-effortless-solution-bringing-health-records-to-iPhone/ [accessed 2019-08-08] [WebCite Cache ID 7348POfOi]
59. Cannon B. MedCity News. 2016. How Cedars Sinai & Dartmouth-Hitchcock Are Using Wearable Tech to Improve Patient Experience URL: https://medcitynews.com/2016/08/cedars-sinai-dartmouth-hitchcock-using-wearable-tech-improve-patient-experience/ [accessed 2018-10-10] [WebCite Cache ID 7348JJwJ2]
60. Epic UserWeb. 2018. URL: https://userweb.epic.com [accessed 2018-10-08] [WebCite Cache ID 731SgRaPq]
61. Siwicki B. Healthcare IT News. 2017. Biggest EHR Challenges for 2018: Security, Interoperability, Clinician Burnout URL: https://www.healthcareitnews.com/news/biggest-ehr-challenges-2018-security-interoperability-clinician-burnout [accessed 2018-10-12] [WebCite Cache ID 737NRLKfP]
62. Overlap Health. 2019. How It Works URL: https://www.overlaphealth.com/how-it-works/ [accessed 2019-05-25] [WebCite Cache ID 78copli6k]
63. O'Reilly K. Philips - United States. 2019. New York-Presbyterian Hospital Expands Virtual Care with Philips URL: https://www.philips.com/content/corporate/en_AA/about/news/archive/standard/news/press/2019/20190124-newyork-presbyterian-hospital-expands-virtual-care-with-philips.html/ [accessed 2019-08-08] [WebCite Cache ID 78coSxSeT]
64. Vivify Health®. 2018. URL: https://www.vivifyhealth.com [accessed 2018-10-10] [WebCite Cache ID 7347wu0kb]
65. Validic Delivers One Connection to a World of Mobile Health Data. 2018. URL: https://validic.com [accessed 2018-10-10] [WebCite Cache ID 7347pmzqs]
66. Doximity. 2018. Call Patients Privately, Without *67: Your Solution to Calling Patients on-the-go URL: https://www.doximity.com/clinicians/download/dialer [accessed 2018-10-10] [WebCite Cache ID 7345eHto3]
67. Xealth - Connecting Digital Health. 2018. URL: https://www.xealth.io [accessed 2018-10-10] [WebCite Cache ID 7344srFgt]
68. Redox | Simple, Scalable Healthcare Data Exchange. 2018. URL: https://www.redoxengine.com/ [accessed 2018-10-08] [WebCite Cache ID 731TW219a]
69. Conversa Health. 2018. URL: https://conversahealth.com/ [accessed 2018-10-08] [WebCite Cache ID 731TM6Odn]
70. Human API. 2018. URL: https://www.humanapi.co/ [accessed 2018-10-08] [WebCite Cache ID 731TFPh3Z]
71. Glooko | Diabetes Remote Monitoring | Population. 2019. URL: https://www.glooko.com/ [accessed 2019-05-25] [WebCite Cache ID 78cp14jR8]
72. Oscar | A new kind of health insurance company. 2018. URL: https://www.hioscar.com/app/ [accessed 2018-10-08] [WebCite Cache ID 731TAXWjO]
73. UnitedHealthcare. 2018. UnitedHealthcare Motion URL: https://www.uhc.com/employer/programs-tools/unitedhealthcare-motion [accessed 2018-10-08] [WebCite Cache ID 731T5HxpL]
74. Humana. 2018. Go365 Wellness & Rewards Program URL: https://www.humana.com/go365/ [accessed 2019-08-08]
75. John Hancock Insurance Home. 2018. URL: https://www.johnhancockinsurance.com/vitality-program.html [accessed 2018-10-08] [WebCite Cache ID 731SqoED4]
76. Pennic F. HIT Consultant. 2015. Novant Health To Integrate Epic MyChart with FitBit & HealthKit URL: https://hitconsultant.net/2015/08/21/novant-health-to-integrate-epic-mychart-with-fitbit-healthkit/ [accessed 2018-10-11] [WebCite Cache ID 736DS10JM]
77. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YF. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018;1:23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]

Abbreviations

AI: artificial intelligence
API: application programming interface
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act of 1996
IT: information technology
UI: user interface

Edited by G Eysenbach; submitted 19.11.18; peer-reviewed by V Mishra, C Fernández; comments to author 31.03.19; revised version received 26.05.19; accepted 21.07.19; published 11.09.19.

Please cite as:
Dinh-Le C, Chuang R, Chokshi S, Mann D
Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions
JMIR Mhealth Uhealth 2019;7(9):e12861
URL: https://mhealth.jmir.org/2019/9/e12861/
doi:10.2196/12861
PMID:31512582

Original Paper

Using Wearable Physiological Monitors With Suicidal Adolescent Inpatients: Feasibility and Acceptability Study

Evan Kleiman[1,2], PhD; Alexander J Millner[2,3], PhD; Victoria W Joyce[2], BA; Carol C Nash[2], MS; Ralph J Buonopane[2], PhD; Matthew K Nock[2,3], PhD

[1]Rutgers, The State University of New Jersey, Department of Psychology, Piscataway, NJ, United States

[2]Franciscan Hospital for Children, Brighton, MA, United States

[3]Harvard University, Department of Psychology, Cambridge, MA, United States

Corresponding Author:
Evan Kleiman, PhD
Rutgers, The State University of New Jersey
Department of Psychology
53 Avenue E
Room 627
Piscataway, NJ, 08854
United States
Phone: 1 8484452345
Email: evan.kleiman@rutgers.edu

Abstract

Background: Wearable physiological monitoring devices enable the continuous measurement of human behavior and psychophysiology in the real world. Although such monitors are promising, their availability does not guarantee that participants will continuously wear and interact with them, especially during times of psychological distress.

Objective: This study aimed to evaluate the feasibility and acceptability of using a wearable behavioral and physiological monitor, the Empatica E4, to continuously assess a group of suicidal adolescent inpatients.

Methods: Participants (n=50 adolescent inpatients) were asked to wear an Empatica E4 on their wrist for the duration of their inpatient stay. In addition to assessing behavioral metadata (eg, hours worn per day), we also used qualitative interviews and self-report measures to assess participants' experience of wearing the monitor.

Results: Results supported the feasibility and acceptability of this approach. Participants wore the monitor for an average of 18 hours a day and reported that despite sometimes finding the monitor uncomfortable, they did not mind wearing it. Many of the participants noted that the part of the study they enjoyed most was contributing to scientific understanding, especially if it could help people similar to them in the future.

Conclusions: These findings provide promising support for using wearable monitors in clinical samples in future studies, especially if participants are invested in being part of a research study.

(JMIR Mhealth Uhealth 2019;0(0):e0) doi:10.2196/13725

KEYWORDS

feasibility studies; wearable electronic devices; adolescent, hospitalized; self-injurious behavior; qualitative research

Introduction

Objective measures of physiological factors such as electrodermal activity (EDA) and heart rate (HR) have existed for over 100 years. Shortly after their development, researchers started using these tools in laboratory settings to examine the association between physiology and emotion (for historical reviews, see studies by AlGhatrif et al [1], Fye [2], and Neumann and Blanton [3]). Recent technological advances have enabled the study of human behavior and psychophysiology outside the laboratory, in the real world, using research-grade wrist-worn physiological monitors. These monitors enable the continuous, extended, real-world assessment of many of the same constructs once only possible to assess in laboratory settings over short assessment periods [4,5]. This offers great promise to improve our understanding, prediction, and prevention of factors related to psychological phenomena of interest. One area where technology may be particularly useful

is suicidal thoughts and behaviors, which are highly prevalent among adolescents (15% of all adolescents each year seriously consider suicide [6]) and present increased risk for suicide death, which account for 8.5% of all deaths among adolescents and young adults globally [7]. For example, through other technology-based studies (ie, studies using mobile phone app-based ecological momentary assessment), we know that suicidal thinking varies rapidly throughout the day and is associated with times of intense psychological distress (ie, high-arousal negative affect) [8]. We also know from laboratory studies that physiological signals such as EDA and HR map on to psychological distress [9-13]. We do not know, however, whether these EDA and HR are associated with suicidal thinking as it occurs in real time. Having this information would allow us to create interventions that can be trigged based on these physiological signals and delivered just in time as needed.

Public familiarity with wearables has increased in recent years with availability of consumer-grade monitors (eg, devices made by Fitbit and Apple). Unfortunately, these commercial monitors are limited for research use, given that they far are less accurate than gold-standard laboratory-based monitors [14-16] and often do not include sensors to measure important psychophysiological variables such as EDA. Compared with consumer-grade devices, scientific-grade wearable monitors are more accurate but also more expensive, bulkier, and less user-friendly than consumer-grade wearables. Thus, we cannot infer that high acceptability of consumer-grade wearables will translate to research-grade monitors. Accordingly, before such widescale research is possible, it is important to assess the feasibility and acceptability of using these monitors in real-world research. The goal of this study was to assess feasibility and acceptability of a commonly used research-grade wearable physiological monitor (Empatica E4, Empatica Srl) to continually assess behavior and psychophysiology among a clinically severe group of adolescents—those hospitalized for suicidal thoughts and behaviors—over the course of their hospital stay. The Empatica E4 is a physiological monitor that is worn on the wrist like a watch and records several streams of physiological data, including EDA and HR (through an optical sensor), as well as temperature and movement (through an accelerometer).

There are 3 important reasons to evaluate the feasibility and acceptability of these monitors within adolescent and psychiatric populations. First, previous studies examining the feasibility of wearable physiological monitors such as the E4 have been among adult samples that are relatively psychologically healthy, such as people who suffer from migraines [17] and tourists visiting a new city [18]. These studies cannot tell us whether adolescents (who may find new technology more acceptable or may be more self-conscious about the aesthetics of a wearable monitor) and those with more severe psychopathology (whose psychopathology may create competing demands for cognitive resources) would find these monitors acceptable. Second, nearly all studies have only collected data for a short period of time (eg, 10-20 min [19,20]) and cannot tell us whether participants would find it acceptable to use the monitor over far longer periods (eg, days, weeks, or months). Third, some newer wearable monitors are equipped with an event marker button that participants can use to report the experience of some psychological events/outcomes of interest. Such a feature allows researchers to examine physiological data leading up to (and following) events of interest. However, there has been no exploration of whether it is feasible to ask adolescent participants to press an event marker button during times of transient psychological events (eg, intense distress and severe suicidal thinking).

To our knowledge, only 1 study directly addressed feasibility/acceptability of the E4 [21]. This study found that a group of adults with schizophrenia and a control group were able to follow the instructions for using the E4 and rated it highly on a composite measure of acceptability. Although this information is useful, it leaves unaddressed several questions about the feasibility and acceptability of using such monitors with adolescents in acute psychiatric distress and over longer periods. Thus, here, we were interested in 3 questions relating to feasibility and acceptability that have not been addressed by previous studies.

First, we wanted to examine whether participants would wear the monitor the majority of each day over the course of multiple days. A monitor that can assess psychophysiology continuously throughout the day is only useful if participants are willing to wear the monitor over this period. Previous work has generally lasted only a few hours and therefore could not examine whether participants are willing to continuously wear the monitor over days or weeks. Second, we were interested in whether participants would interact with the monitor (ie, use the self-initiated button press). Simply collecting behavioral and psychophysiological data does not enable researchers to document (and predict) behavioral or cognitive outcomes (eg, psychological distress, suicidal thoughts, and hallucinations). However, such a button is only useful if participants remember to press the button when they should and do not press the button when they should not. One previous study presents some data on whether adolescent participants could reliably press an event marker on a wearable monitor (actigraphy watch [22]) each time they laid down to rest/sleep or got up from resting/sleeping. Although the authors of this study did not report actual compliance rates regarding the use of the event marker, they noted that “the event marker button was reliably used during the first few days but, afterward, some participants neglected to use it.” The experience of transient psychological events such as intense distress may be more (or less) memorable than times before and after rest, and thus, it is unknown whether participants will use the event marker at the same rate as this previous study. Third, we were interested in what participants liked (or disliked) about wearing the monitor. The only study to explicitly address participants’ opinions on the monitor [21] reported only that approximately 80% of their entire sample rated the monitor as *good* or *excellent* but did not report what participants specifically liked or disliked about the monitor.

Methods

Participants

Data were drawn from the first 50 participants from a larger, ongoing study of suicidal adolescent inpatients assessing the risk of harm to self or others using wearable ambulatory

monitoring. Participants were recruited from a large urban inpatient psychiatry unit. Inclusion criteria for the study were (1) admitted for risk of harm to self (eg, severe suicidal ideation, suicide attempt, and nonsuicidal self-injury), (2) being aged 12 to 19 years, and (3) having at least one wrist with unbroken skin where the wristband could be placed.

Recruitment and Data Collection Procedures

The study took place during participants' inpatient stay. Owing to hospital policy, we were not able to compensate participants.

Consent

We recruited and consented participants as close to hospital admission as possible. For potential participants aged younger than 18 years, we first approached parents/guardians to get their written consent and then approached the participant to get their written assent. We directly approached potential participants who were aged 18 years or older and received written consent. All study procedures were approved by the governing hospital and university institutional review boards.

Baseline Measures and Wearable Monitor Training

After providing informed consent, participants completed a brief set of self-report measures as part of the larger study but not relevant to this study and so not discussed further here (eg, measures of emotion regulation and impulsiveness). Next, participants completed a brief training session on how to properly wear and use the E4 and received a laminated 1-page information sheet with the same instructions to serve as a personal reference.

Monitoring Period

For the duration of the inpatient stay, we asked participants to wear an E4 on their dominant wrist as often as possible (eg, during the day and while sleeping) as long as the monitor was not at risk of getting wet (eg, during showers). The E4 has an event marker button that can be used to *tag* events defined by the research team. We asked participants to press the marker button on the E4 whenever they felt distressed, which we defined as "Feeling so upset or angry that you have an urge to hurt yourself or someone else or to break something." We made sure that participants were aware that no one was actively monitoring when they pressed the button (ie, pressing the button would not signal the clinical team to come help them, and we would not share the study information with the clinical team). Each day, a study staff member (during the workweek) or a clinical staff member (during the weekend) switched each participant's E4 for a fully charged monitor.

Daily Check-In Surveys

Each weekday, a study staff member approached the participant to conduct a brief check-in about any problems with the E4 that occurred since the last check-in. Staff members also assessed (1) whether the participant recalled missing any occasions when they believe they should have pressed the button but did not and (2) whether the participant accidentally pressed the button. If a participant recalled missing a button press or accidentally pressed the button, we assessed when and why this occurred.

Discharge

Shortly before their discharge from the hospital, participants completed 2 sets of open-ended questions aimed at assessing their experiences wearing the E4. First, they completed a 12-item questionnaire regarding satisfaction with the E4, modeled after other measures of comfort with wearable devices, specifically the Wearable Computer Comfort Rating Scale developed by Knight and Baber [23]. Items on this measure assessed (1) concerns about appearance when wearing the device (eg, "I felt anxious wearing the device"), (2) the physical feel of the device (eg, "The device was uncomfortable to wear"), (3) whether the device affected movement (eg, "the device made it hard to sleep at night"), and (4) general worries about taking care of the device (eg, "I worried about taking care of the device"). All items were on a 0 (never) to 10 (all the time) scale. Second, participants completed 4 open-ended qualitative questions: (1) "What did you like MOST about wearing the device?" (2) "What did you like LEAST about wearing the device?" (3) "How did you feel when wearing the device?" and (4) "Is there anything you would change about the device?" Due to reasons unrelated from the study (eg, discharge came quicker than expected), 3 out of the 50 participants were not able to complete the qualitative assessment.

The Wearable Monitor

Overview

The Empatica E4 (Empatica Srl) is a research-grade wrist-worn behavioral and psychophysiological monitor. Its case is 44 mm long (~1.73 inches), 40 mm wide, and 16 mm deep. This means that it is larger than commercially available wearable monitors (eg, the Fitbit Charge HR 2 is 22.86 mm long, 12.7 mm wide, and 11.0 mm deep). It has 4 main sensors: (1) a light-emitting diode–based photoplethysmograph (PPG) used to derive HR from blood flow, (2) a pair of silver-plated EDA/skin conductance sensors, (3) a 3-axis accelerometer, and (4) an infrared thermopile used to determine temperature. The E4 collects these data in real time and stores them on the onboard flash memory (which can hold ~60 hours of data, at 1 MB per hour). The E4 is then connected to a computer through a universal serial bus cradle and synchronized to a secure cloud server through the *E4 Connect* software. The E4 has a 250 mAh battery (lasting ~36 hours) that charges through the synchronizing cradle. The E4 also offers Bluetooth streaming, which can transmit data to the cloud using a compatible mobile phone as a gateway. We did not use this option because it would not have been feasible for long-term use owing to increased battery consumption as a result of using the Bluetooth radio and limitations in mobility (participants would need to be <30 feet from a mobile phone at any time).

Placement of E4 on Dominant Wrist

The larger goal of this study was to test whether the physiological and behavioral data collected from the E4 can accurately predict episodes of self-directed and other-directed violence. Thus, we made the decision about which wrist to wear the E4 based on our goal of collecting this data stream. There is currently a debate in the field about which wrist is optimal for assessing EDA. A body of work suggests that high arousal

negative emotion (eg, distress) can be detected through examining asymmetry between the left and right sides of the body [24], with more pronounced signals coming from the dominant side. Given that wearing 2 monitors at once would be cumbersome for participants, we elected to have participants wear the monitor on their dominant wrist, which would likely provide the most pronounced changes in EDA in response to distress. This is also in line with the manufacturer's recommendation [25]. It is important to note, however, that the dominant wrist likely produces noisier data (because of motion artifacts) than the nondominant wrist. Thus, researchers should examine the trade-off between signal and noise when choosing which wrist to use in their own work.

Logistics of Charging and Synchronizing E4

Each participant was assigned 2 wristbands but only wore 1 wristband at a time. While the participant wore 1 wristband, the other was placed on a dock to charge and synchronize data. We assigned participants 2 wristbands because charging and synchronizing the E4 can, in some cases, take more than an hour. Thus, having just 1 E4 per participant would have meant that we would lose more than an hour of data per day while the monitor was charging. As the E4 software can only synchronize 2 E4s at a time, it would have been cumbersome to have only 1 monitor per participant because it would have required study staff to rotate the E4 on the synchronizing cradle throughout the day. To accommodate a larger number of participants (we had up to 10 simultaneously), we constructed a charging and synchronizing station that we could place the monitors on when they were not being used, which was kept in a research office adjacent to the inpatient unit. The monitors would charge simultaneously and synchronize consecutively (ie, when 1 E4 was finished synchronizing, the computer moved on to the next one).

Analytic Strategy

How Often Will Participants Wear the Monitors?

To answer this question, we calculated 2 features from the E4's recording length metadata: (1) number of days that each participant wore the monitor for at least some amount of time and (2) number of hours during each day that participants wore the monitor. From these features, we calculated the intraclass correlation (ICC) of hours per day wearing the monitor. This allowed us to examine how much of the variability in time wearing the band was because of between-person differences (ie, whether some participants consistently tended to wear the band for more or less time than other participants) versus within-person differences (ie, whether participants wore the band a lot more or less frequently on some days than other days). We also compared these features by attrition status (ie, between those who did and did not drop out of the study).

Do Participants Correctly Use the Event Marker?

To answer this question, we calculated 3 sets of features: (1) number of button presses extracted from the raw data from the tags.csv file on the E4, (2) the number of missed button presses (and the reasons for missed press) extracted from the daily participant surveys, and (3) the number of accidental button presses (and reasons for the accidental press) extracted from the daily surveys. We also examined the ICC for the number of button presses each day, again to determine whether some patients tended to press the button more often than others or if patients pressed the buttons more on some days than others.

Do Participants Like Wearing the Monitor?

To answer this question with the quantitative data, we calculated the means and SDs for each item on the comfort assessment measure. We explored the association between each item on the measure and length of time wearing the band each day. We also compared responses on the measure between those who did and did not drop out of the study. Both sets of comparisons were corrected for multiple comparisons using a Bonferroni correction. We used a 3-step process to answer this question using qualitative data. First, the first author did a line-by-line read-through of the responses to the 4 qualitative questions and developed a codebook (described below). Second, the first, second, and last authors independently coded the responses based on this codebook. Third, we resolved any discrepancies to reach a final consensus. As with the quantitative data, we also examined associations between the qualitative data (ie, differences by whether participants did or did not endorse a qualitative category) and length of time wearing the monitor each day (using a *t* test) and in dropout status (using a chi-square test).

Results

Descriptive Statistics

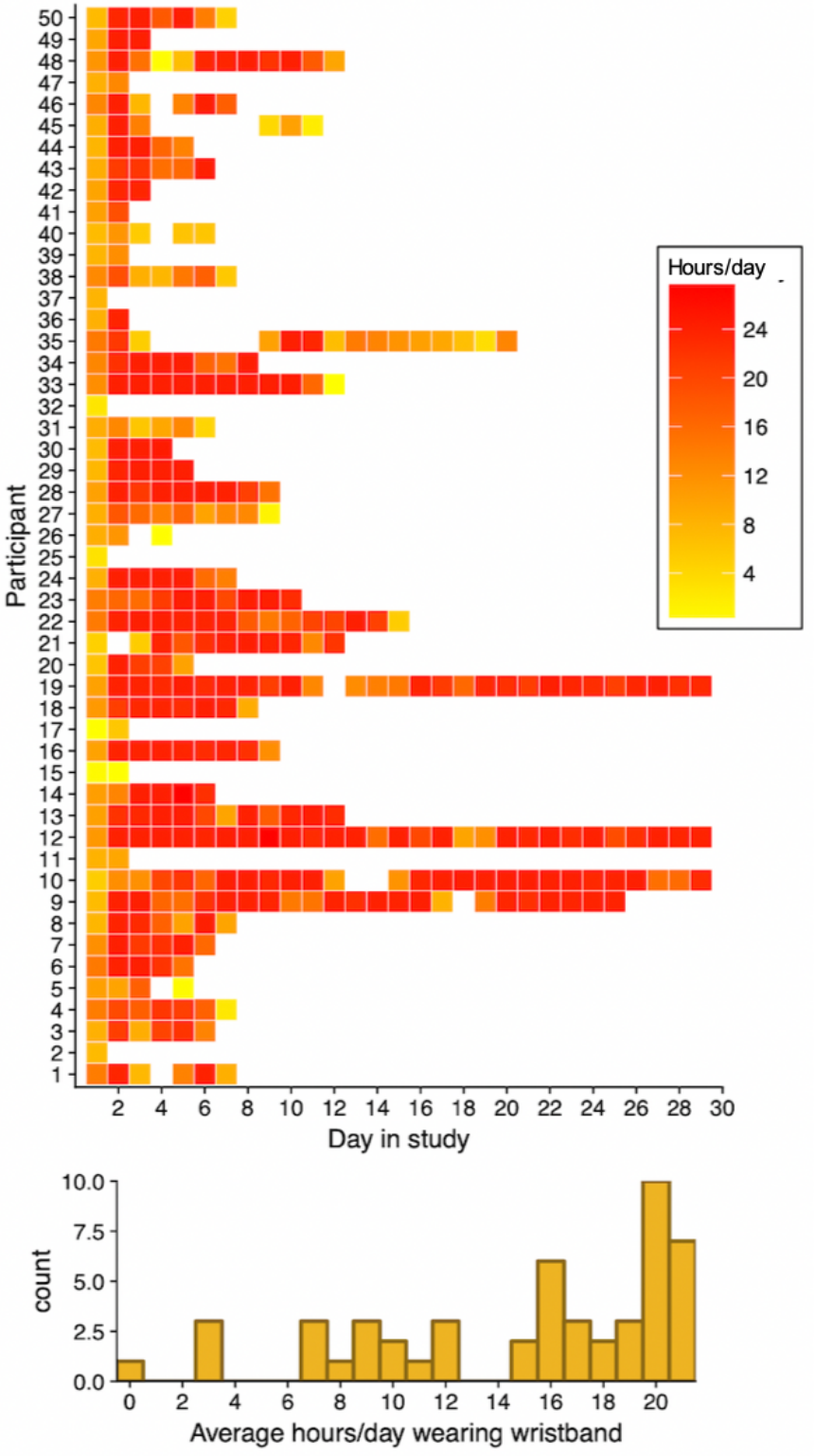
Participants' age ranged from 12.5 to 18.6 years (mean 16.3, SD 1.6), 78% (39/50) of the sample was female and 92% (46/50) of the sample was white (4%, 2/50 was Asian and the remainder indicated that they identified with another race). Participants together provided 487 total days of data (mean 9.74 days per participant, SD 13.81 days, range 1-76 days). The average length of stay in the hospital (ie, day of intake to day of discharge) was 10.7 days (SD 13.86 days, range 1-77 days). There were no demographic differences by age or sex (there were too few cases to examine racial differences) on any key study variables including hours wearing the monitor per day (age: r=0.05, P=.77 sex: $t_{11.12}$*(two-tailed)*=0.13, P=.90), number of button presses (age: r=0.10, P=.58; sex: $t_{23.94}$*(two-tailed)*=1.26, P=.22), and dropout status (age: $t_{5.81}$*(two-tailed)*=0.09, P=.93; sex: χ^2_1=0.8, P=.38).

Do Participants Wear the Monitor?

Participants wore the E4 at some point during the day (ie, any nonzero amount of data for the day) for 464 of a possible 487 total study days (95.3% of all days; see Figure 1, top panel). The majority (15/23; 65%) of days with no data were from 1 participant (#10 in the figures), although this participant was in the study for the longest amount of time (76 days). As can be seen in the bottom panel of Figure 1, there was considerable variability in the average number of hours each participant wore the band each day. On days when participants wore the E4, they did so for an average of 18.3 hours (SD 6.3). Excluding the first and last days of the study, when participants would not have been able to wear the monitor for 24 hours, participants wore the monitor for an average of 20.3 hours (SD 5.3). There was

no association between day in study and hours worn (slope [*b*]=−0.10; *P*=.46), suggesting that participants did not wear the monitor any more or any less as their time in the study increased. When visually inspecting the scatterplot of this association, there was no clear nonlinear effect, suggesting that participants did not likely wear the band more toward the middle of the study than at the end. Regarding variability in hours worn per day, there was more within-person (ie, day-to-day) variability in hours worn per day than there was between-person variability (ICC=0.31, 95% CI 0.21-0.43). This means that one-third of the variability was because of some participants tending to wear the E4 longer than others, whereas two-thirds of the variation in the amount of time wearing the monitor was accounted for by day-to-day variation within people.

Figure 1. Total daily hours worn each day (top) and histogram of average time worn each day per participant (bottom). White squares in top panel: band not worn that day. Participants marked in gray: dropouts. For clarity, range truncated to 1 to 30 in top panel (7 values [1.5% of all responses >30]). Count in bottom panel refers to number of participants.



Dropout

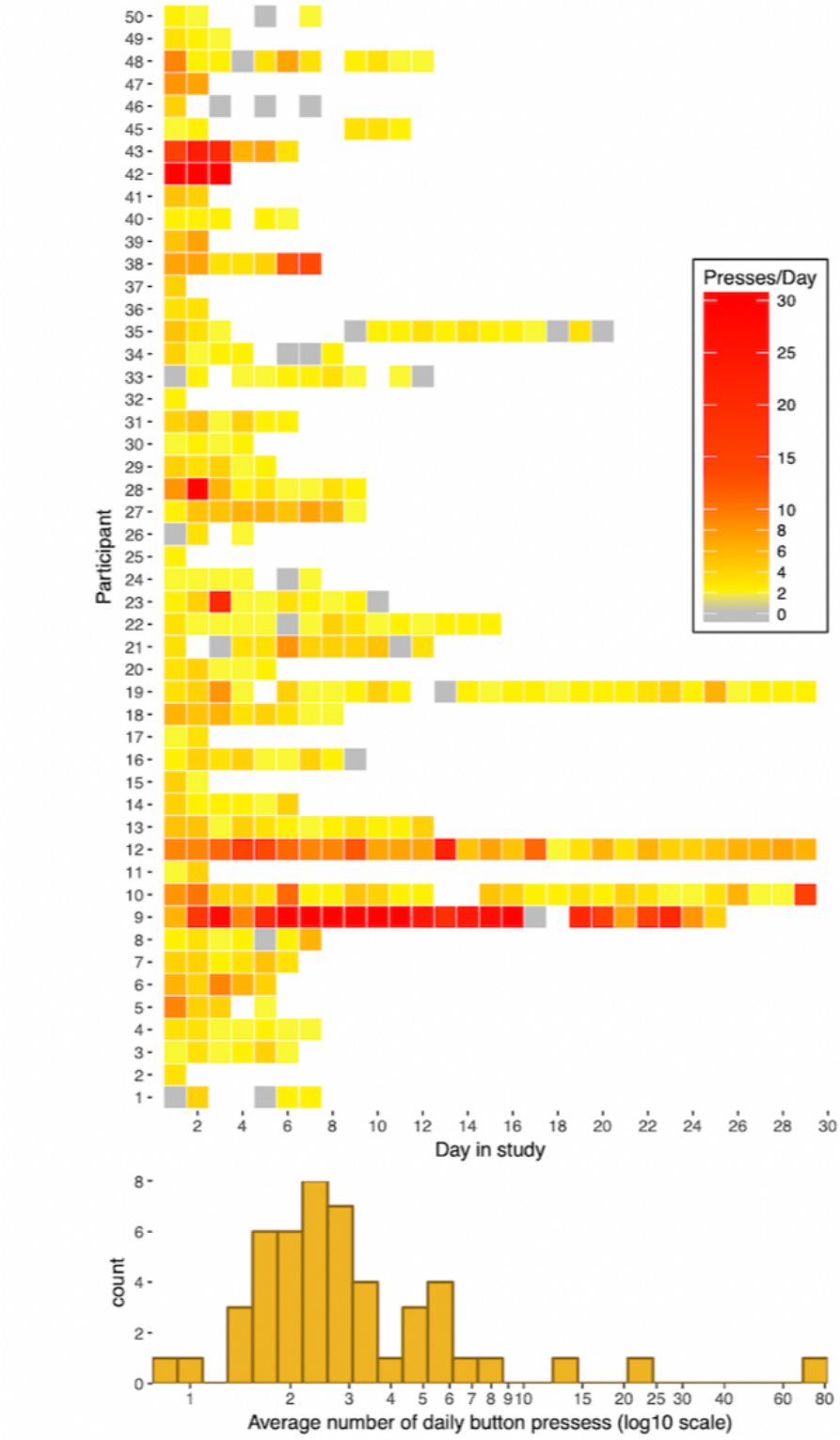
Overall, 7 participants discontinued wearing the monitor before they were discharged. As expected, those who dropped out of the study participated for significantly fewer days (mean 2.1, SD 1.4) than those who continued the study until they were about to be discharged from the hospital (mean 11.2, SD 14.7; $t_{45.66}$=3.92, *P*<.001) and for significantly fewer hours each day when they were in the study (mean 7.4, SD 6.3) than those who continued in the study (mean 18.7, SD 6.5; $t_{13.88}$=6.61, *P*<.001).

Do Participants Correctly Use the Event Marker?

There were 2159 button presses (ie, uses of the event marker) recorded during the study, which occurred during 435 of the 464 days (93.8%) during which participants wore the monitor. Participants pressed the button on average 4.9 times per day (SD 9.3, range 0-140). The top panel of Figure 2 shows a plot of the daily number of button presses for each participant during each day they were in the study.

As can be seen in the bottom panel of Figure 2, there was considerable variability in the average number of presses per participant per day ranging from 0.8 presses per day to 77.7 presses per day. In contrast to the amount of time wearing the monitor, there was more between-person variability in average number of button presses per day than there was within-person variability (ICC=0.63, 95% CI 0.52-0.73). This means that most of the variability in button presses was between-individuals, and each participant tended to stay near their own average throughout the study. As would be expected, when looking within each day regardless of subject, there was a small but significant association between hours worn per day and number of button presses (*b*=0.09, 95% CI 0.01-0.16; *P*=.04). When examining the association between day in study and number of times pressing the button, there was a small but statistically significant association *(b*=−0.10, 95% CI −0.15 to −0.05; *P*<.001), suggesting that as participants were in the study for a longer period of time, they pressed the button slightly less.

Figure 2. Total daily button presses (top) and histogram of average button presses per participant (bottom). White squares in top panel: band not worn that day. Participants marked in gray: dropouts. For clarity, range truncated to 1 to 30 days in top panel (7 values [1.5% of all responses >30]). Count in bottom panel refers to number of participants.



Missed Presses

There were 40 total missed button presses reported during the study. Of 40 missed presses, 17 came from 1 participant. The remaining 23 missed presses (57% of all presses) came from 14 participants (mean 1.7 presses per participant, SD 1.3, range 1-6). Regarding reasons for missed presses, nearly all the missed presses (32/40; 80% of presses) were because of the participant forgetting to press the button. The remainder of missed presses were because they were not wearing the monitor or the battery had died (4/40; 10% of presses) or because of an inability to pick a specific point when feeling distressed (2/40; 5% of presses) or misunderstanding that they were supposed to press the button (2/40; 5% of presses).

Accidental Presses

There were 10 accidental button presses reported during the study (0.46% of all presses). No participant had more than 1 accidental button press. The reasons participants accidentally pressed the button fell into 2 broad categories: (1) another patient pressed their button (6/10 presses) and (2) the result of trying to turn off the monitor by holding down the button, but accidentally releasing it too early (3/10). For the 1 remaining accidental press, the patient did not recall the circumstances.

Do Participants Like Wearing the Monitor?

Quantitative Measures

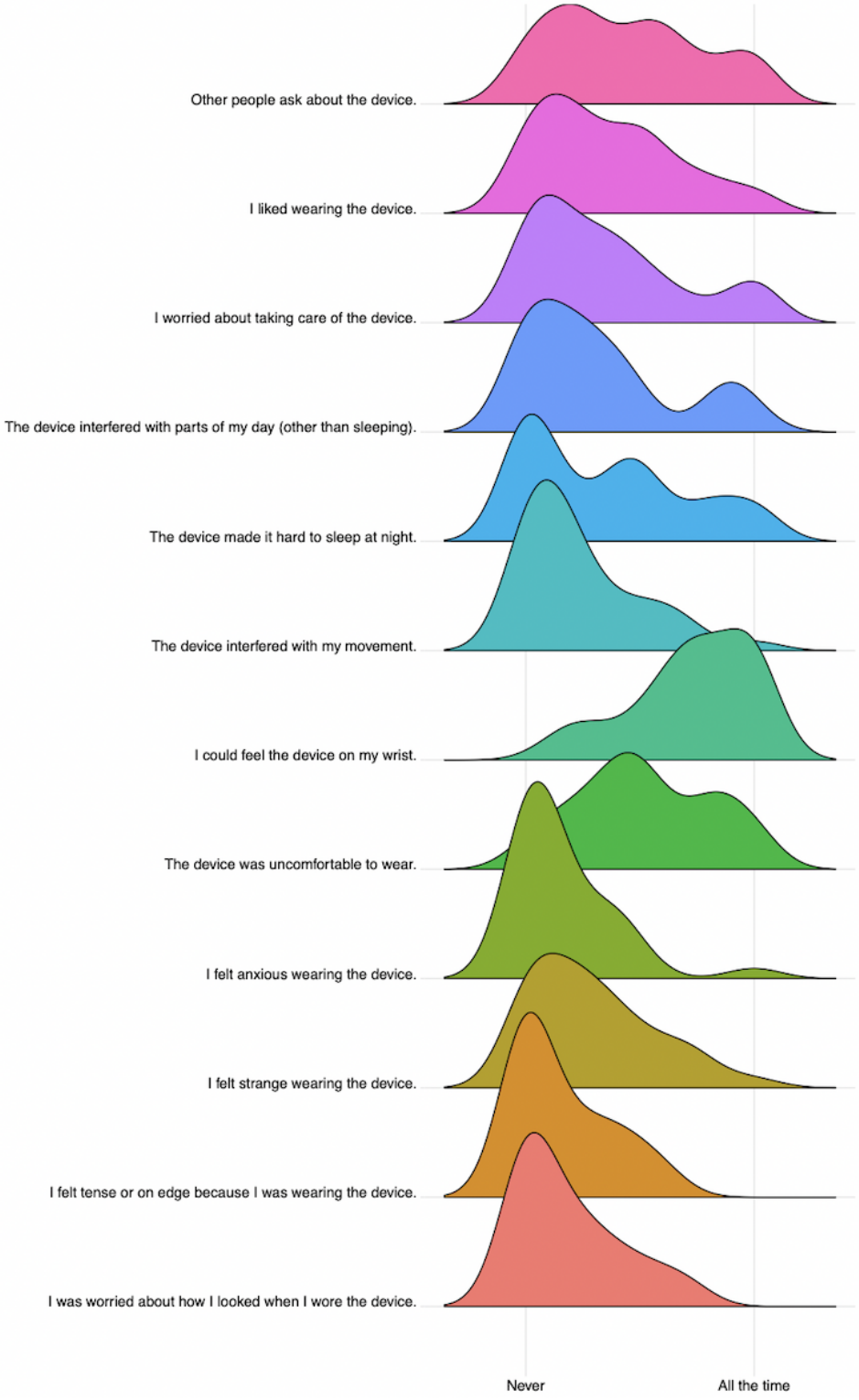
A summary of responses to the quantitative questions about device comfort is presented in Table 1, and a visualization of the distribution of responses is shown in Figure 3. On average, across nearly all items assessing comfort, participants tended to rate their discomfort wearing the monitor below 5 out of 10 (with 10 meaning more discomfort). As can be seen in the middle columns of Table 1, there was a significant negative correlation between hours worn and ratings of how uncomfortable the monitor was, such that the more participants rated the monitor as uncomfortable, the fewer hours they wore it. No other correlations were significant after correcting for multiple comparisons. As can be seen in the rightmost columns of Table 1, after correcting for multiple comparisons, there were only 2 significant findings: (1) those who dropped out were significantly more likely to rate the device as uncomfortable and (2) were more likely to note that they could feel the device. The qualitative data illustrated these quantitative findings well. One participant who dropped out said, “I really wanted to keep it on so I could help but it was too uncomfortable,” and another said, “I could feel the silver [EDA electrodes] rubbing.”

Table 1. Quantitative assessment of wearable monitor comfort.

Item	Descriptive, mean (SD)	Correlation with time worn		Attrition status			
		R value	*P* value	Did not drop out, mean (SD)	Dropped out, mean (SD)	Comparison *t* test *(df)*	*P* value
I was worried about how I looked when I wore the device.	1.87 (2.26)	−0.14	.45	1.93 (2.29)	1.33 (2.31)	0.42 *(2.44)*	.71
I felt tense or on edge because I was wearing the device.	1.59 (2.05)	−0.04	.82	1.66 (2.13)	1.00 (1.00)	0.94 *(4.24)*	.40
I felt strange wearing the device.	2.91 (2.67)	−0.19	.26	2.65 (2.35)	5.00 (4.40)	−1.05 *(3.22)*	.36
I felt anxious wearing the device.	1.63 (2.22)	0.07	.73	1.78 (2.29)	0.33 (0.58)	2.61 *(12.25)*	.02
The device was uncomfortable to wear.	5.23 (2.90)	−0.49[a]	.001[a]	4.53 (2.60)[a]	8.86 (1.07)[a]	−7.30[a]*(22.64)*	.001[a]
I could feel the device on my wrist.	7.28 (2.62)	−0.26	.08	6.87 (2.63)[a]	9.57 (0.79)[a]	−5.24[a]*(33.11)*	.001[a]
The device interfered with my movement.	2.35 (2.55)	−0.13	.47	2.04 (2.13)	3.83 (3.87)	−1.10.*(5.67)*	.31
The device made it hard to sleep at night.	3.53 (3.50)	−0.07	.68	3.50 (3.38)	3.67 (4.46)	−0.09.*(6.12)*	.93
The device interfered with parts of my day.	3.15 (3.20)	−0.31	.08	2.67 (2.95)	5.33 (3.67)	−1.66.*(6.51)*	.14
I worried about taking care of the device.	3.51 (3.28)	0.02	.90	3.59 (3.20)	3.00 (4.12)	0.31 *(4.73)*	.77
I liked wearing the device.	3.60 (2.96)	0.30	.05	3.66 (2.97)	3.29 (3.15)	0.29 *(8.10)*	.78
Other people ask about the device.	4.46 (3.27)	0.33	.04	4.69 (3.14)	2.80 (4.09)	1.00 *(4.70)*	.37

[a]Values are significant after Bonferroni correction (0.05/24=0.002).

Figure 3. Distribution of quantitative response to wearable comfort measure.



Qualitative Measures

The initial read-through of the qualitative responses yielded 13 codes across 3 categories: complaints about the monitor (7 codes), positive/neutral statements about the monitor (3 codes), and positive states about the study itself (3 codes). The results of the qualitative analyses are shown in the middle columns of Table 2. Our initial reliability across all codes was acceptable (kappa=0.77, SD 0.20, range 0.041-1.00). The 3 raters were able to come to a consensus on all the discrepancies. The rightmost column of Table 2 shows the frequency with which each code was endorsed and example statements from each code.

Table 2. Results of qualitative analyses (N=47).

Category	Example	κ[a] value	*Z* value	Endorsed, n (%)
Complaints about monitor				
Discomfort/uncomfortable in general	"It felt extremely uncomfortable."	0.68	8.08	22 (47)
The monitor was too big/clunky	"It was bulky and inconvenient."	0.9	10.74	17 (36)
It should have a clock	"I would probably make it have a clock."	0.92	10.89	11 (23)
The material was uncomfortable	"I'd make the wristband out of a thinner material."	0.41	4.83	7 (15)
Discomfort sleeping/at night	"Was uncomfortable during sleep."	0.92	10.93	5 (11)
EDA[b] sensors were uncomfortable	"Circular sensors are too big and rub too much."	1	11.87	3 (6)
The monitor does not look good	"It could be a little more sleek and comfortable."	0.59	6.95	1 (2)
Positive/neutral statements about device				
Felt OK when wearing monitor	"No different than normal."	0.47	5.62	35 (7)
Could tolerate negatives	"It isn't very comfortable but I managed."	0.58	6.84	5 (11)
Something positive about monitor's looks	"It looked cool."	0.76	9.05	5 (11)
Positive about paradigm/study				
Liked helping in a research study	"I felt I was helping."	0.86	10.17	28 (60)
Liked expressing distress	"I could press the button when I was in distress."	0.94	11.15	7 (15)
Helped become aware of distress	"I was able to be more alert and attentive to when I was having a hard time."	1	11.87	3 (6)

[a]Kappa from initial coding round.

[b]EDA: electrodermal activity.

The most commonly endorsed codes were "feeling OK when wearing the monitor" (74.5% of the sample) and "liked helping in a research study" (59.6%). Interestingly, although the study only asked participants to monitor and express their distress to the extent needed to remember to press the button, some participants reported that they liked the monitoring because it allowed them to express distress (14.89%) or become more aware of distress (6.38%). On balance, nearly half of the participants reported some discomfort while wearing the device (46.8%), with the most common complaint that the device was too large (36.2%). As can be seen in Table 3, there were no significant differences between those who did or did not endorse any of the qualitative categories on hours worn per day or attrition status.

Table 3. Differences in daily hours worn and attrition status by qualitative category endorsement (codes for "device does not look good" and "helped become aware of distress" were not included in these analyses because of low frequency of endorsement).

Category	Hours worn				Attrition			
	Endorsed, mean (SD)	Did not endorse, mean (SD)	*t* test[a]	*P* value	Did not drop out, n (%)	Dropped out, n (%)	χ^2 value *(df)*	*P* value
Discomfort/uncomfortable in general	14.05 (6.98)	15.96 (5.49)	0.88	.39	19 (45)	4 (57)	0.35 *(1)*	.55
Discomfort sleeping/at night	16.19 (5.25)	15.01 (6.32)	−0.42	.69	5 (12)	0 (0)	0.98 *(1)*	.32
EDA[b] sensors were uncomfortable	11.28 (9.55)	15.5 (5.83)	0.75	.52	2 (5)	1 (14)	0.86 *(1)*	.35
The monitor was too big/clunky	16.37 (4.73)	14.41 (6.85)	−1.00	.32	16 (37)	2 (29)	0.21 *(1)*	.65
The material was uncomfortable	11.03 (8.13)	15.83 (5.64)	1.27	.26	5 (12)	2 (29)	1.21 *(1)*	.27
It should have a clock	15.99 (5.40)	14.85 (6.46)	−0.52	.61	10 (22)	2 (29)	0.12 *(1)*	.73
Felt OK when wearing monitor	15.70 (5.49)	13.75 (7.70)	−0.73	.48	33 (78)	4 (57)	1.30 *(1)*	.25
Could tolerate negatives	17.90 (1.83)	14.79 (6.43)	−2.11	.05	5 (13)	0 (0)	0.98 *(1)*	.32
Something positive about monitor's looks	14.07 (4.40)	15.28 (6.38)	0.49	.64	4 (10)	1 (14)	0.12 *(1)*	.73
Liked expressing distress	12.13 (8.75)	15.64 (5.65)	0.87	.43	5 (13)	2 (29)	1.21 *(1)*	.27
Liked helping in a research study	16.32 (5.46)	13.14 (6.93)	−1.41	.17	25 (58)	5 (71)	0.48 *(1)*	.49

[a]All values have 1 degree of freedom *(df)*.

[b]EDA: electrodermal activity.

Discussion

Principal Findings

This study examined and found support for the feasibility and acceptability of using wearable behavioral and psychophysiological monitors to continuously collect objective data from adolescents with clinically severe psychiatric problems. There are several key findings from this study. First, participants were compliant with wearing the monitor, doing so on average more than 18 hours per day. Second, participants were compliant with instructions to use the event marker when distressed and were able to do so independently and without prompting. Third, participants found wearing the monitor to be acceptable and liked wearing them because it was part of a research study (eg, helping researchers understand psychiatric conditions that might someday be used to help others in a similar position). We discuss these main findings in greater detail below.

Do Participants Wear the Monitor?

The first aim of the study was to assess how often participants wore the monitor. We found that participants wore the E4 at some point nearly every day and did so, on average, more than 18 hours a day. We also found that the variability in hours worn per day has more to do with daily-level factors than with person-level factors. These data reflect that it may not be that there are certain types of participants who wear wristband more than others, but rather certain days where any given participant is more likely to wear the wristband than other days. Accordingly, future research could focus on reasons for this day-to-day variability to optimize participant compliance (eg, by finding factors that can identify the types of days where participants are more or less likely to wear the wristband). Overall, these data suggest that it is possible to use wearable monitors to collect continuous, objective data from clinically severe adolescents as they navigate their daily lives, and as such, to collect exponentially larger and more ecologically valid data than what has been possible in laboratory studies where recordings might be for a few hours at most while performing benign experimental tasks. This opens up myriad possibilities for better understanding the phenomenology and prediction of a range of clinical outcomes such as depression, anxiety, psychotic experiences, and suicidal and violent thoughts/behaviors.

Do Participants Correctly Use the Event Marker?

The second aim of the study was to assess whether participants interacted with the monitor. We asked participants to press the event marker on the monitor when they were feeling distressed and found that they were generally compliant in doing so. Participants rarely accidentally pressed the button (approximately 1 in 216 button presses was an accident) and rarely forgot to press the button (participants forgot to press the button approximately 1 out of every 59 times they should have). This is particularly impressive, given the circumstances under which we asked participants to press the button. Being on an inpatient unit can be unfamiliar and stressful and being highly distressed (which is when they were asked to press the button) is an affect state, which may be particularly difficult to self-monitor and respond to. Indeed, the basis for this larger study these data are drawn from is to attempt to identify distress using the E4's sensors so that interventions could be developed that do not rely on patients needing to monitor for distress. Regarding the frequency of the button presses each day, we found that although there was day-to-day variability in the times each participant pressed the button each day, there was more variability from participant to participant, suggesting that some participants tended to use the event marker more than others. It is unclear whether this is best explained by between-participant variability in conscientiousness in pressing the button versus in the likelihood of actually experiencing

distress. Future work examining the correspondence between button presses and actual physiological arousal (ie, as an objective indicator of distress) will help to address this question. We also found that as the number of days in the study increased, the number of times pressing the button each day slightly decreased. This could be a marker of an effective treatment (ie, participants were feeling less distressed) or a marker of study fatigue. Given how infrequently participants did not press the button when they should have, it is probably more likely that this is a marker of treatment efficacy rather than study fatigue.

Do Participants Like Wearing the Monitor?

The third aim of the study was to assess both quantitively and qualitatively what participants liked and did not like about participating in a study using wearable physiological monitors. We found that participants tended to report that the monitor did not interfere much with their movement and that they *felt OK* when wearing the monitor. Many participants reported that one of the aspects they liked the most about the study was being in a study where they could contribute to knowledge acquisition and/or help others similar to them. On balance, nearly half of the sample noted that the monitor was generally uncomfortable, with more than one-third of participants noting the monitor was too bulky. It is also notable that more than 15% of participants said that they liked how the monitor helped them be more aware of their distress, although this was not an explicit goal of the study. This is interesting because it may suggest the viability of interventions using these monitors to help participants self-monitor and manage their own distress in a more explicit manner.

Taken together, these findings suggest that doing research with these wearable monitors is feasible, especially when participants are motivated to be in a study. These findings also suggest, however, that the feasibility of using wearable physiological monitors for research may not translate to applied settings where the incentive of being in a research study is not there to motivate compliance. In these cases, there would be fewer incentives to balance out the negatives about the size of the monitor. Thus, more comfortable monitors may be required to obtain similar compliance to what we found in this study. For example, the Empatica Embrace is a currently available consumer monitor that is approximately 30% smaller than the E4 (30×30×10 mm and 44×40×16 mm, respectively). Although not explicitly designed for research use, it can be used in some research applications (eg, studies concerned with movement and skin conductance, but not HR because it does not have a PPG sensor). Moreover, it is reasonable to expect that smaller research-grade monitors will be available from a variety of manufacturers in the near future.

Limitations, Strengths, and Future Directions

These findings should be interpreted in the context of 3 key limitations. First, although we assessed the *amount* of data received, we did not assess the *quality* of data received. Not all data collected from any wearable monitor are usable (eg, because of motion artifacts). Evaluating the quality of the data obtained in such studies is an important step for future research, and this is something that we will be undertaking in our ongoing work in this area. Second, we used recording length as a proxy for time wearing the E4. If the E4 is turned on, it records even if it is not being worn. Thus, it is possible that the data reported here on time that the monitor was worn may be overestimated. There are 2 reasons why the probability of time worn overestimations is quite low. One is that participants were told to turn off the E4 whenever it was not being worn. Another is that data were collected on a psychiatric inpatient unit where unit staff observes patients at least once every 5 min. If unit staff saw an unworn E4 that was powered on, they would turn off the monitor. A third limitation, related to the previous point, is that these data were collected on an inpatient unit where participants were continually monitored by clinical staff and visited nearly daily by research staff, possibly making them more likely to wear the monitors. Thus, findings from this study may not generalize to other settings where there is less intensive adult supervision (eg, schools and home) or to other samples (eg, adults). Examining the generality of these results in other settings and samples is another important future step for research in this area. Finally, although our study provided a rich description of how often participants wore the monitor and interacted with it, the study was unable in some cases to provide explanations of why participants wore a device more on some days than others. Similar to many studies whose goal is description, future studies should explore possible explanations for the phenomena described in this study.

In conclusion, the clearest implication from this study is that it is feasible to conduct research where participants wear physiological monitors for an extended period (ie, days or weeks). This implication is in line with other studies of samples of adults with less severe psychopathology [21]. The study's findings are important because studies that use wearable behavioral and psychophysiological monitors over an extended period have great promise to help researchers understand how constructs of interest to psychological scientists operate in everyday life. This is especially true for studies that combine these data streams with other streams of data such as medical records, passive mobile phone sensing, and ecological momentary assessment. Our ability to conduct studies similar to this is only just beginning, and it is undoubtable that wearable technology will become even more advanced in the coming years, making studies even more feasible in the future.

Acknowledgments

Research reported in this publication was supported by the National Institute of Mental Health (R21MH115293; EMK) and the Chet and Will Griswold Suicide Prevention Fund (MKN).

Conflicts of Interest

None declared.

References

1. AlGhatrif M, Lindsay J. A brief review: history to understand fundamentals of electrocardiography. J Community Hosp Intern Med Perspect 2012;2(1) [FREE Full text] [doi: 10.3402/jchimp.v2i1.14383] [Medline: 23882360]
2. Fye WB. A history of the origin, evolution, and impact of electrocardiography. Am J Cardiol 1994 May 15;73(13):937-949. [doi: 10.1016/0002-9149(94)90135-x] [Medline: 8184849]
3. Neumann E, Blanton R. The early history of electrodermal research. Psychophysiology 1970 Jan;6(4):453-475. [doi: 10.1111/j.1469-8986.1970.tb01755.x] [Medline: 4907439]
4. Dolgin E. Technology: dressed to detect. Nature 2014 Jul 9;511(7508):S16-S17. [doi: 10.1038/511S16a]
5. Poh MZ, Swenson NC, Picard RW. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans Biomed Eng 2010 May;57(5):1243-1252. [doi: 10.1109/TBME.2009.2038487] [Medline: 20172811]
6. Cash SJ, Bridge JA. Epidemiology of youth suicide and suicidal behavior. Curr Opin Pediatr 2009 Oct;21(5):613-619 [FREE Full text] [doi: 10.1097/MOP.0b013e32833063e1] [Medline: 19644372]
7. Cha CB, Franz PJ, Guzmán EM, Glenn CR, Kleiman EM, Nock MK. Annual research review: suicide among youth - epidemiology, (potential) etiology, and treatment. J Child Psychol Psychiatry 2018 Apr;59(4):460-482 [FREE Full text] [doi: 10.1111/jcpp.12831] [Medline: 29090457]
8. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J Abnorm Psychol 2017 Aug;126(6):726-738. [doi: 10.1037/abn0000273] [Medline: 28481571]
9. Lensvelt-Mulders G, Hettema J. Genetic analysis of autonomic reactivity to psychologically stressful situations. Biol Psychol 2001 Sep;58(1):25-40. [doi: 10.1016/S0301-0511(01)00099-0] [Medline: 11473793]
10. Reinhardt T, Schmahl C, Wüst S, Bohus M. Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the mannheim multicomponent stress test (MMST). Psychiatry Res 2012 Jun 30;198(1):106-111. [doi: 10.1016/j.psychres.2011.12.009] [Medline: 22397919]
11. Lazarus RS, Speisman JC, Mordkoff AM, Davison LA. A laboratory study of psychological stress produced by a motion picture film. Psychol Monogr 1962;76(34):1-35. [doi: 10.1037/h0093861]
12. Garralda ME, Connell J, Taylor DC. Psychophysiological anomalies in children with emotional and conduct disorders. Psychol Med 1991 Nov;21(4):947-957. [doi: 10.1017/S0033291700029937] [Medline: 1780407]
13. Nock MK, Mendes WB. Physiological arousal, distress tolerance, and social problem-solving deficits among adolescent self-injurers. J Consult Clin Psychol 2008 Feb;76(1):28-38. [doi: 10.1037/0022-006X.76.1.28] [Medline: 18229980]
14. 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]
15. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep 2015 Aug 1;38(8):1323-1330 [FREE Full text] [doi: 10.5665/sleep.4918] [Medline: 26118555]
16. Shoval N, Schvimer Y, Tamir M. Real-time measurement of tourists' objective and subjective emotions in time and space. J Travel Res 2017 Feb 16;57(1):3-16. [doi: 10.1177/0047287517691155]
17. Siirtola P, Koskimäki H, Mönttinen H, Röning J. Using sleep time data from wearable sensors for early detection of migraine attacks. Sensors (Basel) 2018 Apr 28;18(5):pii: E1374 [FREE Full text] [doi: 10.3390/s18051374] [Medline: 29710791]
18. Shoval N, Schvimer Y, Tamir M. Tracking technologies and urban analysis: adding the emotional dimension. Cities 2018 Feb;72(4):34-42 [FREE Full text] [doi: 10.1016/j.cities.2017.08.005]
19. Corino VD, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. Physiol Meas 2017 May;38(5):787-799. [doi: 10.1088/1361-6579/aa5dd7] [Medline: 28151434]
20. Spann CA, Schaeffer J, Siemens G. Expanding the Scope of Learning Analytics Data: Preliminary Findings on Attention and Self-Regulation Using Wearable Technology. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017 Presented at: LAK'17; March 13-17, 2017; Vancouver, British Columbia, Canada p. 203-207. [doi: 10.1145/3027385.3027427]
21. Cella M, Okruszek Ł, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2018 May;195:537-542 [FREE Full text] [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
22. Huynh C, Guilé JM, Breton JJ, Desrosiers L, Cohen D, Godbout R. Is it possible to study sleep-wake patterns in adolescent borderline personality disorder? An actigraphic feasibility study. Int J Adolesc Med Health 2010;22(4):547-560. [doi: 10.1515/IJAMH.2010.22.4.547] [Medline: 21404885]
23. Knight JF, Baber C. A tool to assess the comfort of wearable computers. Hum Factors 2005;47(1):77-91. [doi: 10.1518/0018720053653875] [Medline: 15960088]

24. Picard RW, Fedor S, Ayzenberg Y. Multiple arousal theory and daily-life electrodermal activity asymmetry. Emotion Review 2015 Mar 2;8(1):62-75. [doi: 10.1177/1754073914565517]
25. Empatica Support. 2019. What Should I Know to Use EDA Data in My Experiment? URL: http://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment- [accessed 2019-02-15] [WebCite Cache ID 76CtiiyIb]

Abbreviations

EDA: electrodermal activity
HR: heart rate
ICC: intraclass correlation
PPG: photoplethysmograph

Edited by C Dias; submitted 15.02.19; peer-reviewed by C Carrion, D Levine, V Agyapong, M DeJonckheere, R Shalaby, K Rosen, G Myreteg; comments to author 11.04.19; revised version received 22.06.19; accepted 13.07.19; published 24.09.19.

Please cite as:
Kleiman E, Millner AJ, Joyce VW, Nash CC, Buonopane RJ, Nock MK
Using Wearable Physiological Monitors With Suicidal Adolescent Inpatients: Feasibility and Acceptability Study
JMIR Mhealth Uhealth 2019;0(0):e0
URL: http://mhealth.jmir.org/2019/0/e0/
doi:10.2196/13725
PMID:31586364

Corrigenda and Addenda

Figure Correction: Digital Pain Drawings Can Improve Doctors' Understanding of Acute Pain Patients: Survey and Pain Drawing Analysis

Nour Shaballout[1], MD; Anas Aloumar[1], MD; Till-Ansgar Neubert[1]; Martin Dusch[2*], Dr Med; Florian Beissner[1*], Dr Phil Nat

[1]Somatosensory and Autonomic Therapy Research, Institute for Diagnostic and Interventional Neuroradiology, Hannover Medical School, Hannover, Germany

[2]Section Pain Medicine, Clinic of Anaesthesiology and Intensive Care Medicine, Hannover Medical School, Hannover, Germany

[*]these authors contributed equally

Corresponding Author:
Florian Beissner, Dr Phil Nat
Somatosensory and Autonomic Therapy Research
Institute for Diagnostic and Interventional Neuroradiology
Hannover Medical School
Carl-Neuberg-Strasse 1
Hannover, 30625
Germany
Phone: 49 511 53508413
Email: Beissner.Florian@mh-hannover.de

Related Article:

Correction of: https://mhealth.jmir.org/2019/1/e11412/

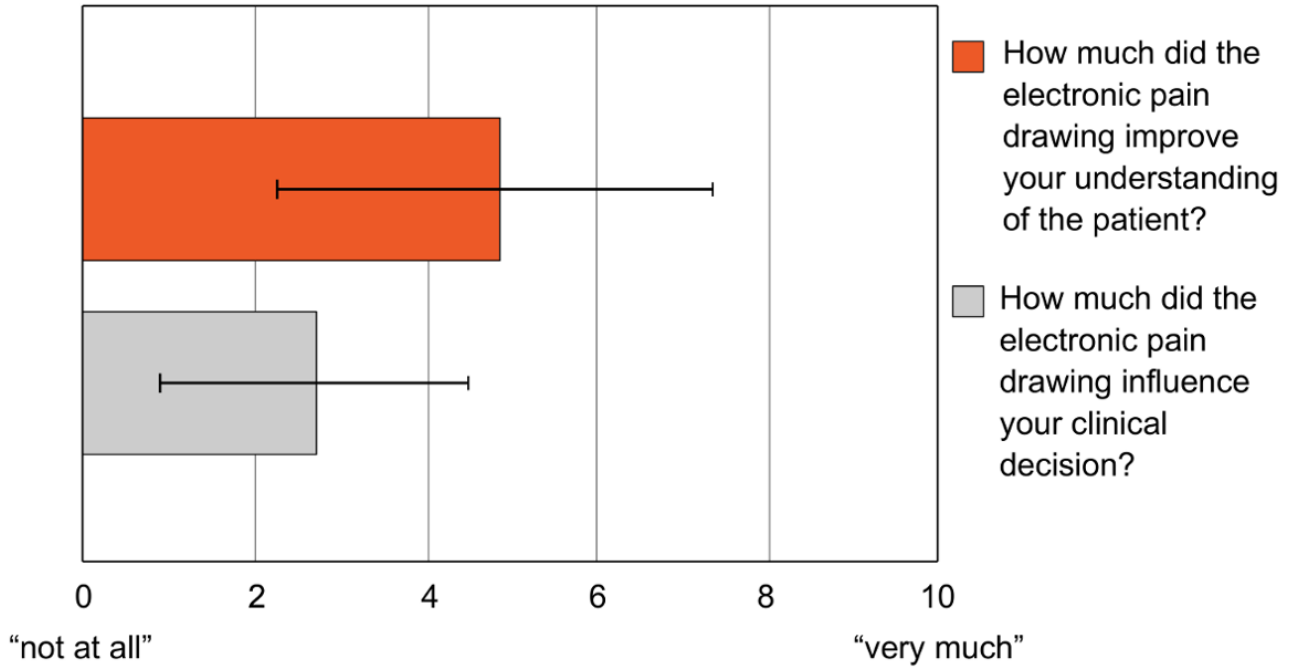
(JMIR Mhealth Uhealth 2019;7(9):e16017) doi:10.2196/16017

In "Digital Pain Drawings Can Improve Doctors' Understanding of Acute Pain Patients: Survey and Pain Drawing Analysis" by Shaballout et al (JMIR Mhealth Uhealth 2019;7(1):e11412), the authors inadvertently reversed the legends in the bar plot of Figure 2.

A revised version of Figure 2 has been uploaded with the correct legend wherein the upper (red) bar is denoted by "How much did the electronic pain drawing improve your understanding of the patient?" and the lower (grey) bar is denoted by "How much did the electronic pain drawing influence your clinical decision?"

The correction will appear in the online version of the paper on the JMIR website on September 27, 2019, 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.

Figure 2. Impact of knowing patients' pain drawings (PDs) on understanding of the pain and clinical decision making as rated by the doctors. Patients' PDs significantly improved the doctors' understanding of the pain and to a lesser but still significant extent influenced their clinical decision.



Submitted 27.08.19; this is a non–peer-reviewed article;accepted 29.08.19; published 27.09.19.

Please cite as:
Shaballout N, Aloumar A, Neubert TA, Dusch M, Beissner F
Figure Correction: Digital Pain Drawings Can Improve Doctors' Understanding of Acute Pain Patients: Survey and Pain Drawing Analysis
JMIR Mhealth Uhealth 2019;7(9):e16017
URL: http://mhealth.jmir.org/2019/9/e16017/
doi:10.2196/16017
PMID:31573932

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org