JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 1 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Reviews

Original Papers

Corrigenda and Addenda

Review

Unraveling Mobile Health Exercise Interventions for Adults: Scoping Review on the Implementations and Designs of Persuasive Strategies

Karlijn Sporrel[1], MSc; Nicky Nibbeling[2], MSc, PhD; Shihan Wang[3,4], MSc, PhD; Dick Ettema[1], IR, PhD, Prof Dr; Monique Simons[5], MSc, PhD

[1]Faculty of Geosciences, Utrecht University, Utrecht, Netherlands

[2]Department of Applied Psychology, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[3]Institute of Informatics, University of Amsterdam, Amsterdam, Netherlands

[4]Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

[5]Social Sciences, Consumption and Healthy Lifestyles, Wageningen University & Research, Wageningen, Netherlands

Corresponding Author:
Karlijn Sporrel, MSc
Faculty of Geosciences
Utrecht University
PO Box 80115
Princetonlaan 8a
Utrecht, 3584 CB
Netherlands
Phone: 31 642514287
Email: k.sporrel@uu.nl

Abstract

Background: It is unclear why some physical activity (PA) mobile health (mHealth) interventions successfully promote PA whereas others do not. One possible explanation is the variety in PA mHealth interventions—not only do interventions differ in the selection of persuasive strategies but also the design and implementation of persuasive strategies can vary. However, limited studies have examined the different designs and technical implementations of strategies or explored if they indeed influenced the effectiveness of the intervention.

Objective: This scoping review sets out to explore the different technical implementations and design characteristics of common and likely most effective persuasive strategies, namely, goal setting, monitoring, reminders, rewards, sharing, and social comparison. Furthermore, this review aims to explore whether previous mHealth studies examined the influence of the different design characteristics and technical operationalizations of common persuasive strategies on the effectiveness of the intervention to persuade the user to engage in PA.

Methods: An unsystematic snowball and gray literature search was performed to identify the literature that evaluated the persuasive strategies in experimental trials (eg, randomized controlled trial, pre-post test). Studies were included if they targeted adults, if they were (partly) delivered by a mobile system, if they reported PA outcomes, if they used an experimental trial, and when they specifically compared the effect of different designs or implementations of persuasive strategies. The study methods, implementations, and designs of persuasive strategies, and the study results were systematically extracted from the literature by the reviewers.

Results: A total of 29 experimental trials were identified. We found a heterogeneity in how the strategies are being implemented and designed. Moreover, the findings indicated that the implementation and design of the strategy has an influence on the effectiveness of the PA intervention. For instance, the effectiveness of rewarding was shown to vary between types of rewards; rewarding goal achievement seems to be more effective than rewarding each step taken. Furthermore, studies comparing different ways of goal setting suggested that assigning a goal to users might appear to be more effective than letting the user set their own goal, similar to using adaptively tailored goals as opposed to static generic goals. This study further demonstrates that only a few studies have examined the influence of different technical implementations on PA behavior.

Conclusions: The different implementations and designs of persuasive strategies in mHealth interventions should be critically considered when developing such interventions and before drawing conclusions on the effectiveness of the strategy as a whole. Future efforts are needed to examine which implementations and designs are most effective to improve the translation of theory-based persuasive strategies into practical delivery forms.

(JMIR Mhealth Uhealth 2021;9(1):e16282) doi:10.2196/16282

KEYWORDS

mobile health; physical activity; goals; feedback; rewards; reminder systems; social support; adult

Introduction

Physical activity (PA) mobile health (mHealth) interventions, such as interventions delivered by wearable technologies, SMS messages, and mobile apps, have potential for supporting PA behavior [1-5]. Yet, while some PA mHealth interventions successfully increase PA, others do not [2,6,7], and it is unclear why this is the case. A possible explanation for this discrepancy in effectiveness is the various ways in which the persuasive strategies are being incorporated in PA mHealth interventions [8].

Persuasive strategies (or behavior change techniques [8]) are theoretically underpinned elements of interventions, such as goal setting or rewards, intended to foster a positive behavior or attitude change toward PA. Over the last decade, several taxonomies of persuasive strategies have been developed [8-12], including a taxonomy specifically for PA and dietary interventions (ie, the CALO-RE taxonomy [9]). These taxonomies allow for a clear and consistent description of interventions [13], and they have been frequently adopted for designing and evaluating interventions for behavior change [14-18].

Although the persuasive strategies from the taxonomies are commonly used to inform the study design, they do not contain a guideline to operationalize the strategies [19-21]. Consequently, the same persuasive strategy can be shaped differently in different exercise interventions [20,22]. For instance, the technical implementation of the strategy cue or prompt [9] can be delivered via a mobile phone as a text or sound message, by means of a flashing light or even as a vibration. However, it can also be delivered via an email or an actual phone call. Furthermore, the design characteristics of the message can also differ; messages can be framed differently (eg, positive or negatively framed) [23,24], and the messages can be short or long and generic or tailored [23,24]. As a result of these diverse implementations and designs, interventions might evoke different user responses, even though they use the same persuasive strategy. This renders it difficult to draw a conclusion about the effectiveness of the persuasive strategy at the theoretical level.

Several studies have argued that design characteristics influence the effectiveness of the strategy, such as the use of different social media features [25] or the content of messages [23,24]. However, the technical implementation of these strategies has received little attention so far (eg, the device used or the accessibility of the strategy). This is surprising, as technical implementation can influence user experience, usability, and intervention exposure, which, in turn, likely influences the effect of the intervention [18,26,27]. Thus far, only one review has examined whether both design characteristics and technical implementations could impact the effectiveness of digital exercise or dietary interventions on the persuasive strategy *feedback* [22]. A great variety in implementation forms of feedback was found, for instance, regarding the accessibility of feedback (continuous access or daily messages) and the form of feedback (visual or not). Moreover, the findings indicated that not all types of feedback were equally effective in changing PA behavior [22].

Schembre et al [22] limited their study to feedback; however, other strategies likely face the same diversity in design characteristics and technical implementations. Therefore, this scoping review sets out to explore the designs and implementations of other promising persuasive strategies in mHealth PA interventions for adults and explores whether previous mHealth studies examined the influence of the different design characteristics and technical operationalizations of common persuasive strategies on the effectiveness of the intervention to persuade the user to engage in PA. As analyzing all strategies is beyond the scope of this study, the analysis is limited to the most common and evidence-based strategies for PA [7,18], namely, monitoring, goal setting, reminders, rewards, and 2 social strategies (sharing and social comparison).

Methods

Approach

A nonsystematic literature search was performed to identify original research papers that examined the selected persuasive strategies in the context of PA mHealth interventions. A nonsystematic search was deemed appropriate because the objective of this study is to gain insights into the various operationalizations of strategies and their influence on the effectiveness of the strategy and *not* to provide a complete overview of current literature or a list of effective implementations and characteristics of the strategies. The PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) criteria were used to guide the reporting of the methods and findings (Multimedia Appendix 1) [28]. The protocol for this review was not registered.

Search Strategy and Study Selection

Most papers were identified by checking the references of (recent) reviews in the same field and in the author's personal libraries. As the research team is multidisciplinary, papers from the fields of behavior change, computer science, and

gamification were included. When interesting papers were identified (either by previous reviews or in the library of authors), extensive snowball searches were performed (ie, the references of all interesting papers were checked for other relevant papers). We extended the snowball search with additional quick searches in Google Scholar when the initial search resulted in a small number of studies for a specific operationalization of a persuasive strategy (eg, sharing information on social media). The search included terms that refer to mHealth (eg, mobile devices, PA apps), PA behavior (eg, exercise, walking, running), the persuasive strategy (eg, sharing, social media, Facebook, Twitter), and adults. These additional searches increased the number of identified papers to a limited extent.

Studies were included if they (1) were (partly) delivered by a mobile system (eg, pedometer, SMS, mobile app), (2) reported PA outcomes (self-reported or objectively measured), (3) used an experimental trial (eg, randomized controlled trial, factorial design, pretest-posttest design), (4) examined at least one of the selected persuasive strategies, and (5) described this strategy in sufficient detail. Finally, (6) studies were only included when they specifically compared the effect of different designs or implementations of persuasive strategies. Therefore, papers were excluded if they only examined the effectiveness of the intervention as a whole (with multiple strategies). Studies were also excluded if they targeted children or individuals with a chronic disease (eg, patients with cardiovascular diseases, mental disorders, etc), as these target groups may have different needs [26]. There were no restrictions on publication year, sample size, or study duration.

Data Charting

To systematically analyze the included studies, a data charting list was developed in multiple review rounds with input from all coauthors (consistent with the guidelines for writing a scoping review [29]). The final data chart comprised 3 sections, namely (1) study characteristics, (2) technical implementations and design characteristics of persuasive strategies, and (3) study results (Multimedia Appendix 2). The study characteristics included information on the methodology of the study and other factors that can influence study outcomes (ie, characteristics of the participants and contextual factors [30]).

The *technical implementations and design characteristics* section was inspired by the Behavioral Intervention Technology (BIT) model [20]. The BIT model is grounded in 3 well-respected design models [11,31-33] and includes principles from both behavioral theories and technological features. The BIT delivery elements (subdivided into delivery systems and elements) and BIT workflow informed the *technical implementation* category (Multimedia Appendix 2). The BIT characteristics of the elements informed the *design characteristic* category. Most design characteristics were inferred from previous reviews [22-25] and theories of behavior change [33-35], as the BIT framework does not provide a detailed list of design characteristics. Furthermore, during the data extraction period, the chart was updated when new implementations or design characteristics were identified.

The final data chart section covered the *study results*. We examined whether there was a higher amount of PA compared with a control group (without the strategy) and compared with another intervention arm (with a different operationalization of the strategy, see Multimedia Appendix 2). The PA outcome measurement used was the step count of the participants, unless the paper did not measure this. In that case, the main PA outcome measurement was used instead, as described in this study. The outcomes were classified as *positive* when the PA outcome measurement of the study was significantly more effective, *neutral* when no effect was found, and *negative* when the implementation resulted in worse PA outcomes.

The general study characteristics were extracted for the study as a whole. The type of persuasive strategy and its design and technical implementation were extracted separately for each intervention arm because this differed between arms in the same study. One researcher (KS) performed the data extraction of all the included papers. To ensure that the data extraction was performed correctly for all persuasive strategies (in line with Levac et al [29]), the second reviewer (SW) performed data extraction of at least 20% of each of the persuasive strategies (22.6% on average, SD 0.9%). The interrater reliability was high (93.5% agreement).

Results

Overview

The search yielded 29 original research papers (85 intervention arms) [36-64]. An overview of the study characteristics can be found in Multimedia Appendix 3 [36-64]. The results of the individual persuasive strategies are presented below. For each strategy, first, a short description or definition is provided. Second, a summary of the identified implementations and designs in the intervention arms is provided (a complete overview can be found in Multimedia Appendix 3). Third, an overview is given of which implementations have proven to be an effective *addition* in an intervention. Finally, the findings of studies that *compared* designs and implementations of strategies are provided.

Persuasive Strategy: Monitoring

Monitoring involves keeping track of your behavior or behavior outcomes [9]. Traditionally, users had to actively track their own PA behaviors (eg, by means of questionnaires and self-logging); however, nowadays, PA tracking can also be performed passively by using mobile devices, without posing a burden to the user. Almost all of the included studies used monitoring of behavior (28 studies and 82 intervention arms), apart from [54].

Design and Technical Implementations

With regard to the technical implementation, 7 different delivery systems were identified that enabled monitoring (for instance, mobile apps [n=46] and SMS functions [n=4]). Most interventions used a combination of 2 systems (mean 1.68). The elements that were used were either related to active self-monitoring (ie, data entry field [n=42]) or passive recordings of behaviors (n=77; eg, accelerometer [n=77], GPS [n=2]). Although most studies used passive tracking, users often had

to track their behavior themselves as well. With respect to the design characteristics, many different behavior types were monitored (n=8) and 4 specific monitoring characteristics were identified (such as the option to correct automatically logged data). More details on the various implementations and design characteristics can be found in Multimedia Appendix 3.

Effect of Including Monitoring to the Intervention

None of the included studies examined whether adding monitoring to the intervention increased the effectiveness of the intervention.

Comparison of Designs and Technical Implementation of Monitoring

One study compared self-logging alone with self-logging in combination with automatic tracking (ie, wearing a second tracking device) regarding its effect on PA. No difference was found in the effectiveness of the 2 technical implementations [41].

Persuasive Strategy: Goal Setting

Goal setting is a strategy in which the individual either sets a goal or gets a goal assigned. CALO-RE distinguishes between unspecific behavioral goals, behavioral outcome goals, and action plans [9] (also referred to as implementation intentions [65]). It is a commonly used strategy [18], as reflected by the presence of the strategy in most of the included papers (n=26). Not all studies described goal setting in sufficient detail [41,42,61], and these were, therefore, excluded from this part of the analysis. The results of the remaining 23 studies (66 intervention arms) are listed below.

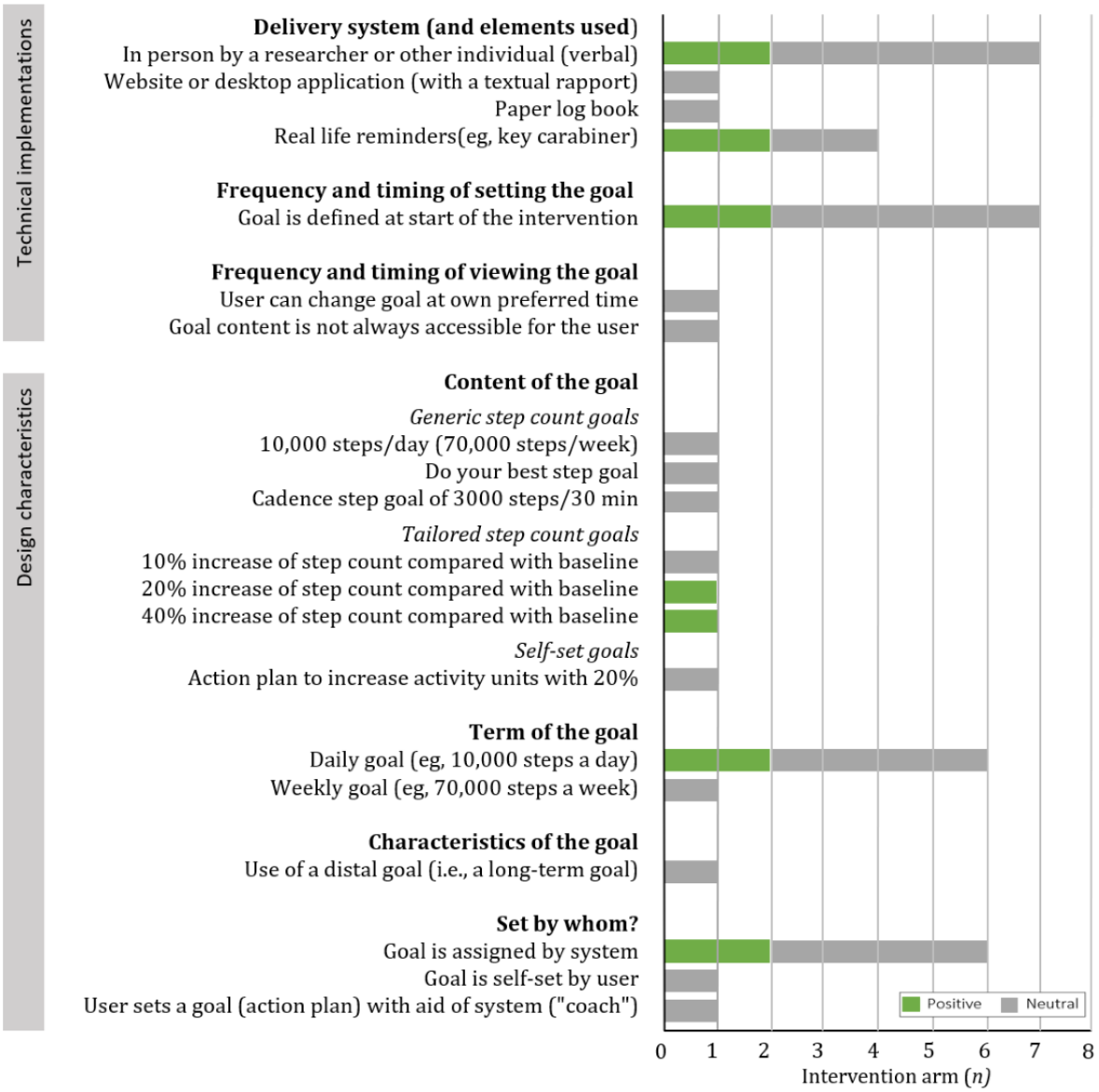
Design and Technical Implementations

Regarding the technical implementation, 8 different delivery systems were used to deliver goal setting, of which the researchers themselves were the most common deliverers (n=27). Frequently used elements were reports (n=19) and textual notifications (n=21). Mostly, the goal is only set at the offset of the intervention (n=60), but some systems also changed the goal on a daily (n=3) or weekly (n=6) basis. We found a great variety in the design principles for goal setting; in total, 29 different goal types were identified in the 66 intervention arms. The goal types differed for goal difficulty (eg, 7000 or 10,000 steps) and targeted behavior (eg, step count or floor count; Multimedia Appendix 3). Most of the goals were assigned to the user (n=44); however, in some interventions, the user was instructed to set her own goal (n=16) or the user could choose a goal from a list of suggestions (n=9). Furthermore, 12 specific design characteristics of goal setting were identified, such as tailoring by the system (n=20) and using metaphoric goals (eg, Climb the Eiffel Tower [39]; n=6).

Effect of Including Goal Setting to the Intervention

In total, 3 studies (7 arms) examined whether including goal setting improved the effectiveness of the intervention. In 2 of the 7 arms, goal setting appeared to be an effective addition [48,50], whereas in the remaining 5, no effect of including goal setting was found [48,50,58]. As can be seen in Figure 1, there is no clear trend in the data regarding effective technical implementations. For the design characteristics of goal setting, it appears that *tailored goals* [50] are generally effective, whereas generic goals do not seem to increase PA behavior [48,58]. However, tailored easy goals (eg, 10% increase compared with baseline) did not result in more PA compared with a group without goal setting [50] (notably, this study lasted for only 1 week). Self-set moderate-to-vigorous PA (MVPA) goals, even with a coaching system for developing action plans, did not result in long-term (48 weeks) PA change compared with a group without goal setting [58]. Thus, these results suggest that some operationalizations of goals increase the effectiveness of the intervention but not all.

Figure 1. The effectiveness of the technical implementations and design characteristics of goal setting compared with receiving no goals in 7 intervention arms (3 different studies).



Comparison of Designs and Technical Implementations of Goal Setting

With 8 studies comparing design characteristics of goal setting [36,37,43,46,48,50,55,57,62], goal setting is one of the most extensively examined strategies. However, no studies have compared various technical implementations of goal setting. The results demonstrate that an effective design for long-term behavior change (4 months) is the use of automatically adaptive and tailored goals compared with a 10,000 steps per day goal [36,62]. Interestingly, a static goal seems to be more effective at the initiation phase of the intervention [36]. A second efficient design is the use of more *difficult* (eg, 40% increase compared with the baseline level) tailored step goals compared with *easier* (eg, 10% increase) tailored step goals [37,50]. However, there is some discrepancy in the right difficulty level; one study found that a step count increase of 20% and 40% of steps was better than a 10% increase [50], whereas another study found that only a 100% increase was better than a 10% increase [37].

A third design that seems effective is using a list of *context-aware* activity goals compared with a list of general not location-specific goals [55]. The context-aware goals were based on previously logged activities and frequent locations of the user to generate location-based goals that were tailored to the individual's previous behaviors. Finally, the effectiveness of different goal characteristics was examined in one study [46]. Although the differences between the goals did not reach significance, there was a trend demonstrating that participants performed most steps per week if they chose the recommended step goal (mean 42.195), followed by the metaphoric goal (mean 35.462), whereas the self-set goal resulted in the least number of steps (mean 31.774). In contrast, most individuals choose to

set a manual goal, which suggests that they prefer this goal setting type.

Persuasive Strategy: Reminders for PA

Reminders are cues, prompts, or triggers that *push* the user to perform a certain behavior, such as a notification or an email. Only reminders that were used to remind the user to engage in PA were considered. Reminders intended to remind the user to set a goal or to wear the device were not taken into account [40,50,59]. Reminders are used relatively frequently, almost half of the included studies incorporated this strategy (n=13, implemented in 34 intervention arms).

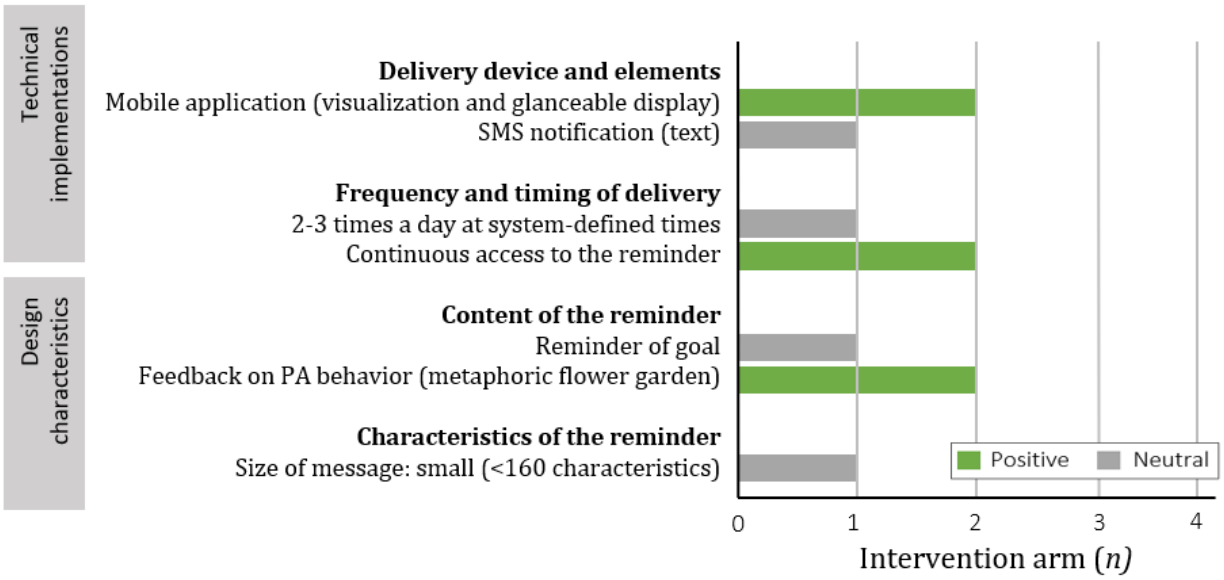
Design and Technical Implementations

Regarding technical implementations, 6 different delivery systems were identified (eg, mobile apps [n=13] and email [n=4]) and 6 different delivery elements, such as text notifications (n=24) and visualizations (n=5). The most notable differences in the technical implementations concern the frequency and the timing in which reminders were provided to the users. For instance, in some studies, the user received more than 15 messages per day [42], whereas in other interventions, the user received only one message a week [43] (Multimedia Appendix 3). In contrast to the other investigated strategies, the design characteristics of reminders were often not described in detail. For instance, the framing [24,66], tailoring [23], and size of the messages were seldom reported. Moreover, in some studies, the content of the reminder was not described (n=4).

Effect of Including Reminders to the Intervention

Only 2 studies (2 intervention arms) examined the effectiveness of receiving a reminder compared with not receiving a reminder, of which one found a positive effect [41] and one found no effect [61]. The design and implementation of the reminders used in these studies and its effectiveness on PA promotion are shown in Figure 2. A study that found a positive effect used a glanceable display, which is a constant reminder that resides on the background of the phone while simultaneously providing the user with information on his or her activity level [41]. The second study demonstrated that 3 SMS notifications a day reminding the user of his or her goal were not effective for increasing PA, at least, not for longer than 1 week [61]. At the end of the intervention, various participants reported that they stopped reading the messages, as the messages were impersonal and automated.

Figure 2. The effectiveness of the technical implementations and design characteristics of reminders compared with receiving no reminders in 3 intervention arms (2 different studies). PA: physical activity.



Comparison of Designs and Technical Implementations of Reminders

One study examined the effect of the timing of reminders (ie, technical implementation) on the effectiveness of the intervention [42]. The study results were conflicting. On the one hand, individuals increased their steps more if they received reminders at random times (about 10 a day) compared with receiving context-aware reminders (eg, just after eating or after prolonged sedentary time; no significance values were reported) [42]. On the other hand, the participants liked and accepted the context-aware reminders better than random reminders. The authors argue that these conflicting findings might be the result of the short duration of the study and a few participants (n=19), which increases the likelihood of factors (weather and busy calendar) influencing the results. A second study examined whether the content of reminders influences its effectiveness (ie, the design characteristics) [54]. No differences in step count were found between receiving reminders of the participants' action plans or of their general goal.

Persuasive Strategy: Rewards

Rewards are reinforcers of behavior that can be given for attempts to reach a goal and for reaching the goal [9]. In line

with the CALO-RE taxonomy [9], only rewards that incentivize performing PA behaviors were considered and not rewards that incentivize study participation. A total of 12 studies (29 different intervention arms) met these criteria.

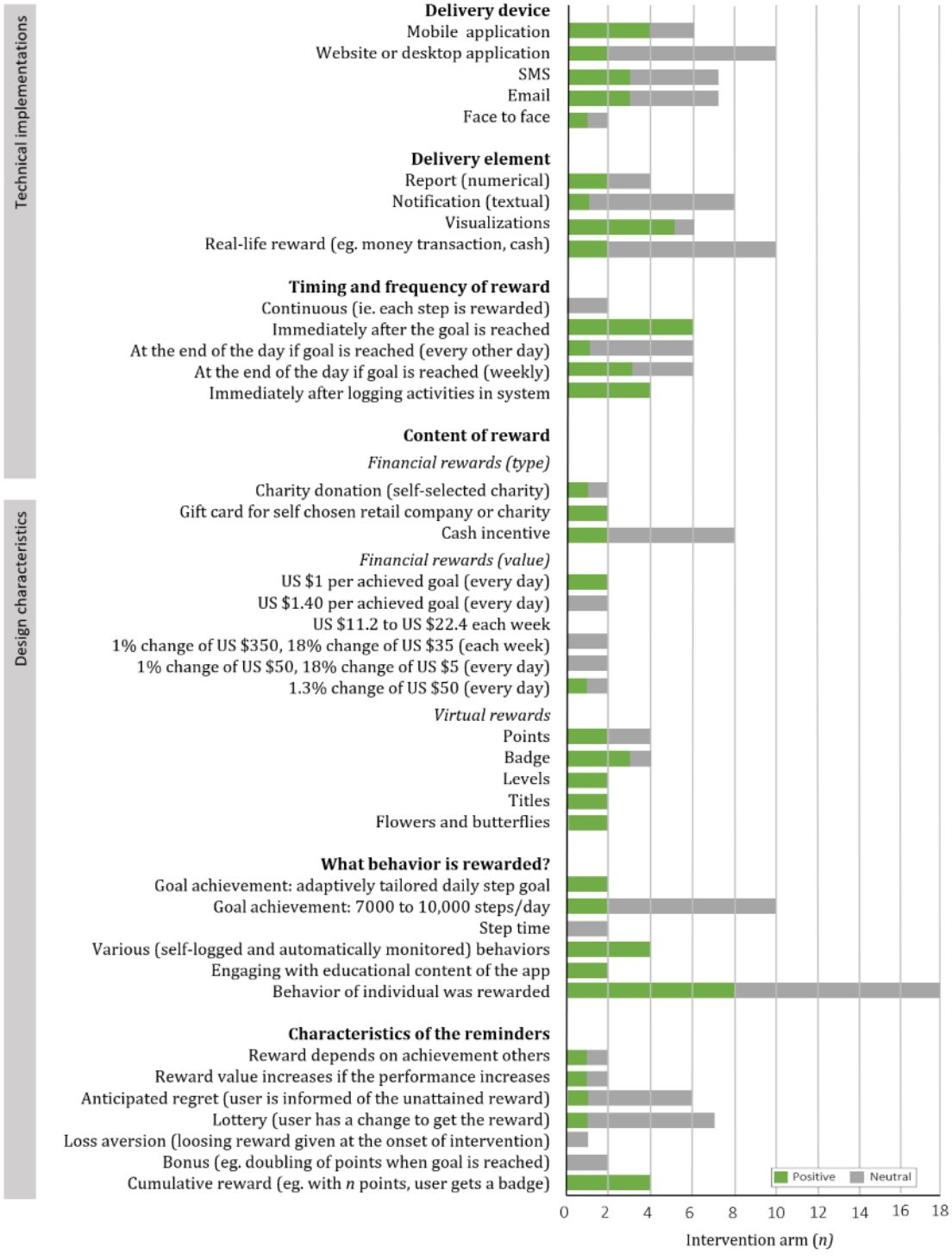
Design and Technical Implementations

Regarding technical implementations, 6 different delivery systems were used, such as websites (n=14). Furthermore, 7 different elements were identified, the most frequent element being visualizations (n=13). Most participants received rewards either immediately after they achieved their goal (n=14) or with a short delay (n=10). In a few studies, participants were rewarded for performing PA (eg, each step taken) and not necessarily for reaching their goal (n=6). The design characteristics varied greatly between interventions (Multimedia Appendix 3). For instance, we identified 16 different reward contents (eg, points, US $1 per achieved goal [n=2]), 9 different behaviors that were rewarded (eg, achieving 7000 steps per day [n=8]) and 7 reward characteristics (eg, receiving the reward depended on somebody else [n=8]).

Effect of Including Rewards to the Intervention

Of the identified studies (18 intervention arms), 8 examined whether including rewards increases the effectiveness of the intervention [36,41,43,44,51-53,63]. Of the 18 arms, 8 arms demonstrated that adding a reward resulted in more PA than the same intervention without rewards [36,41,43,51], whereas the remaining 10 did not have a significant effect on PA [43,51-53,63]. Some technical implementations of rewards appear to be effective additions to interventions (Figure 3), such as the use of visualizations of the rewards [41,43] and receiving the reward immediately after the goal is attained [36,41,43]. Implementations that do not seem to result in more PA are rewarding each step (ie, efforts toward reaching a goal) [63]. Furthermore, interventions with the design characteristic of *cumulative rewards* (eg, with enough points the user receives a badge) appear to be effective in motivating individuals to engage in PA [36,41,43,51]. Taken together, it seems that some operationalizations of rewards are effective additions to an intervention, whereas others are not.

Figure 3. The effectiveness of the technical implementations and design characteristics of rewards compared with receiving no rewards in 18 intervention arms (8 different studies). PA: physical activity.



Comparison of Designs and Technical Implementations of Rewards

A total of 3 studies compared different design characteristics of (financial) rewards [44,51,52]. Finkelstein et al [44] demonstrated that adding a cash reward on top of virtual rewards (from Fitbit) increases step count, and they also showed that cash rewards are equally effective in changing the stepping behavior as charity rewards. Notably, the MVPA minutes (main outcome of the study) was higher in the group that received a cash reward [44]. Furthermore, when the financial reward was no longer offered (after 12 weeks), the step count declined. Other studies found no significant differences between rewarding the behavior of the individual, the behavior of a team, or a combination of both [51] and between receiving a financial reward, a loss aversion reward, or a lottery-based reward [52].

Persuasive Strategy: Sharing

Sharing is a social strategy in which users can actively offer and receive social support from others. It can be provided digitally (eg, sharing on Facebook) or in a real-life setting (eg, group meetings [48]). In total, 9 studies (15 intervention arms) were identified that included this strategy in their intervention.

Design and Technical Implementations

Regarding the technical implementation of sharing, 4 different delivery systems were used: mobile apps (n=8) and face-to-face delivery (n=3). In total, 8 different delivery elements were identified, such as reports and messaging functions (Multimedia Appendix 3). Various design characteristics were also identified. For instance, several content types could be shared, such as PA data (n=5) and competition results (n=2). In addition, the relationships between individuals who shared information with each other differed between studies. Sometimes, the users shared their information with strangers (n=5), whereas in other interventions, individuals shared their information with acquaintances (n=4) or acquaintances and strangers (n=2).

Effect of Including Sharing to the Intervention

In total, 4 studies (4 intervention arms) examined whether adding sharing strategies to the intervention increased its effectiveness [40,49,56,60]. One study found a positive effect [56], and the other 3 found no effect of sharing on PA [40,49,60] (Figure 4). The interventions differed in devices that were used for sharing. In a study with positive results, individuals (frequent Facebook users) used Facebook [56], whereas studies with no effect used websites or a mobile app (specifically developed for the particular study) [40,49,60]. A Facebook-delivered intervention is likely to be effective because it is well integrated into the individual's life. In contrast, using an additional intervention website might pose a barrier for the user, which can explain why they were neither effective nor frequently used [49,60].

Figure 4. The effectiveness of the technical implementations and design characteristics of sharing compared with receiving no sharing function in 4 intervention arms (4 different studies). PA: physical activity.

Comparison of Designs and Technical Implementations of Sharing

As none of the included studies compared different designs or technical implementations of sharing to examine which implementation or design is more effective in increasing PA, no conclusions can be drawn on effective operationalization of sharing to increase PA behavior.

Persuasive Strategy: Social Comparison

A second social strategy is facilitating *social comparison* [9], which includes competition, collaboration, and social norm information. In total, 9 of the included studies enabled social comparisons (17 intervention arms and 11 did not include social comparison).

Design and Technical Implementations

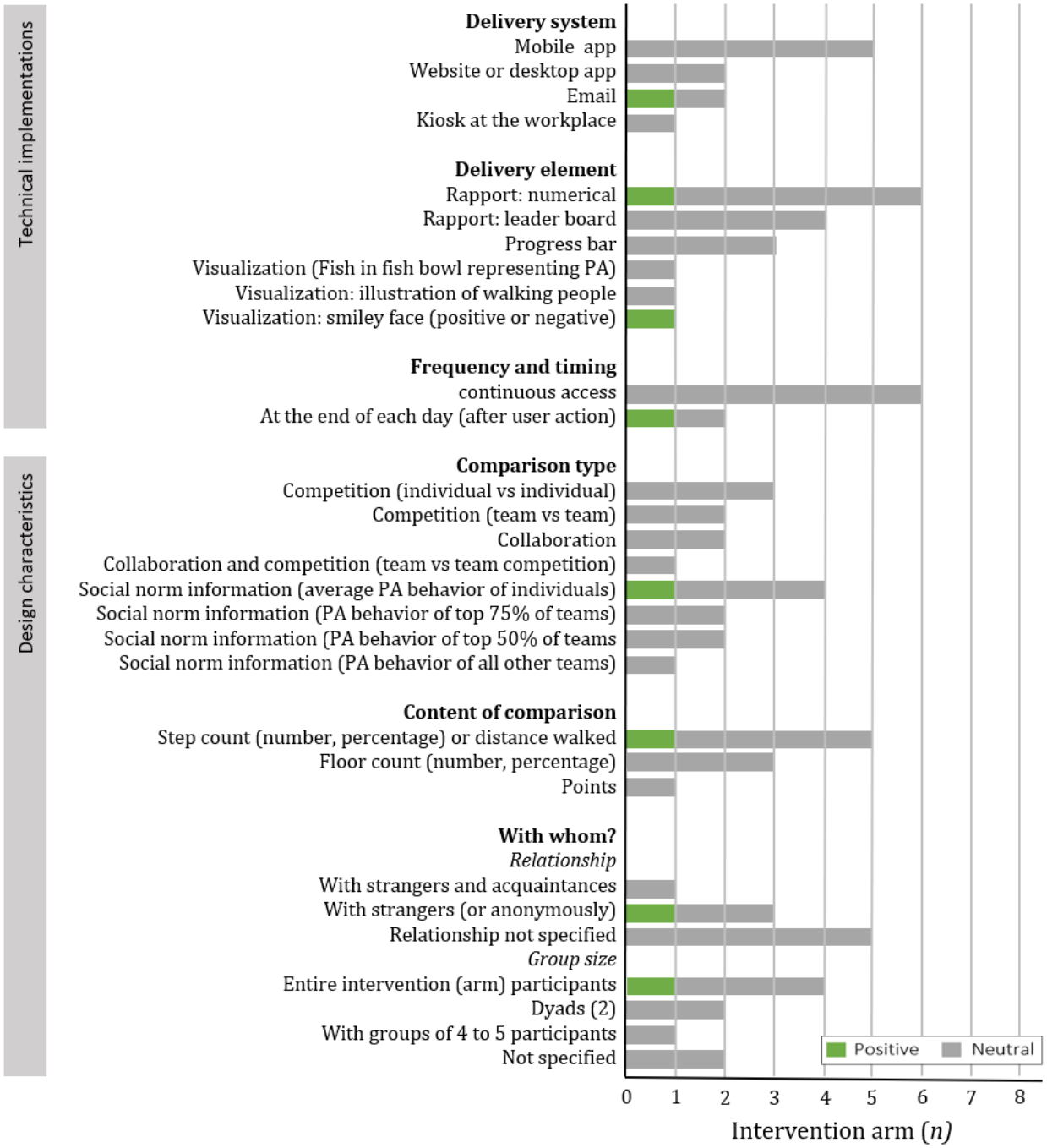
A total of 4 different delivery systems were used, of which mobile apps were most frequently used (n=8). To enable social comparison, almost all studies used reports (n=16) and some included visualizations (n=3; Multimedia Appendix 3). In most studies, the user could view the comparison at all times (n=10); however, sometimes the participant received a comparison message at fixed times (for instance, once a week [n=4]). Regarding the design characteristics, different types of comparisons were identified, including competition (ie, individual vs individual [n=5] and team vs team [n=2]),

collaboration between individuals (n=4), and various forms (n=3) of social norm information (eg, the individual's average step count compared with the step count of all users) [45,47,59,60]. Furthermore, 3 different PA behaviors were compared across the interventions and 7 different social compositions (eg, between 2 friends or with a group of strangers) were identified.

Effect of Including Social Comparison to the Intervention

In total, 6 studies (8 intervention arms) examined whether adding a comparison with an intervention increased the effectiveness of the intervention [39,45,47,59,63,64]. One study (1 intervention arm) found that including social strategies resulted in more PA [59], although it did not significantly increase PA behavior in the remaining 7 arms (Figure 5) [39,45,47,63,64]. The effective implementation involved weekly reports (emails) to inform the user if he or she performed more steps than the average study participants. If the user did, the email contained a positive smiley face. Otherwise, a negative smiley face was placed in the report [59].

Figure 5. The effectiveness of the technical implementations and design characteristics of social comparison compared with receiving no comparison in 9 intervention arms (6 different studies). PA: physical activity.



Comparison of Designs and Technical Implementations of Social Comparison

A total of 2 studies were identified that compared the effect of different design characteristics of social comparisons on PA behavior [39,59]; however, no study examined the technical implementations. One of these studies found that competition was less effective than collaboration or a combination of collaboration and competition (in the short term, ie, 2 weeks [39]). In a second study, the step count of the individual's team (4 individuals) was compared with either the top 25% or the top 50% of the other teams. No significant differences were found between the groups [59], suggesting that this characteristic does not influence the effectiveness of the intervention.

Discussion

Principal Findings

The main objectives of this study were to examine (1) the variation in the technical implementations and design characteristics of common persuasive strategies and (2) if previous mHealth studies examined the influence of the different operationalizations of these persuasive strategies on the effectiveness of the intervention to persuade the user to engage in PA. Similar to previous work on feedback [22], we found that the technical implementations and design characteristics of the examined persuasive strategies vary greatly between studies. For instance, 29 different goal types were identified in the 29 included studies. The goals differed in terms of behavior type (eg, step count, time walked) and difficulty (eg, 10,000 steps per day, 7000 steps per day).

Only a few of the identified implementations and designs were thoroughly examined for effectiveness, especially regarding the technical implementation. Interestingly, the studies that examined this showed that some implementations and designs are more effective in increasing PA compared with others. By performing an in-depth analysis of these studies, this review provides important insights into which implementation types and designs are promising for PA mHealth interventions, as listed below. Furthermore, this study highlights that more research on effective technical implementations and design characteristics of persuasive strategies is essential.

Monitoring

Monitoring was not only the most frequently used strategy in the selected studies but also the least investigated one regarding implementation types and designs. Only one study compared technical implementations, which found no significant difference regarding intervention effectiveness between automatic tracking alone, or both automatic tracking and (active) self-monitoring [41]. Thus, although self-monitoring is arguably more demanding for the user, these results suggest that this additional effort does not influence the effectiveness of the intervention. A possible explanation is that individuals perceive wearing a second device as inconvenient or not pretty (eg, when wearing a dress) [40], which counterbalances the positive effects of the ease of use of automatic tracking. More research is needed to better understand how monitoring can be implemented to successfully support PA behavior.

Goal Setting

The results suggest that the operationalization of goal setting influences the effectiveness of the intervention to increase PA behavior. In line with the Goal Setting Theory [34], this review demonstrates that goals that are tailored to the physical capabilities of the user and/or the context of the user are likely more effective in increasing PA than generic goals [36,37,50,55,62]. Next, challenging (but doable) goals appear to be more effective than easy goals. However, there is no consensus on how challenging the goal should be. Is a small increase of baseline steps a difficult enough goal (eg, 20% increase [50]) or is a large increase required (eg, a 100% increase [37])?

In contrast to the Goal Setting Theory [34] and other theories, such as the Self-Determination Theory (SDT) [67], positing that self-set goals are likely to be more effective, the results of this review indicate that adding self-set goals to an intervention does not increase effectiveness [58] and that goals set by the system might be more promising than self-set goals in increasing PA behavior [68]. However, people will likely choose to set their own goal when offered the choice [68]. Therefore, there might be a discrepancy between what the user wants and what is most effective. Whether a combination of both self-set and assigned goals is effective should be determined in future studies. Future efforts are also needed to explore the technical implementations of gamification, as none of the selected studies examined this.

Reminders

One of the included studies in this study demonstrated that a promising approach to implement reminders is the use of glanceable displays, which is a constant but gentle reminder that resides at the background of the device [41]. However, as most implementation types of reminders were not compared with each other, it is unknown if certain designs of glanceable displays are more attractive and effective than others (eg, *flowers or a robot army* [41]) or if other types of reminders (such as text messages) are equally effective as glanceable displays.

With respect to the design characteristics of reminder text messages, tailoring the timing of the messages might improve the intervention [4,33,69]. One study examined this and found that optimally timed messages were perceived as more useful than randomly timed messages, although it did not result in more PA [42]. Next, regarding the content of the reminders, the results of this study suggest that *generic* reminders do not result in more PA compared with receiving no messages [61]. People perceive these messages as impersonal, *boring* [61] and stop reading them [70,71]. Tailored message content is likely more appropriate [23], as demonstrated in similar research domains [72]. However, none of the papers selected for this study examined this. Taken together, future studies should be conducted to better understand the effectiveness of tailoring (of both the content and the timing) of messages to increase PA behavior.

Rewards

This study suggests that the design and implementation of rewards are important factors that motivate individuals to engage

in PA. For instance, the results suggest that rewards are effective if they are cumulative (eg, with enough points, the participant receives a badge). Furthermore, receiving an immediate reward (eg, immediately after attaining the goal) is likely effective, whereas receiving delayed rewards seems less effective. This is in line with previous research and behavioral economics theory [35], which explain that immediate rewards are perceived as more valuable than rewards given in the future [73,74].

Although several studies have examined effective reward types (eg, comparing monetary rewards with charity donations), multiple important questions remain unanswered. For instance, it is neither known whether the value of the financial reward influences its effectiveness, nor is it known if certain types of virtual rewards (eg, badges, points, levels) are more effective than others. The use of rewards is sometimes criticized, as they mainly promote the extrinsic motivation to perform PA [75,76] and can inhibit the intrinsic motivation associated with lasting behavior change. Consequently, users likely stop performing PA after the reward is removed, as demonstrated in several of the selected studies of this study [44,51,52]. Therefore, it is important to gain a better understanding of how to design rewards that foster feelings of intrinsic motivation [77].

Sharing

Sharing and giving social support are argued to foster feelings of relatedness to others, which is an important determinant of intrinsic motivation to perform PA according to the SDT [67,75]. However, of the 4 studies that examined the effect of adding sharing to the intervention, only 1 study found that having access to a sharing function (with Facebook) significantly increased the step count [56]. A possible explanation for the ineffectiveness of the other 3 implementations of sharing is that they used a delivery system that is not well integrated in people's lives and are, therefore, not frequently used [40,49,60]. However, as no study has directly compared the effectiveness of different delivery systems, future research should examine whether the delivery system of social sharing indeed influences the effectiveness of the intervention. Other technical implementation or design characteristics also require future investigation, such as with whom the data are shared or the type of data that are shared (such as tips from peers or the performed PA behavior).

Social Comparison

In contrast to previous work on PA (mHealth) interventions [7,14], we found that most interventions that included social comparisons did not result in significantly more PA compared with the same intervention without social comparison. There may be various reasons for this, such as the relatively small sample sizes used in the studies [38,47] because the strategy was not optimally operationalized or it could be that the strategy itself is not effective. As only 2 studies compared different design characteristics and no study examined the technical implementation of social comparisons, knowledge on effective operationalizations remains to be limited. However, it does seem that collaboration enhances participation in PA above competition (in line with previous research [38,78]). Notably, it might not be competition itself but the overemphasis on winning that can be counterproductive [78]. In general, it is thought that competition can increase PA when everybody has a reasonable chance to win [79], as also indicated in the exit interviews of Chen and Pu [38].

Strengths, Limitations, and Recommendations

The important strengths of this study are the in-depth systematic analysis of multiple persuasive strategies by using a framework and the exploration of both the technical implementation and design characteristics of the interventions. By doing so, we went beyond examining which persuasive strategies are effective by also examining how these strategies can be effectively implemented and designed. Furthermore, by only including studies that evaluated the isolated effect of the selected persuasive strategies, an overview of implementations and designs could be provided that have been shown to be effective.

However, this study has some limitations. First, inherent to performing scoping reviews, we did not consider the quality of the included studies, which could have biased the results. In addition, as no systematic search was performed, important literature could have been overlooked. To minimize this risk, we performed cross-reference checks of previous (recent) reviews and snowball searches. In light of these limitations, the results of this study should be treated with caution. Second, no corrections for other factors that can influence the effectiveness of the intervention, such as the study duration [36], the geographical location of the study [30], or user friendliness of the intervention [27]. To illustrate, it is likely that the same intervention is not equally effective in different contexts (such as in different countries or seasons). Although we recognize the importance of these factors, it was beyond the scope of this paper to examine this. A third limitation is that the influence of other persuasive strategies of the intervention was not examined. For instance, it is possible that the investigated strategy did not increase PA behavior, even though its operationalization was good, as the other included persuasive strategies of the intervention were not set up properly.

Furthermore, the data extraction was limited to the amount of details in which the researchers described their intervention. For instance, most papers did not describe how they developed the reminder system (eg, whether they used a pool of messages or framed the messages positively or negatively). Therefore, we call on future studies to report the implementations and designs in more detail. For instance, authors can include screenshots of the apps or websites [43], videos of the app, and/or a user manual. To document mHealth interventions, researchers can use the framework provided in Multimedia Appendix 3 and, for instance, the work of Schembre et al [22], Elaheebocus et al [25], and/or Hoffman et al [80].

We realize that the number of implementation types is so diverse that it would cost considerable time and resources to investigate the individual design characteristics and technical implementations of the strategies. A promising approach to reduce this burden on researchers is to build a database in which mHealth PA interventions are described at a granular level, covering both the implementation characteristics of the included persuasive strategies and the study characteristics (as possible confounders). Advanced statistical testing (eg, Meta-Cart analyses [81]) and machine learning techniques [82] allow

identifying which implementation characteristics seem effective for which target group.

Conclusions

Mobile exercise interventions have the potential to increase PA behaviors of individuals [1,6,7]. However, there is a limited understanding of how to effectively develop exercise interventions and its components (ie, their persuasive strategies). To increase this understanding, it is important to examine the operationalization of persuasive strategies and to evaluate its impact on the effectiveness of the intervention. The results of this study highlight the great variation in which monitoring, goal setting, reminders, rewards, sharing, and social comparison are being operationalized. Moreover, the findings of this study suggest that how a conceptual persuasive strategy is being translated into a practical delivery form can influence the effectiveness of the PA intervention. Thus, the operationalization of strategies in mHealth interventions should be critically considered when developing such interventions and before drawing conclusions on the effectiveness of the strategy as a whole. To advance the research field, future research should go beyond evaluating *which* persuasive strategies are effective by also examining *how* these strategies can be effectively implemented and designed.

Acknowledgments

The authors thank regieorgaan Stichting Innovatie Alliantie and Nederlands Wetenschappelijk Onderzoek for funding this research under grant 629.004.013. The authors are especially thankful to Marije Deutekom and Ben Kröse for their role in obtaining funding for this research and for their input in this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 107 KB - mhealth_v9i1e16282_app1.docx]

Multimedia Appendix 2
The data chart used for extracting information from the selected papers consisting of (1) general study characteristics, (2) technical implementations and design characteristics of persuasive strategies, and (3) study results.
[DOCX File , 16 KB - mhealth_v9i1e16282_app2.docx]

Multimedia Appendix 3
The characteristics of the included studies and the design characteristics and technical implementations of monitoring, goal setting, reminders, rewards, social support, and social comparisons.
[XLSX File (Microsoft Excel File), 70 KB - mhealth_v9i1e16282_app3.xlsx]

References

1. Hekler EB, Klasnja P, Froehlich JE, Buman MP. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.: ACM; 2013 Presented at: CHI '13; May 02, 2013; Paris, France p. 3307-3316 URL: https://tinyurl.com/y2eg8hsf
2. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
3. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
4. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
5. 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]
6. 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]
7. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]

8. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
9. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
10. 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]
11. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Sys 2009;24. [doi: 10.17705/1CAIS.02428]
12. Mollee J, Middelweerd A, Kurvers R, Klein M. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
13. Teixeira PJ, Marques MM. Health behavior change for obesity management. Obes Facts 2017;10(6):666-673 [FREE Full text] [doi: 10.1159/000484933] [Medline: 29237167]
14. 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]
15. French DP, Olander EK, Chisholm A, Mc SJ. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
16. 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]
17. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van LW, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
18. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
19. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RAC, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2015 Oct 15:1-16. [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
20. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
21. Wang Y, Fadhil A, Lange J, Reiterer H. Integrating taxonomies into theory-based digital health interventions for behavior change: a holistic framework. JMIR Res Protoc 2019 Jan 15;8(1):e8055 [FREE Full text] [doi: 10.2196/resprot.8055] [Medline: 30664477]
22. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
23. Pope JP, Pelletier L, Guertin C. Starting Off on the Best Foot: a review of message framing and message tailoring, and recommendations for the comprehensive messaging strategy for sustained behavior change. Health Commun 2018 Dec;33(9):1068-1077. [doi: 10.1080/10410236.2017.1331305] [Medline: 28622007]
24. Latimer AE, Brawley LR, Bassett RL. A systematic review of three approaches for constructing physical activity messages: what messages work and what improvements are needed? Int J Behav Nutr Phys Act 2010 May 11;7:36 [FREE Full text] [doi: 10.1186/1479-5868-7-36] [Medline: 20459779]
25. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
26. Schoeppe S, Alley S, Van LW, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
27. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
29. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

30. van Stralen MM, De Vries H, Mudde AN, Bolman C, Lechner L. Determinants of initiation and maintenance of physical activity among older adults: a literature review. Health Psycol Rev 2009 Dec;3(2):147-207. [doi: 10.1080/17437190903229462]
31. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
32. Oinas-Kukkonen H, Harjumaa M. A systematic framework for designing and evaluating persuasive systems. In: Oinas-Kukkonen H, Hasle P, Harjumaa M, Segerstâhl K, Øhrstrøm P, editors. Persuasive Technology. Berlin: Springer; 2008:164-176.
33. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. New York: Association for Computing Machinery; 2009 Presented at: Persuasive '09; April 26-29; Claremond, USA p. 1-7. [doi: 10.1145/1541948.1541999]
34. Tosi HL, Locke EA, Latham GP. A theory of goal setting and task performance. Acad Manage Rev 1991 Apr;16(2):480-483. [doi: 10.2307/258875]
35. Hursh SR. Behavioral economics. J Exp Anal Behav 1984 Nov;42(3):435-452 [FREE Full text] [doi: 10.1901/jeab.1984.42-435] [Medline: 16812401]
36. 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]
37. Chapman GB, Colby H, Convery K, Coups EJ. Goals and social comparisons promote walking behavior. Med Decis Making 2016 May;36(4):472-478. [doi: 10.1177/0272989X15592156] [Medline: 26139447]
38. Chen Y, Pu P. HealthyTogether: exploring social incentives for mobile fitness applications. In: Proceedings of the Second International Symposium of Chinese CHI. New York: Association for Computing Machinery; 2014 Presented at: Chinese CHI '14; April 2014; Toronto, Canada p. 25-34 URL: https://dl.acm.org/doi/10.1145/2592235.2592240 [doi: 10.1145/2592235.2592240]
39. Ciman M, Donini M, Gaggi O, Aiolli F. Stairstep recognition and counting in a serious Game for increasing users' physical activity. Pers Ubiquit Comput 2016 Sep 22;20(6):1015-1033. [doi: 10.1007/s00779-016-0968-y]
40. Consolvo S, Everitt K, Smith I, Landay JA. Design requirements for technologies that encourage physical activity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2006 Presented at: CHI'06; April 22-27 2006; Montréal, Canada p. 457-466 URL: https://doi.org/10.1145/1124772.1124840 [doi: 10.1145/1124772.1124840]
41. Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich J, LeGrand L, et al. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. In: Proceedings of the 10th International Conference on Ubiquitous Computing. New York: Association for Computing Machinery; 2008 Presented at: UbiComp '08; September, 2008; Seoul, Korea p. 54-63. [doi: 10.1145/1409635.1409644]
42. Ding X, Xu J, Wang H, Chen G, Thind H, Zhang Y, editors. WalkMore: Promoting walking with just-in-time context-aware prompts. In: 2016 IEEE Wireless Health (WH). New York: IEEE; 2016 Presented at: 2016 IEEE Wireless Health; Oct 25-27; Bethesda, USA p. 1-8 URL: https://ieeexplore.ieee.org/abstract/document/7764558 [doi: 10.1109/WH.2016.7764558]
43. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone 'app'-delivered randomized factorial trial targeting physical activity in adults. J Behav Med 2017 Oct;40(5):712-729. [doi: 10.1007/s10865-017-9838-y] [Medline: 28255750]
44. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, 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]
45. Harries T, Eslambolchilar P, Rettie R, Stride C, Walton S, van Woerdon HC. Effectiveness of a smartphone app in increasing physical activity amongst male adults: a randomised controlled trial. BMC Public Health 2016 Dec 02;16:925 [FREE Full text] [doi: 10.1186/s12889-016-3593-9] [Medline: 27590255]
46. Herrmanny K, Ziegler J, Dogangün A. Supporting users in setting effective goals in activity tracking. In: Meschtscherjakov A, De Ruyter B, Fuchsberger V, Murer M, Tscheligi M, editors. Lecture Notes in Computer Science. Cham: Springer; Mar 11, 2016.
47. Lin JJ, Mamykina L, Lindtner S, Dalajoux G, Strub HB. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In: Lecture Notes in Computer Science, vol 4206. Berlin: Springer; 2006 Presented at: UbiComp 2006: International Conference on Ubiquitous Computing; September 17-21, 2006; Orange County, USA p. 261-278. [doi: 10.1007/11853565_16]
48. Marshall SJ, Nicaise V, Ji M, Huerta C, Haubenstricker J, Levy SS, et al. Using step cadence goals to increase moderate-to-vigorous-intensity physical activity. Med Sci Sports Exerc 2013 Mar;45(3):592-602. [doi: 10.1249/MSS.0b013e318277a586] [Medline: 23059868]

49. Monroe CM, Bassett DR, Fitzhugh EC, Raynor HA, Thompson DL. Effect of adding online social support tools to an adult walking program: a pilot randomized controlled trial. Health Promot Pract 2017 Jan;18(1):84-92. [doi: 10.1177/1524839915626674] [Medline: 26895847]
50. Moon D, Yun J, McNamee J. The effects of goal variation on adult physical activity behaviour. J Sports Sci 2016 Oct;34(19):1816-1821. [doi: 10.1080/02640414.2016.1140218] [Medline: 26860430]
51. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Eberbach K, et al. Individual versus team-based financial incentives to increase physical activity: a randomized, controlled trial. J Gen Intern Med 2016 Mar 14. [doi: 10.1007/s11606-016-3627-0] [Medline: 26976287]
52. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Heuer J, et al. Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. Ann Intern Med 2016 Mar 15;164(6):385-394. [doi: 10.7326/M15-1635] [Medline: 26881417]
53. Patel MS, Volpp KG, Rosin R, Bellamy SL, Small DS, Fletcher MA, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. Am J Health Promot 2016 Jul;30(6):416-424. [doi: 10.1177/0890117116658195] [Medline: 27422252]
54. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]
55. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York: Association for Computing Machinery; 2015 Presented at: UbiComp '15; 07-11 September, 2015; Osaka, Japan p. 707-718. [doi: 10.1145/2750858.2805840]
56. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The efficacy of a walking intervention using social media to increase physical activity: a randomized trial. J Phys Act Health 2015 Jun;12 Suppl 1:S18-S25. [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]
57. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
58. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. 'Go4Life' exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]
59. Wally CM, Cameron LD. A randomized-controlled trial of social norm interventions to increase physical activity. Ann Behav Med 2017 Oct;51(5):642-651. [doi: 10.1007/s12160-017-9887-z] [Medline: 28213634]
60. Walsh G, Golbeck J. StepCity: a preliminary investigation of a personal informatics-based social game on behavior change. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2014:2371-2376.
61. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: A randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
62. Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28 [FREE Full text] [doi: 10.2196/mhealth.9117] [Medline: 29371177]
63. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014 Jul 5;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
64. Prestwich A, Conner M, Morris B, Finlayson G, Sykes-Muskett B, Hurling R. Do web-based competitions promote physical activity? Randomized controlled trial. Psychol Sport Exerc 2017 Mar;29:1-9. [doi: 10.1016/j.psychsport.2016.11.003]
65. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
66. Lim JS, Noh G. Effects of gain-versus loss-framed performance feedback on the use of fitness apps: mediating role of exercise self-efficacy and outcome expectations of exercise. Comput Human Behav 2017 Dec;77:249-257. [doi: 10.1016/j.chb.2017.09.006]
67. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
68. Herrmanny K, Beckmann N, Nachbar K, Sauer H, Ziegler J, Dogangün A. Using psychophysiological parameters to support users in setting effective activity goals. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2016 Presented at: CHI EA '16; May, 2016; San Jose, USA p. 1637-1646 URL: https://dl.acm.org/doi/10.1145/2851581.2892378 [doi: 10.1145/2851581.2892378]
69. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, USA: Morgan Kaufmann Publishers; 2003.

70. den Akker HO, Cabrita M, Op den Akker R, Jones VM, Hermens HJ. Tailored motivational message generation: a model and practical framework for real-time physical activity coaching. J Biomed Inform 2015 Jun;55:104-115 [FREE Full text] [doi: 10.1016/j.jbi.2015.03.005] [Medline: 25843359]
71. de Vries RAJ, Truong KP, Zaga C, Li J, Evers V. A word of advice: how to tailor motivational text messages based on behavior change theory to personality and gender. Pers Ubiquit Comput 2017 Jun 15;21(4):675-687. [doi: 10.1007/s00779-017-1025-1]
72. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]
73. Laibson D. Golden eggs and hyperbolic discounting. Q J Econ 1997 May 01;112(2):443-478. [doi: 10.1162/003355397555253]
74. Frederick S, Loewenstein G, O'donoghue T. Time discounting and time preference: a critical review. J Econ Lit 2002 Jun;40(2):351-401. [doi: 10.1257/jel.40.2.351]
75. Deci EL, Ryan RM. The 'What' and 'Why' of goal pursuits: human needs and the self-determination of behavior. Psychol Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
76. Bakker F, Oudejans R. Sportpsychologie: 5th edition. Nieuwegein: Arko Sports Media; 2012.
77. Lewis ZH, Swartz MC, Lyons EJ. What's the point?: a review of reward systems implemented in gamification interventions. Games Health J 2016 Apr;5(2):93-99. [doi: 10.1089/g4h.2015.0078] [Medline: 26812253]
78. Roseth CJ, Johnson DW, Johnson RT. Promoting early adolescents' achievement and peer relationships: the effects of cooperative, competitive, and individualistic goal structures. Psychol Bull 2008 Mar;134(2):223-246. [doi: 10.1037/0033-2909.134.2.223] [Medline: 18298270]
79. Weinberg RS, Gould DS. Foundations of sport and exercise psychology. 6th edition. Leeds, United Kingdom: Human Kinetics Publishers; 2014.
80. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014;348:g1687 [FREE Full text] [Medline: 24609605]
81. Dusseldorp E, van GL, van BS, Verheijden MW, van EP. Combinations of techniques that effectively change health behavior: evidence from Meta-CART analysis. Health Psychol 2014 Dec;33(12):1530-1540. [doi: 10.1037/hea0000018] [Medline: 24274802]
82. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157-1182 https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf.

Abbreviations

BIT: Behavioral Intervention Technology
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SDT: Self-Determination Theory

Edited by G Eysenbach; submitted 26.09.19; peer-reviewed by W Goodman, I Hochberg; comments to author 06.01.20; revised version received 03.02.20; accepted 23.03.20; published 18.01.21.

Please cite as:
Sporrel K, Nibbeling N, Wang S, Ettema D, Simons M
Unraveling Mobile Health Exercise Interventions for Adults: Scoping Review on the Implementations and Designs of Persuasive Strategies
JMIR Mhealth Uhealth 2021;9(1):e16282
URL: http://mhealth.jmir.org/2021/1/e16282/
doi:10.2196/16282
PMID:33459598

Original Paper

Validity of Wrist-Wearable Activity Devices for Estimating Physical Activity in Adolescents: Comparative Study

Yingying Hao[1], MEd; Xiao-Kai Ma[1], MEd; Zheng Zhu[1], MEd; Zhen-Bo Cao[1], PhD

School of Kinesiology, Shanghai University of Sport, Shanghai, China

Corresponding Author:
Zhen-Bo Cao, PhD
School of Kinesiology
Shanghai University of Sport
399 Changhai Road
Shanghai, 200438
China
Phone: 86 2165508160
Email: caozb_edu@yahoo.co.jp

Abstract

Background: The rapid advancements in science and technology of wrist-wearable activity devices offer considerable potential for clinical applications. Self-monitoring of physical activity (PA) with activity devices is helpful to improve the PA levels of adolescents. However, knowing the accuracy of activity devices in adolescents is necessary to identify current levels of PA and assess the effectiveness of intervention programs designed to increase PA.

Objective: The study aimed to determine the validity of the 11 commercially available wrist-wearable activity devices for monitoring total steps and total 24-hour total energy expenditure (TEE) in healthy adolescents under simulated free-living conditions.

Methods: Nineteen (10 male and 9 female) participants aged 14 to 18 years performed a 24-hour activity cycle in a metabolic chamber. Each participant simultaneously wore 11 commercial wrist-wearable activity devices (Mi Band 2 [XiaoMi], B2 [Huawei], Bong 2s [Meizu], Amazfit [Huamei], Flex [Fitbit], UP3 [Jawbone], Shine 2 [Misfit], GOLiFE Care-X [GoYourLife], Pulse O2 [Withings], Vivofit [Garmin], and Loop [Polar Electro]) and one research-based triaxial accelerometer (GT3X+ [ActiGraph]). Criterion measures were total EE from the metabolic chamber (mcTEE) and total steps from the GT3X+ (AGsteps).

Results: Pearson correlation coefficients r for 24-hour TEE ranged from .78 (Shine 2, Amazfit) to .96 (Loop) and for steps ranged from 0.20 (GOLiFE) to 0.57 (Vivofit). Mean absolute percent error (MAPE) for TEE ranged from 5.7% (Mi Band 2) to 26.4% (Amazfit) and for steps ranged from 14.2% (Bong 2s) to 27.6% (Loop). TEE estimates from the Mi Band 2, UP3, Vivofit, and Bong 2s were equivalent to mcTEE. Total steps from the Bong 2s were equivalent to AGsteps.

Conclusions: Overall, the Bong 2s had the best accuracy for estimating TEE and total steps under simulated free-living conditions. Further research is needed to examine the validity of these devices in different types of physical activities under real-world conditions.

(JMIR Mhealth Uhealth 2021;9(1):e18320) doi:10.2196/18320

KEYWORDS

wrist-wearable activity devices; accelerometer; energy expenditure; step counts; free-living

Introduction

Since the turn of the 21st century, physical inactivity has increasingly become a global public health issue among youth [1]. In 2010, 81% of adolescents aged 11 to 17 years worldwide failed to achieve the World Health Organization–recommended amounts of moderate to vigorous physical activity (60 minutes or more per day). Of this proportion, girls were less active than boys (87% vs 78%, respectively) [2,3]. Similarly, nearly 70% of adolescents are categorized as insufficiently active, with girls having a higher prevalence of insufficient activity than boys (72% vs 68%, respectively) in China [4-6]. This is a serious issue [7] as physical inactivity in adolescence is associated with adult inactivity [8].

Physical inactivity is one of the leading risk factors for mortality, adding to the burden of noncommunicable diseases (NCDs) and

affecting general health worldwide [9]. Physical inactivity among adolescents is significantly associated with many major health conditions, such as obesity, diabetes, and cardiovascular disease [10]. Young adults who are physically inactive during adolescence are also more likely to be overweight or obese than are their physically active counterparts [11].

Several behavior change methods exist to encourage youth to become more physically active. Self-monitoring of physical activity (PA) with activity devices is helpful to improve the PA levels of adolescents [12]. However, knowing the accuracy of activity devices in adolescents is necessary to identify current levels of PA and assess the effectiveness of intervention programs designed to increase PA.

Rapid advancement in the science and technology of wrist-wearable activity devices offers considerable potential for clinical applications, which may serve as cost-effective and attractive intervention methods for PA improvement apps. It is ideal to measure an adolescent's PA during their usual living conditions to assess when and how long they are active and inactive in a typical day. In this context, measuring PA over the 24-hour day should not be limited to specific activities that can be measured in a laboratory; instead, it should be measured during free-living conditions [13]. Free-living conditions are different from laboratory settings as they offer a wider array of activities and situations for activity devices to measure PA. Accordingly, free-living validity information is important for researchers, fitness coaches, and consumers to choose the most appropriate activity device for their needs [14].

A few studies have examined the accuracy of wrist-wearable activity devices under free-living conditions [14-18]. Dominick [15] and Reid et al [14] reported that compared with the GT3X+ (ActiGraph LLC), the Flex (Fitbit Inc) can estimate total step counts accurately in the free-living conditions, but Chu [16] and Sushames et al [17] showed that the Flex overestimated total step counts with error rates of 15.5% to 47.2%. Other researchers have determined the validity of wrist-, waist-, and arm-wearable devices to monitor the total energy expenditure (TEE) under free-living conditions [18-20]. Brooke et al [19] found that TEE estimated by the Flex, FuelBand (Nike Inc), and Charge HR (Fitbit Inc) were similar to TEE obtained from the arm-worn SenseWear (BodyMedia) and Armband Mini (BodyMedia), but the Shine 2 (Misfit), UP3 (Jawbone), and Vivofit (Garmin Ltd) overestimated TEE with error rates of 15.2%, 22.8%, and 24.5%, respectively [19]. In addition, Dannecker et al [18] found that Fitbit devices significantly underestimated EE with an error rate of 28%. And Ferguson et al [20] found significant differences in TEE obtained from the Shine, UP, and Pulse O2 (Withings) compared with the SenseWear. To date, it appears that no studies have evaluated the accuracy of total step counts and TEE for a large number of wrist-worn activity devices simultaneously. Further, few studies have examined the accuracy of the devices for estimating physical activities in a metabolic chamber that can simulate free-living conditions and estimate energy expenditure of physical activity and TEE, especially in adolescents [14-17,19,20].

Considering this limited evidence, additional research is needed to determine the validity of wrist-wearable activity devices over long periods of time in controlled free-living conditions for adolescents. Hence, the study aimed to determine the validity of 11 wrist-wearable activity devices to monitor total step counts and TEE in adolescents under the stimulated free-living conditions.

Methods

Participants

Nineteen (10 male and 9 female) inactive and healthy participants aged 14 to 18 years volunteered to participate in the study. Participants were recruited from middle schools and community settings located within a 50 kilometer area of Shanghai University of Sport through online advertising, leaflets, and word of mouth. Inclusion criteria included free of metabolic disorders affecting energy expenditure and conditions that influence the ability to perform daily PA, a BMI from 18.5 kg/m^2 to 23.9 kg/m^2, and no attempt to lose weight within the past 2 years. Exclusion criteria included individuals with cardiovascular disease or musculoskeletal injury within the past 6 months and with acute illness, unstable chronic conditions, neurological disorders, and cognitive disorders. Each participant provided written informed consent, and all procedures were approved by the ethical committee of Shanghai University of Sport. The data were collected from December 2017 to June 2018.

Procedures

Participants completed 2 study visits. We asked the participants to refrain from vigorous physical activities on the day before each experiment. At the first visit, participants gave informed consent, had their weight, percent body fat, height, and maximum oxygen uptake (VO_2max) measured while in a fasting state (12 hours postprandial). Participants also completed the long form of the International Physical Activity Questionnaire [21] to determine information about their lifestyle habits. Each participant's energy intake in the metabolic chamber was calculated by multiplying the basal metabolic rate (BMR) predicted by using revised Harris-Benedict equation by 1.55, which was the PA level assumed for a standardized day.

At the second visit, each participant was given 12 wrist-wearable activity devices to wear for 24 hours in the metabolic chamber. We selected these devices based on domestic and foreign sales rankings and the attention of the interrelated research field. Nine were worn on their nondominant wrist in a random order (GT3X+, Flex, Vivofit, B2 [Huawei Technologies Co Ltd], UP3 [Jawbone], Shine 2 [Misfit], Loop [Polar Electro], Pulse O2 [Withings], Mi Band 2 [XiaoMi], and three were worn on their dominant wrist in random order (Amazfit [Huami Corp], Bong 2s [Meizu], GOLiFE Care-X [GoYourLife Inc]). Characteristics of the activity devices are described in Table 1.

Table 1. Activity devices details, set up parameters and analysis software.

Device	Retail price ($)	Steps	Distance	Energy expenditure		Sleep time		Active time	Wear site	Setup parameters	Software	
				Basal	Activity	Deep	Light				Setup	Analysis
GT3X+ (ActiGraph)	249.00	x[a]	—[b]	—	x	—	—	x	hip, wrist	H[c], W[d], sex, DOB[e], 30 Hz, 60 s epoch	Actilife V6.0	Actilife V6.0
Amazfit (Huami)	43.46	x	—	x	x	x	x	—	wrist	H, W, sex, DOB	Midong iPad app	Midong iOS app
Bong 2s (Meizu)	18.75	x	—	—	x	x	x	x	wrist	H, W, sex, DOB	Bong iPad app	Bong iOS app
Flex (Fitbit)	130.52	x	x	+[f]	+	—	—	x	wrist	H, W, sex, DOB	Fitbit iPad app	Fitbit iOS app
Vivofit (Garmin)	72.53	x	x	x	x	x	x	—	wrist	H, W, sex, DOB	Connect iPad app	Connect iOS app
GOLiFE Care-X (GoYourLife)	28.78	x	x	+	+	x	x	—	wrist	H, W, sex, DOB	GOLiFE Fit iPad app	GOLiFE Fit iOS app
B2 (Huawei)	116.13	x	x	—	x	x	x	x	wrist	H, W, sex, DOB	Huawei wearable iPad app	Huawei wearable iOS app
UP3 (Jawbone)	159.74	x	x	x	x	x	x	x	wrist	H, W, sex, DOB	UP iPad app	UP iOS app
Shine 2 (Misfit)	116.13	x	x	+	+	x	x	x	wrist	H, W, sex, DOB	Misfit iPad app	Misfit iOS app
Loop (Polar Electro)	142.44	x	x	+	+	x	x	x	wrist	H, W, sex, DOB	Polar iPad app	Polar iOS app
Pulse O2 (Withings)	137.8	x	x	+	+	x	x	—	wrist	H, W, sex, DOB	Withings iPad app	Withings iOS app
Mi Band 2[g] (XiaoMi)	21.66	x	x	—	x	x	x	x	wrist	H, W, sex, DOB	Xiaomi Sport iPad app	Xiaomi Sport iOS app

[a]x: feature present.

[b]—: feature absent.

[c]H: height.

[d]W: weight.

[e]DOB: date of birth.

[f]+: sum of basal and activity energy expenditures.

[g]Device no longer on the market.

Each participant stayed in the metabolic chamber alone for 24 hours to measure TEE in a simulated free-living environment. Moreover, the researchers would remind the participants to perform daily physical activities (eg, watching TV, sleeping, eating lunch) according to the schedule of activities. The schedule of activities performed in the metabolic chamber is shown in Table 2. Since daily PAs are performed frequently for short durations in actual life, each activity was limited to a period of 30 minutes, except for doing housework and radio gymnastics, which were 10 minutes long.

Table 2. Schedule of activities during the metabolic chamber stay.

Timetable	Activity
19:40	Enter chamber
20:00-22:00	Watch TV
22:00-22:45	Measure RMR[a]
22:45-23:00	Prepare to sleep
23:00-07:00	Sleep
07:00-07:15	Prepare to measure BMR[b]
07:15-08:00	Measure BMR
08:00-08:15	Eat breakfast
08:15-08:45	Listen to music
08:45-09:15	Read
09:15-10:00	Watch videos
10:00-10:10	Do housework
10:10-10:20	Do video calisthenics
10:20-10:50	Slow walk at the speed of 3.2 km/h
10:50-11:20	Play on the phone
11:20-11:50	Type
11:50-12:05	Eat lunch
12:05-13:00	Midday sleep
13:00-13:30	Read
13:30-14:00	Type
14:00-14:30	Fast walk at the speed of 5.6 km/h
14:30-15:00	Listen to music
15:00-25:30	Read
15:30-16:00	Type
16:00-16:30	Run at the speed of 8 km/h
16:30-17:15	Watch videos
17:15-17:45	Play on the phone
17:45-18:00	Eat dinner
18:00-18:30	Listen to music
18:30-19:00	Write
19:00-19:30	Slow walk at a self-selected speed
19:30-20:00	Watch TV
20:20	Leave chamber

[a]RMR: resting metabolic rate.

[b]BMR: basal metabolic rate.

Materials and Measures

Demographics, Anthropometrics, and Cardiorespiratory Fitness

A digital scale (Takei Kiki Kogyo Co Ltd) was used to measure body weight to the nearest 0.1 kg while participants were dressed in light clothing. Height was measured to the nearest 0.1 cm by using an electronic stadiometer with participants standing barefoot. BMI was computed as kg/ m^2. Percent body fat was measured by dual-energy x-ray absorptiometry (Lunar Prodigy, GE Healthcare).

Total Energy Expenditure

The TEE was measured using a whole metabolic chamber (3.85 m width × 2.85 m depth × 2.5 m height; FHC-20S, Fuji Medical Science Co Ltd), which contains a toilet, wash stand, bed, desk

with chair, and treadmill. Participants can sleep, eat, and do different physical activities in the chamber. The temperature and relative humidity of incoming fresh air were maintained at 25.0°C (±0.5°C) and 50.0% (±3.0%), respectively. The sample air is dehumidified using a gas-sampling unit (SCC-C, ABB Corp) and analyzed using a mass spectrometer (Prima PRO, Thermo Fisher Scientific) [22]. The accuracy of VO^2 and VCO^2 measured by metabolic chamber is 99.8% to 99.9%. Once a month, the accuracy and precision of the respiratory chamber are assessed by 24-hour propane combustion tests. The chamber software allows the measurement of energy expenditure with high-time resolution by detecting changes in activity level [23].

Step Counts

The GT3X+ is the most widely used accelerometer to monitor physical activity. The data are displayed as counts, which represents movement intensity and step counts taken. Lee et al [24] reported that the GT3X+ counted 98.5% of the steps compared with the Yamax Digiwalker SW-701 (Yamasa Tokei Keiki Co Ltd) pedometer during free living. We used the wrist-worn GT3X+ to monitor PA while participants were in the metabolic chamber.

Data Processing

Before data collection, the devices were set up with unique user accounts using the parameters of weight, height, gender, and date of birth. Data from the devices were recorded at the beginning and end of each session. Data were downloaded from each device-specific app and uploaded to an iPad (Apple Corp). Step counts from the GT3X+ were downloaded and analyzed using ActiLife 6 software. The Mi Band 2, B2, and Bong 2s yielded estimates of activity EE without accounting for the resting metabolic rate according to the manufacturer's instructions. To facilitate direct comparisons, we calculated the resting energy expenditure for each participant using the following revised Harris-Benedict equation [25]:

- $Male$=88.362+[13.397*$weight(kg)$]+[4.799*$height(cm)$]–(5.677*age)
- $Female$=447.593+[9.247*$weight(kg)$]+[3.098*$height(cm)$]–(4.330*age)

Estimated resting EE values were added to the measured activity EE values from the activity devices to calculate the total EE.

Statistical Analysis

Paired *t* tests were the statistical model adopted for the sample size calculation. The medium ES=0.5 was determined based on the variable of step in the study by Dominick et al [15] (Cohen *d*=0.4). Therefore, we estimated that 17 paired observations would be needed to achieve 80% power to detect the primary outcome variables between the reference devices and activity devices, with 2-sided alpha=.05. To allow for potential withdrawals, 19 participants were randomized.

We analyzed all data using SPSS Statistics version 19.0 (IBM Corp). Data were first checked for normality using standardized skewness and kurtosis values. The results showed that the data in this study were normally distributed. The mean and standard deviation were presented for normally distributed data. Paired *t* tests for normally distributed data were used to analyze differences between the activity devices and the criterion measures: total EE from the metabolic chamber (mcTEE) and step counts from the GT3X+ (AGsteps). A significance level of .05 was used to guide statistical decisions.

Pearson correlation analyses were used to determine the association between the summary scores from each device and the criterion measures. Mean bias (estimated values – measured values) was computed to show the overall underestimation or overestimation of TEE and total step counts by each device compared with the criterion measures at the group level. Mean absolute percentage error (MAPE, [estimated values – measured values] / measured values × 100%) was calculated to quantify the differences between the wrist-wearable activity devices and the criterion measures at the individual level. MAPE accounts for each individual participant's error while avoiding cancellation of errors from underestimation and overestimation [26]. Bland-Altman statistics were performed to determine the 95% limits of agreement to further evaluate individual variations in a more systematic way for each device compared with the criterion measures.

Paired *t* tests are designed to test for differences rather equivalence. The failure to reject the null hypothesis of no difference simply cannot be used to infer agreement or equivalence. Therefore, equivalence testing is used to statistically examine measurement agreements between devices and criterion measures at the group level [26]. Since there are no definitive guidelines to follow to determine the accuracy of the equivalence tests, we selected a 10% error zone. The devices are considered to be equivalent to the criterion measure (with 95% precision) if the 90% confidence interval for a mean of estimated values falls into the defined equivalence zone [27].

Results

Nineteen participants met the eligibility criteria, agreed to participate, and completed the study. Participants' ages ranged from 14 to 18 (mean 17.3 [SD 1.3]) years. BMI ranged from 17.8 to 24.4 (mean 20.5 [SD 1.8]) kg/m^2, and percent body fat ranged from 6.1% to 36.8% (mean 24.0% [SD 9.7%]). The information from the long form of the International Physical Activity Questionnaire confirmed that participants were physically inactive (mean moderate to vigorous PA 95-150 minutes per week). All participants were right hand dominant.

The Pearson correlation coefficient between the wrist-wearable activity devices and the criterion measures for TEE and step counts are displayed in Table 3. All wrist-wearable activity devices were strongly correlated with mcTEE with correlations ranging from *r*=.78 (Shine 2, Amazfit; *P*<.001) to *r*=.96 (Loop; *P*<.001) for TEE. Only the Flex and Vivofit were significantly correlated with AGsteps with *r*=.54 and *r*=.57, respectively (*P*<.05).

Table 3. The Pearson correlation coefficient between wrist-wearable activity devices and criterion measures for total energy expenditure (kcal) and step counts.

Device	McTEE[a]	*P* value	AGsteps[b]	*P* value
Amazfit (Huami)	0.78	<.001	0.45	.06
Bong 2s (Meizu)	0.85	<.001	0.44	.07
Flex (Fitbit)	0.92	<.001	0.54	.02
Vivofit (Garmin)	0.85	<.001	0.57	.01
GOLiFE (GoYourLife)	0.88	<.001	0.20	.42
B2 (Huawei)	0.87	<.001	0.40	.10
UP3 (Jawbone)	0.87	<.001	0.46	.05
Shine 2 (Misfit)	0.78	<.001	0.26	.29
Loop (Polar Electro)	0.96	<.001	0.44	.07
Pulse O2 (Withings)	0.86	<.001	0.22	.38
Mi Band 2 (XiaoMi)	0.91	<.001	0.42	.09

[a]McTEE: total energy expenditure from the metabolic chamber.

[b]AGsteps: total steps from the GT3X+ (ActiGraph).

The mean, standard deviation, and bias between wrist-wearable activity devices and the criterion measures are displayed in Table 4. For TEE, there were no significant differences between the Mi Band 2, UP3, Vivofit, and Bong 2s with mcTEE (*P*>.05). The Flex, Shine 2, and Loop overestimated TEE significantly as noted by the positive bias values ranging from 7.0% (Loop) to 19.0% (Shine 2; *P*<.05). On the contrary, Amazfit, GOLiFE, B2, and Pulse O2 underestimated TEE significantly as noted by the negative bias values ranging from 5.6% (GOLiFE) to 26.6% (Amazfit; *P*<.05). For step counts, there were no significant differences between the Bong 2s, GOLiFE, and Pulse O2 with AGsteps (*P*>.05). The remaining devices overestimated the AGsteps significantly as noted by the positive bias values ranging from 9.7% (Shine 2) to 24.3% (Loop; *P*<.05).

Table 4. Mean, standard deviation, and bias between wrist-wearable activity devices and criterion measures for total energy expenditure (kcal) and step counts (n=19)a.

Device	TEE_P (kcal)[b]	Bias	*P* value	Step count	Bias	*P* value
Amazfit	1496.6 (249.1)	–542.0 (188.2)	<.001	11,910.3 (1864.3)	1766.6 (1753.8)	<.001
Bong 2s[c]	2037.7 (208.8)	–0.9 (164.5)	.98	9586.4 (1600.6)	–557.4 (1602.6)	.16
Flex	2325.6 (272.2)	287.0 (118.7)	<.001	12,228.2 (1377.3)	2084.0 (1327.0)	<.001
Vivofit	2040.5 (290.4)	1.9 (162.2)	.96	12,411.7 (1396.8)	2267.9 (1300.6)	<.001
GOLiFE	1925.0 (246.8)	–113.6 (144.9)	<.001	11,111.8 (2374.5)	968.1 (2497.2)	.12
B2[b]	1922.8 (258.1)	–115.9 (146.7)	<.001	12,193.9 (1246.1)	2050.1 (1456.7)	<.001
UP3	1970.5 (282.4)	–68.1 (148.5)	.06	12,031.5 (1430.4)	1887.7 (1464.4)	<.001
Shine 2	2426.5 (324.4)	387.9 (209.7)	<.001	11,127.2 (1590.4)	983.4 (1820.4)	.04
Loop	2181.8 (312.1)	143.2 (92.4)	<.001	12,613.0 (1785.6)	2469.2 (1714.2)	<.001
Pulse O2	1886.0 (261.4)	–152.6 (154.2)	<.001	11,107.1 (1984.9)	963.3 (2160.9)	.08
Mi Band 2[c]	1979.1 (239.0)	–59.5 (128.5)	.06	11,986.3 (1487.9)	1842.6 (1560.2)	<.001

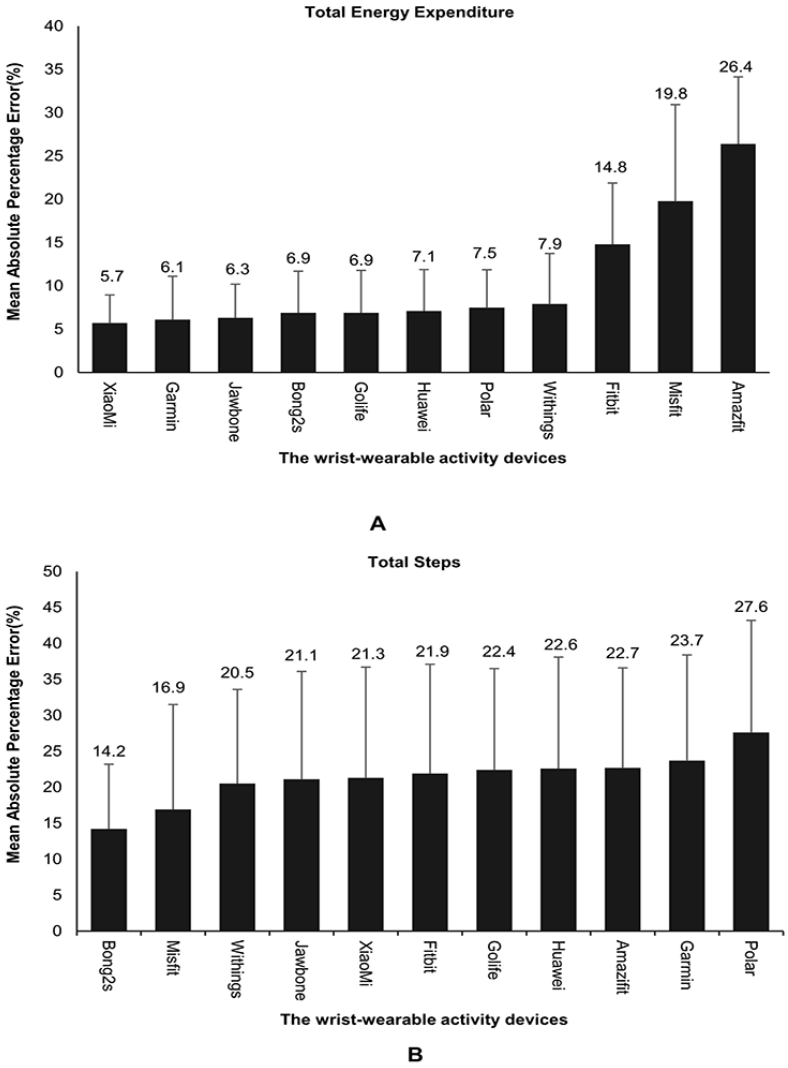
[a] Criterion values: McTEE 2038.6 (299.8) kcal; AGsteps 10143.8 (1396.5).

[b]TEE_P: predicted total energy expenditure.

[c]Add rest energy expenditure.

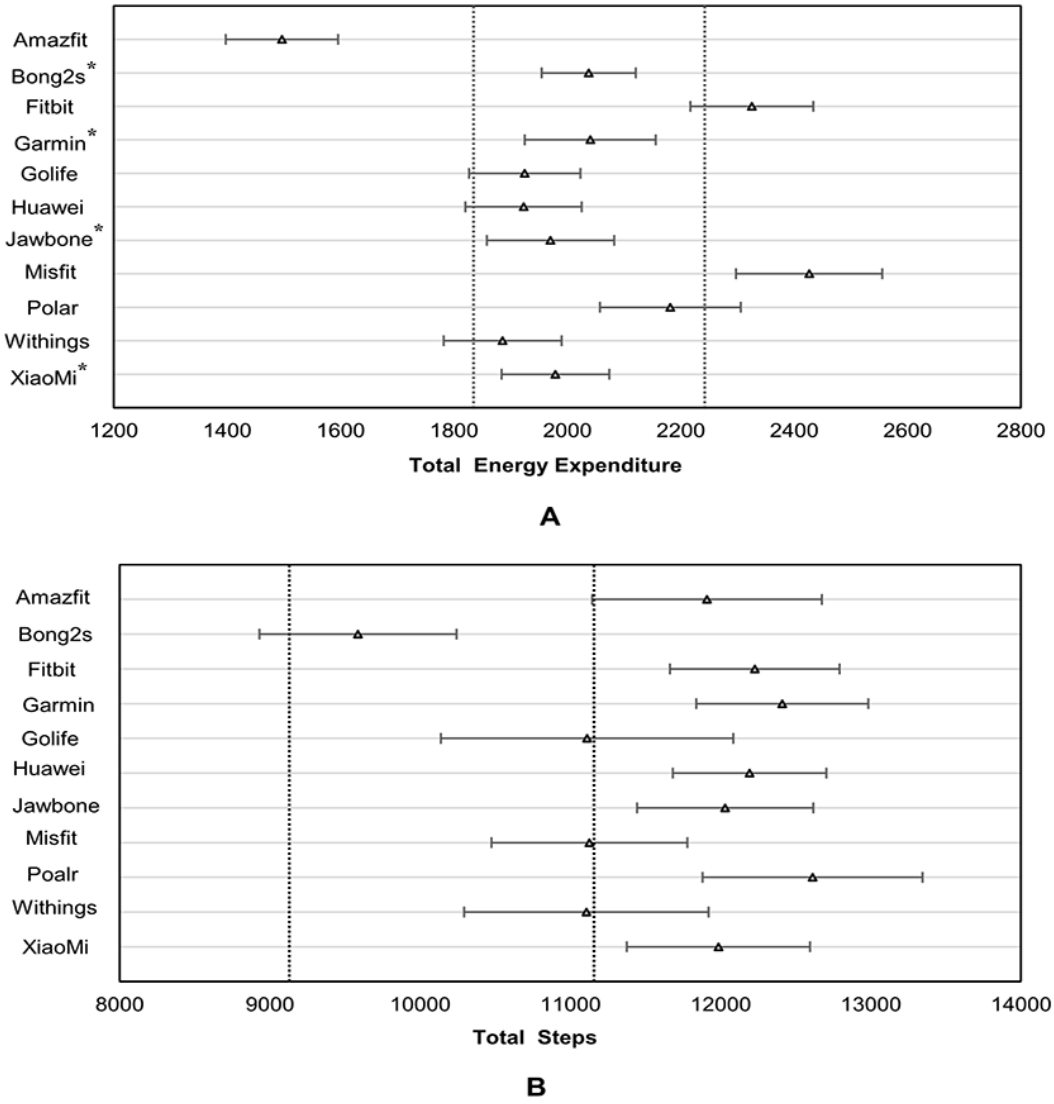
MAPEs for the various devices are illustrated in Figure 1. For TEE, the magnitude of MAPE was least for the Mi Band 2 (5.7%) and highest for the Amazfit (26.4%; Figure 1A). For step counts, the magnitude of MAPE was least for the Bong 2s (14.2%) and highest for the Loop (27.6%; Figure 1B).

Figure 1. Mean absolute percentage error for total energy expenditure and steps estimated by wrist-wearable activity devices.



Equivalence test results are displayed in Figure 2. For TEE, the calculated 90% confidence interval from the Mi Band 2, UP3, Vivofit, and Bong 2s fell within the equivalence zone, indicating equivalence with mcTEE at the group level. The B2 and GOLiFE were close to the equivalence zone (Figure 2A). For step counts, no device was equivalent with AGsteps, however the Bong 2s was closest to the equivalence zone (Figure 2B). All the Bland-Altman scatter plots displayed no systematic bias for all wrist-wearable activity devices (Multimedia Appendix 1).

Figure 2. Agreement on total energy expenditure (kcal) and step counts between criterion measured and devices on 95% equivalence testing. Dashed lines indicate the equivalence zone from criterion measured. Dark lines indicate the 90% confidence interval of estimated values from the devices. *Within the equivalence zone. Δ: mean value estimated by activity devices.



Discussion

Principal Findings

This study aimed to determine the validity of 11 wrist-wearable activity devices for monitoring TEE and total step counts in adolescents during simulated free-living conditions. For TEE, we found that the predicted values by all wrist-wearable activity devices were strongly correlated with TEE obtained from the metabolic chamber and the Mi Band 2, UP3, Vivofit, and Bong 2s measured TEE accurately. For step counts, only the Flex and Vivofit had moderate correlations with the steps obtained by the GT3X+. The Bong 2s, GOLiFE, and Pulse O2 steps were similar to AGsteps. Overall, the wrist-activity devices listed above tended to show good validity when monitoring TEE but not in monitoring step counts at the individual and group levels.

For TEE, the UP3 and Pulse O2 underestimated TEE, and the Flex and Shine 2 overestimated TEE. This finding aligns with previous studies [28,29] showing the UP3, Shine 2, FuelBand, and Pulse O2 compared with criterion measures such as the SenseWear and TEE from a metabolic chamber. The MAPE for the Pulse O2 and Shine 2 (10% to 20%) were similar to those obtained by Ferguson [20]. However, for the UP3, the MAPE in this study was 6.3% which differs widely from values observed by Ferguson [20] and Brooke [19] that reported error rates of more than 29.8% and 22.8%, respectively. Murakami showed the UP3 had a MAPE of 13% compared with TEE from the metabolic chamber and an error rate of more than 20% compared with doubly labeled water. However, Murakami [28] reported using a different reference standard for TEE obtained from the metabolic chamber than the one used in this study. It should be noted, however, that the metabolic chamber had higher accuracy and precision for total daily energy expenditure than doubly labeled water according to the study by Melanson et al [29]. The comparisons may have more accuracy when considering the metabolic chamber as the gold standard for measuring TEE.

This is the first study to examine the validity of Mi Band 2, B2, and Bong 2s on estimating TEE. All three devices were significantly correlated with the TEE, and the Mi Band 2 and Bong 2s estimated TEE accurately. Since the Mi Band 2, B2, and Bong 2s only provided PA energy expenditure output, the predicted resting metabolic rate using the revised Harris-Benedict equation was added to PA energy expenditure measured by these devices in order to provide a more appropriate comparison with TEE in our study. Accordingly, interpretation of the results for these devices requires caution.

This study found that all of the wrist-wearable activity devices overestimated the AGsteps with the exception of the Bong 2s. It is likely that recording total step counts in a free-living setting over a longer duration (ie, 24 hours) resulted in different findings from studies that measured walking for shorter periods of time [30-32]. However, there are similarities in results with Rosenberger et al [13], who showed the UP3 overestimated total steps on the order of 20%, and by Chu et al [16] and Sushames et al [17], who showed that the Flex overestimated total steps from 15.5% to 47.2% (both $P<.05$). Unlike our study, Dominick et al [15] and Reid et al [14] showed the Flex can monitor total steps accurately. However, this discrepancy may be due to different characteristics of the participants studied. The proportion of female participants was nearly 80% in the previous two studies [14,15]. Ferguson et al [20] and Farina et al [33] found that the UP3 and Shine 2 underestimated total steps by 3% and 11%, respectively. This differed from our study, which showed the UP3 and Shine 2 overestimated total step counts by 16.9% and 21.1%, respectively. A possible reason for the underestimation observed by Ferguson [20] and Farina [33] is that their participants were aged 20 to 84 years while the participants in our study were aged 14 to 18 years. In past studies, older adults were shown to be less active compared with younger people [7]. With the lower activity levels and shorter time for monitoring exercise duration, a relatively small range of movement may be overlooked by sensor [34-37]. Therefore, studies are needed with wrist-wearable activity devices in persons with wide age differences who are measured in similar experimental settings so as to assess the accuracy of wrist-wearable activity devices objectively and widely. Further, few [20,28] or no studies have assessed some of these devices, such as the Pulse O2, Mi Band 2, B2, Bong 2s, Amazfit, and GOLiFE Care-X, as done in this study.

As the criterion measure of step counts, the GT3X+ was worn on the nondominant wrist in this study in order to standardize the study design and minimize the measurement variation introduced by the placement of the devices. Compared with hip-worn accelerometers, wrist-worn accelerometers may be less intrusive, particularly during sleep, and may thus engender higher compliance. Wrist-worn accelerometers have been used to monitor children's and adolescents' physical activity for nearly two decades [38]. In their PA surveillance activities, the National Health and Nutrition Examination Survey previously used a uniaxial accelerometer worn on hip to assess PA (2003-2004 and 2005-2006) but has now changed its protocol, asking participants to wear a triaxial accelerometer on the wrist instead of hip in their 2011-2014 surveillance systems, which include persons aged 6 years and older [39].

As a whole, in this study all wrist-wearable activity devices overestimated step counts by 963 to 2469 steps as compared with the GT3X+. It is noteworthy that users may reduce PA if wrist-wearable activity devices overestimate steps, as this may cause the illusion of achieving the goal of fitness and prevent consumers achieving the goal indirectly. This specific type of information about the accuracy of step monitoring devices may be valuable to consumers considering purchasing such devices. That said, contemporary wrist-wearable activity devices have emphasized wrist locations by the manufacturers for their less obstructive placement and user's convenience in checking their progress throughout the day. Wrist locations also facilitate integration with telecommunications features (ie, smart watch), enable sleep detection, and promote participant compliance [40].

In this study, we found that the price and performance of wrist-wearable activity devices seems to be unrelated. The most inexpensive wrist-wearable activity device, Bong 2s, was one of the best performing activity devices, while more expensive activity devices (Loop, Flex, B2) showed a large difference in accuracy, a finding similar to results in the study by Ferguson et al [20]. It is likely that the addition of smartphone connectivity, intelligence, wearability, and esthetics contribute to higher priced wrist-wearable activity devices.

Strengths and Limitations

Our study has some strengths. First, participants were adolescents aged 14 to 18 years. In all previous studies, samples included adults and older people only. The addition of this study, combined with investigations with a broader age range of participants, can provide more confidence that the results can be generalized to a broader population, especially teenagers who typically have lower levels of physical activity in many societies [4]. Second, we used the metabolic chamber as a gold standard criterion measure for TEE. A high-precision metabolic chamber allowed precise measurement of EE which facilitated the output of credible results. Beyond that, the cubage of the metabolic chamber is 11.4 m^2, similar to a household room. Accordingly, we could simulate a free-living environment to monitor daily behavior in a real-time 24-hour daily life. Third, compared with previous studies, we examined the accuracy of a wide range of wrist-wearable activity devices: Mi Band 2, Flex, UP3, Vivofit, Shine 2, B2, Bong 2s, GOLiFE Care-X, Pulse O2, Amazfit, and Loop. The price of the wrist-wearable activity devices ranged from US $18 to $250, which is suitable for people in different consumer stratums. Collectively, the results in our study can inform decision making about the use of wrist-wearable activity devices.

This study is not without limitations. First, we did not assess the reliability of the wrist-wearable activity devices. Poor reliability can negatively impact validity. In further studies, we need to test the reliability of wrist-wearable activity devices to ensure consistency among the different brands. Second, we need to further test wrist-wearable activity device monitors to assess multiple parameters such as different types of PA EE, distance, time of various intensity, sleep, and so on, which may impact the validity of the devices. Additionally, the results of our research should be carefully considered for application to

overweight and obese people. Finally, according to the time schedule in the metabolic chamber, there were many activities of daily life (eg, listening to music, doing housework, writing), but these data were not revealed in detail in this paper.

Conclusions

In conclusion, the Mi Band 2, UP3, Vivofit, and Bong 2s wrist-worn activity devices estimated TEE accurately both at individual and group level as compared to the TEE obtained in a metabolic chamber. The Bong 2s, GOLiFE, and Pulse O2 were similar to total step counts recorded by the GT3X+ at the individual level. No devices were equivalent with total step counts from the GT3X+ at the group level. With the upgrade and expansion of the measurement abilities of the wrist-wearable activity devices, the research field should regularly assess the accuracy of new devices to ensure that the wrist-wearable activity devices can be used with confidence in scientific research and by practitioners in daily life.

Acknowledgments

We gratefully thank Prof Barbara Ainsworth for her constructive comments and suggestions on our manuscript and our laboratory members (Lin Zhang, Chen Sun) for their kind assistance during data collection. We also thank the study participants for their time and commitment to the study protocol. Without their contribution, this study would not have been possible. This study was funded partly by the Sports Scientific Research Program of the Shanghai Municipal Education Commission (No. HJTY-2016-D31) and the Science and Technology Commission of Shanghai Municipality (No. 16080503300).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bland-Altman scatterplots of steps and total energy expenditure for all wrist-wearable activity devices.
[PDF File (Adobe PDF File), 172 KB - mhealth_v9i1e18320_app1.pdf]

References

1. Ng SW, Popkin BM. Time use and physical activity: a shift away from movement across the globe. Obes Rev 2012 Aug;13(8):659-680 [FREE Full text] [doi: 10.1111/j.1467-789X.2011.00982.x] [Medline: 22694051]
2. Global Health Observatory (GHO) data: prevalence of insufficient physical activity. World Health Organization. URL: http://158.232.12.119/gho/ncd/risk_factors/physical_activity/en/ [accessed 2019-12-28]
3. Overview: Behavioral Risk Factor Surveillance System 2014. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/brfss/annual_data/2014/pdf/Overview_2014.pdf [accessed 2020-12-20]
4. Prevalence of insufficient physical activity among school going adolescents. World Health Organization. URL: http://apps.who.int/gho/data/view.main.2463ADO?lang=en [accessed 2019-12-29]
5. Tudor-Locke C, Ham SA, Macera CA, Ainsworth BE, Kirtland KA, Reis JP, et al. Descriptive epidemiology of pedometer-determined physical activity. Med Sci Sports Exerc 2004 Sep;36(9):1567-1573. [doi: 10.1249/01.mss.0000139806.53824.2e] [Medline: 15354039]
6. Fan X, Cao Z. Physical activity among Chinese school-aged children: national prevalence estimates from the 2016 Physical Activity and Fitness in China—The Youth Study. J Sport Health Sci 2017 Dec;6(4):388-394 [FREE Full text] [doi: 10.1016/j.jshs.2017.09.006] [Medline: 30356592]
7. 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]
8. Tammelin T, Näyhä S, Laitinen J, Rintamäki H, Järvelin MR. Physical activity and social status in adolescence as predictors of physical inactivity in adulthood. Prev Med 2003 Oct;37(4):375-381. [doi: 10.1016/s0091-7435(03)00162-2] [Medline: 14507496]
9. Blazer DG. Social support and mortality in an elderly community population. Am J Epidemiol 1982 May;115(5):684-694. [doi: 10.1093/oxfordjournals.aje.a113351] [Medline: 7081200]
10. Chen Y, Zheng Z, Yi J, Yao S. Associations between physical inactivity and sedentary behaviors among adolescents in 10 cities in China. BMC Public Health 2014 Jul 22;14:744 [FREE Full text] [doi: 10.1186/1471-2458-14-744] [Medline: 25053199]
11. Menschik D, Ahmed S, Alexander MH, Blum RW. Adolescent physical activities as predictors of young adult weight. Arch Pediatr Adolesc Med 2008 Jan;162(1):29-33. [doi: 10.1001/archpediatrics.2007.14] [Medline: 18180409]
12. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]

13. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465. [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
14. Reid RER, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
15. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
16. Chu AHY, Ng SHX, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
17. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of fitbit flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016 Sep;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
18. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. Med Sci Sports Exerc 2013 Nov;45(11):2105-2112 [FREE Full text] [doi: 10.1249/MSS.0b013e318299d2eb] [Medline: 23669877]
19. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017;31(4):1097-1106. [doi: 10.1519/jsc.0000000000001571]
20. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
21. 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]
22. Ma S, Zhu Z, Zhang L, Liu X, Lin Y, Cao Z. Metabolic effects of three different activity bouts during sitting in inactive adults. Med Sci Sports Exerc 2020 Apr;52(4):851-858. [doi: 10.1249/MSS.0000000000002212] [Medline: 31764465]
23. Nguyen-Duy T, Nichaman MZ, Church TS, Blair SN, Ross R. Visceral fat and liver fat are independent predictors of metabolic risk factors in men. Am J Physiol Endocrinol Metab 2003 Jun;284(6):E1065-E1071 [FREE Full text] [doi: 10.1152/ajpendo.00442.2002] [Medline: 12554597]
24. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
25. Roza AM, Shizgal HM. The Harris Benedict equation reevaluated: resting energy requirements and the body cell mass. Am J Clin Nutr 1984 Jul;40(1):168-182. [doi: 10.1093/ajcn/40.1.168] [Medline: 6741850]
26. Welk GJ, Bai Y, Lee J, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 2019 Aug;51(8):1767-1780 [FREE Full text] [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]
27. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. Med Sci Sports Exerc 2018 Apr;50(4):837-845. [doi: 10.1249/MSS.0000000000001481] [Medline: 29135817]
28. Murakami H, Kawakami R, Nakae S, Nakata Y, Ishikawa-Takata K, Tanaka S, et al. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. JAMA Intern Med 2016 Dec 01;176(5):702-703. [doi: 10.1001/jamainternmed.2016.0152] [Medline: 26999758]
29. Melanson EL, Swibas T, Kohrt WM, Catenacci VA, Creasy SA, Plasqui G, et al. Validation of the doubly labeled water method using off-axis integrated cavity output spectroscopy and isotope ratio mass spectrometry. Am J Physiol Endocrinol Metab 2018 Feb 01;314(2):E124-E130 [FREE Full text] [doi: 10.1152/ajpendo.00241.2017] [Medline: 28978547]
30. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: an accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140. [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]
31. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016 Dec;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025] [Medline: 27477705]
32. Storm FA, Heller BW, Mazzà C. Step detection and activity recognition accuracy of seven physical activity monitors. PLoS One 2015;10(3):e0118723 [FREE Full text] [doi: 10.1371/journal.pone.0118723] [Medline: 25789630]
33. Farina N, Lowry RG. The validity of consumer-level activity monitors in healthy older adults in free-living conditions. J Aging Phys Act 2018 Jan 01;26(1):128-135. [doi: 10.1123/japa.2016-0344] [Medline: 28595019]
34. Buckworth J, Lee RE, Regan G, Schneider LK, DiClemente CC. Decomposing intrinsic and extrinsic motivation for exercise: application to stages of motivational readiness. Psychol Sport Exerc 2007 Jul;8(4):441-461. [doi: 10.1016/j.psychsport.2006.06.007]

35. Corder K, Brage S, Ramachandran A, Snehalatha C, Wareham N, Ekelund U. Comparison of two Actigraph models for assessing free-living physical activity in Indian adolescents. J Sports Sci 2007 Dec;25(14):1607-1611. [doi: 10.1080/02640410701283841] [Medline: 17852668]
36. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. Med Sci Sports Exerc 2010 Feb;42(2):368-374 [FREE Full text] [doi: 10.1249/MSS.0b013e3181b3af49] [Medline: 19927022]
37. Crouter SE, Schneider PL, Karabulut M, Bassett DR. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. Med Sci Sports Exerc 2003 Aug;35(8):1455-1460. [doi: 10.1249/01.MSS.0000078932.61440.A2] [Medline: 12900704]
38. Rowlands AV, Rennie K, Kozarski R, Stanley RM, Eston RG, Parfitt GC, et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. Med Sci Sports Exerc 2014 Dec;46(12):2308-2316. [doi: 10.1249/MSS.0000000000000365] [Medline: 24781890]
39. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014 Jul;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
40. Freedson PS, John D. Comment on Comment on "estimating activity and sedentary behavior from an accelerometer on the hip and wrist". Med Sci Sports Exerc 2013 May;45(5):962-963. [doi: 10.1249/MSS.0b013e31827f024d] [Medline: 23594509]

Abbreviations

AGsteps: step counts from the ActiGraph GT3X+
BMR: basal metabolic rate
MAPE: mean absolute percentage error
mcTEE: total energy expenditure from metabolic chamber
NCD: noncommunicable disease
PA: physical activity
TEE: total energy expenditure

Edited by L Buis; submitted 19.02.20; peer-reviewed by T Muto, J Rahman, P Dharmarajan, YS Chen; comments to author 31.08.20; revised version received 20.10.20; accepted 02.12.20; published 07.01.21.

Please cite as:
Hao Y, Ma XK, Zhu Z, Cao ZB
Validity of Wrist-Wearable Activity Devices for Estimating Physical Activity in Adolescents: Comparative Study
JMIR Mhealth Uhealth 2021;9(1):e18320
URL: http://mhealth.jmir.org/2021/1/e18320/
doi:10.2196/18320
PMID:33410757

Original Paper

Digital Contact Tracing Based on a Graph Database Algorithm for Emergency Management During the COVID-19 Epidemic: Case Study

Zijun Mao[1,2], PhD; Hong Yao[1,2], PhD; Qi Zou[1,2], PhD; Weiting Zhang[1,2], MA; Ying Dong[1,2,3], BA

[1]College of Public Administration, Huazhong University of Science and Technology, Wuhan, China

[2]Non-traditional Security Research Center, Huazhong University of Science and Technology, Wuhan, China

[3]School of Law and Humanities, China University of Mining and Technology, Beijing, China

Corresponding Author:
Zijun Mao, PhD
College of Public Administration
Huazhong University of Science and Technology
1037 Luoyu Road
Wuhan, 430074
China
Phone: 86 15871410683
Email: maozijun@hust.edu.cn

Abstract

Background: The COVID-19 epidemic is still spreading globally. Contact tracing is a vital strategy in epidemic emergency management; however, traditional contact tracing faces many limitations in practice. The application of digital technology provides an opportunity for local governments to trace the contacts of individuals with COVID-19 more comprehensively, efficiently, and precisely.

Objective: Our research aimed to provide new solutions to overcome the limitations of traditional contact tracing by introducing the organizational process, technical process, and main achievements of digital contact tracing in Hainan Province.

Methods: A graph database algorithm, which can efficiently process complex relational networks, was applied in Hainan Province; this algorithm relies on a governmental big data platform to analyze multisource COVID-19 epidemic data and build networks of relationships among high-risk infected individuals, the general population, vehicles, and public places to identify and trace contacts. We summarized the organizational and technical process of digital contact tracing in Hainan Province based on interviews and data analyses.

Results: An integrated emergency management command system and a multi-agency coordination mechanism were formed during the emergency management of the COVID-19 epidemic in Hainan Province. The collection, storage, analysis, and application of multisource epidemic data were realized based on the government's big data platform using a centralized model. The graph database algorithm is compatible with this platform and can analyze multisource and heterogeneous big data related to the epidemic. These practices were used to quickly and accurately identify and trace 10,871 contacts among hundreds of thousands of epidemic data records; 378 closest contacts and a number of public places with high risk of infection were identified. A confirmed patient was found after quarantine measures were implemented by all contacts.

Conclusions: During the emergency management of the COVID-19 epidemic, Hainan Province used a graph database algorithm to trace contacts in a centralized model, which can identify infected individuals and high-risk public places more quickly and accurately. This practice can provide support to government agencies to implement precise, agile, and evidence-based emergency management measures and improve the responsiveness of the public health emergency response system. Strengthening data security, improving tracing accuracy, enabling intelligent data collection, and improving data-sharing mechanisms and technologies are directions for optimizing digital contact tracing.

(JMIR Mhealth Uhealth 2021;9(1):e26836) doi:10.2196/26836

KEYWORDS

COVID-19; digital contact tracing; emergency management; graph database; big data; visualization; China

Introduction

The COVID-19 epidemic is still spreading rapidly worldwide. Tens of millions of people have been infected, and the number of infections is still growing rapidly. Most countries or regions are still in states of public health emergency. Although China, the United States, and other countries have successfully developed COVID-19 vaccines, the production and application of the vaccines still have large gaps. Moreover, effective drugs for treatment of COVID-19 have not been successfully developed. Therefore, quickly identifying and tracing individuals infected with COVID-19 and their contacts and adopting active emergency management measures, such as travel restrictions, health monitoring, and home-based quarantine, are necessary for all countries and regions to overcome the COVID-19 epidemic [1].

Contact investigation has always been an important public health strategy and a key process in epidemic emergency management. The scale of COVID-19 infection poses a major challenge to contact investigation [2]. Traditional manual contact investigation is conducted through telephone or face-to-face interviews, and these processes are usually slow, laborious, and inefficient; therefore, it is difficult for health officials and disease control staff to obtain all the information of all contacts [2-5]. In addition, traditional manual contact tracing is not effective in identifying and managing asymptomatic and mildly symptomatic infected individuals, who account for 84% of the total number of infections [6,7]. Traditional manual contact tracing is also unlikely to be implemented for displaced migrants [8]. Some medical experts believe that COVID-19 may develop into a seasonal epidemic, similar to influenza, which creates great challenges for traditional manual contact tracing [9].

Digital contact tracing uses electronic information to identify exposures to infection; it has the potential to address the limitations of traditional contact tracing, such as scalability, notification delays, recall errors, and contact identification in public spaces [10,11]. Current digital technology enables the continuous tracing of individuals and locations using mobile phones, GPS, wireless fidelity (Wi-Fi), and Bluetooth to record and trace the spatiotemporal trajectory of individuals to identify potential contacts [12,13]. These systems can help overcome the limitations of manual contact tracing [12]. An empirical study proved that the widespread use of digital contact tracing can reduce the spread of epidemics, and far fewer individuals are placed in isolation when such systems are applied [12].

The data storage and processing models of digital contact tracing adopted by some countries and regions are generally divided into centralized and decentralized models [14]. In the centralized model, anonymized data are uploaded from people's contact information to centralized servers; then, health authorities can check, notify, and manage previous contacts [14,15]. In comparison, the decentralized model locally stores these key codes and allows users to (1) notify the system if they have tested positive (or not), so the mobile app will upload the last 14 days of locally stored keys to the server; and (2) voluntarily check their risk exposure, that is, whether they have been in contact with someone who may have been infected, by downloading the uploaded keys from the server and matching locally against the stored keys to evaluate their risk of exposure [14].

South Korea and China have adopted the centralized model for digital contact tracing [2]. Korea's Center for Disease Control and Prevention (CDC) built the COVID-19 Smart Management System based on big data, such as security cameras, credit card records, and even the GPS of cars and mobile phones, to trace people who may be exposed to COVID-19 and to implement quarantine measures [16]. The mainland Chinese government has developed a smartphone applet to collect residents' self-reported health characteristics and travel history and to combine the location information generated by the users' specific quick response (QR) codes to trace infected persons and contacts for individual risk assessment. A health code is generated for the identification of personal infection risk; this code serves as a voucher for passage, work resumption, and school resumption [10,16]. The COVID-19 epidemic emergency management strategies in Hong Kong and Taiwan focus on the isolation of imported cases. Hong Kong has designed a compulsory electronic wristband with a mobile app, which all travelers entering Hong Kong are required to wear [16]. The device alerts the authorities when a traveler leaves their designated quarantine place [16]. Taiwan authorities used the National Health Insurance Database and the Immigration and Customs Database to match symptoms and travel experiences to identify people who may have been exposed to COVID-19 [16].

The decentralized model was adopted by countries and regions such as Europe, North America, and Singapore [2]. Google, Apple, the Massachusetts Institute of Technology, Singapore authorities, and some European consortia have developed apps based on smartphone Bluetooth technology that can notify users if they have been in contact with confirmed patients [2,10,17]. These apps use the Bluetooth signal strength to infer the distance between smartphones and define the user's exposure status based on the distance and time of proximity to individuals who are subsequently identified as infected. The apps are adopted with users' permission and do not collect user data in any centralized digital or information system [2,10,17].

App-based digital contact tracing in a decentralized mode can only be effective when used by 40%-70% of smartphone users [3]. The early results of the adoption of Singapore's Bluetooth-based digital contact tracing app showed that only 20% of the population had installed the app as of April 21, 2020, because of its operational complexity, the limitations of the underlying technology, and citizen trust [2]. The proportions of countries in Europe and North America that adopted such apps were even lower [15]. Europe, North America, and other Western regions are relatively adverse to digital contact tracing under the centralized model due to considerations of public freedom and privacy; these countries also rejected strict emergency management measures for similar reasons at the beginning of the COVID-19 epidemic, which led to uncontrolled spread of the epidemic [18,19]. Fortunately, after recognizing the importance of strict emergency management measures, Western countries subsequently adopted measures to address the COVID-19 epidemic and achieved good results. Practices

have proved that governments must take a more active role in COVID-19 epidemic emergency management; digital contact tracing in a centralized model may become the mainstream measure for major epidemic emergency management worldwide.

An empirical study [14] showed that the use of a centralized model for digital contact tracing is more effective, and the minimum utilization or coverage rate of the centralized model in epidemic control is remarkably lower than that of the decentralized model; however, this study ignored the unique problems faced by the centralized model in the implementation process. Taking Mainland China as a typical example, the health code applets developed in Mainland China based on WeChat social software and Alipay payment software have high utilization rates, which may meet the requirements for effective contact tracing. However, the objectivity of health codes is insufficient, and their effectiveness is weakened because of the autonomy of users to report their health characteristics and contact history and the arbitrariness with which users scan designated QR codes in various places to determine location information.

Therefore, this study aimed to overcome the insufficient use and objectivity of existing digital contact tracing–related practices and provide new solutions to further improve the effectiveness and reliability of digital contact tracing. Hainan Province, China, was selected as the case in this study. Hainan Province adopted a centralized model to conduct contact tracing during the COVID-19 epidemic. Moreover, the model gathered multisource epidemic data that relied on the government's big data public service platform, which enabled government agencies to apply graph database algorithms, data visualization, and other digital technologies to determine and trace contacts from hundreds of thousands of epidemic records. This study describes the organizational process, technical process, application prospects and possible obstacles of digital contact tracing in Hainan Province, which may provide a more effective solution and technical support for other countries and regions.

Methods

Case Selection

Hainan Province is one of the most popular tourist destinations among Chinese and even global tourists because of its tropical island scenery. Due to the massive influx of tourists, the flow rate of the population of Hainan Province is very large. Additionally, due to the geographical location of the island, the influx of infected individuals is the main challenge in the emergency management of the COVID-19 epidemic. As of August 7, 2020, Hainan Province reported a total of 171 confirmed individuals infected with COVID-19 [20]. The vast majority of these cases were identified between January 22 and February 19, 2020 (168 cases); only 3 cases appeared in the next 6 months [20]. This outcome proved the success of the emergency management strategy of the COVID-19 epidemic in Hainan Province.

Hainan Province issued the "Regulations on the Development and Application of Big Data in Hainan Province" in October 2019 [21], which clarifies relevant regulations on the development and sharing of big data, application and industry promotion, data security, and legal responsibility at all levels of government agencies. According to the regulations, Hainan Province established the Big Data Administration and the big data public service platform to collect the transaction information system data of government agencies at all levels [21]. These regulations helped Hainan Province become one of the first provinces in China to use digital technology to respond to the COVID-19 epidemic [22]. Hainan Province relied on the government's big data public service platform to gather epidemic multisource big data and used graph database algorithms to determine and trace contacts, which achieved good results. Hainan Province was selected as the case in our research to better demonstrate the innovative ideas and technologies of digital contact tracing development in the context of big data.

Materials

Qualitative Materials

We conducted semistructured interviews in collaboration with the project manager and technical staff of the Big Data Administration who were in charge of Hainan's COVID-19 epidemic digital contact tracing project to understand the details of the whole process. We explained the purpose and content of the interview to the interviewees before the interview, recorded the outlines during the interview, and transcribed the voice recording to text format after the interview.

Quantitative Materials

Under the leadership and coordination of the Command of Hainan Provincial Epidemic Prevention and Control, the Hainan Provincial Big Data Administration collected epidemic data from different agencies, including epidemic investigation records, confirmed infected individuals' information and their spatiotemporal trajectory, high-risk population information, information on close contacts and patients with fever, and the mobile phone signaling data of imported residents or travelers with a history of residence in Hubei Province (patients with COVID-19 were first reported in Wuhan, Hubei Province, and Hubei Province accounted for the majority of cases in China). The information in the specific databases is shown in Table 1. The technical personnel organized by the Hainan Provincial Big Data Administration cleaned and integrated the multisource epidemic data, then uploaded the data to the government's big data public management platform. The platform provided hive data warehouse tools to help authorized personnel manage, extract, query, and analyze data. The time range covered by all the databases was January 21 to February 9, 2020. The collection, storage, and analysis of all data in this study were completed by the administrative officials and technical personnel authorized by the Hainan Provincial Big Data Administration. The authors of this paper had no authority to participate in any specific process. The only information possessed by the authors was the organizational process, technical process, and analysis results provided by the Hainan Provincial Big Data Administration. No data or information in this study involved citizens' sensitive or private information.

Table 1. Data collection list of COVID-19 epidemic digital contact tracing in Hainan Province.

Database	Data provider	Description	Sample size (n)	Collection method
Resident trajectory information database from Hubei Province	Health Committee	Used to understand where citizens have gone and who they have contacted	205,833	The health committee organizes the disease prevention and control center, community health service agencies, and community (village) resident committees to conduct household surveys and telephone verifications.
Confirmed Person Information Database	Public Security Department; Health Committee	Used to understand which patients have died or have recovered and been discharged	163	Reported by medical and health institutions
Information database of high-risk groups and close contacts	Epidemic Prevention and Control Headquarters	Used to understand the close contacts of high-risk individuals in confined spaces	2,269	Summary of Center for Disease Control information
Hospital fever information database	Epidemic Prevention and Control Headquarters	Used to understand who was infected, where the person was treated, whether the infected person recovered, and where the infected person lives	113,606	Hospitals report through the health committee information system
Mobile phone signaling database	Communications operator branch	Used to understand where people moved to and are staying	231,296	Directly provided by China Mobile, China Unicom, China Telecom, and China Broadcasting Network Corp, Ltd

Main Methods

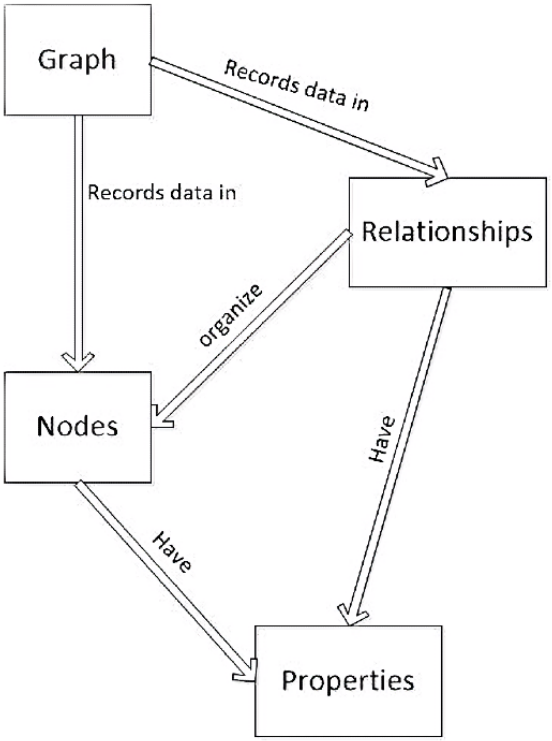
The Graph Database and Its Algorithm

A graph database is a new type of database system based on graph theory and algorithms that efficiently processes complex relational networks. A graph database can efficiently process large-scale, complex, interconnected, and changeable data, and its computational efficiency is far higher than that of a traditional relational database [23]. A graph database is scalable and flexible, is suitable for complex relationship management and relationship query reasoning, and reveals implicit relationships between entities [24].

Most graph databases provide a query language that is suitable for representing graph structures and graph query. Neo4j (Neo4j, Inc) is a Java-based open-source graph database with high performance, high reliability, and strong scalability [23]. The graph database model based on Neo4j is shown in Figure 1. The information modeling of Neo4j includes three structural units: nodes, relationships, and properties [25]. Each node in the database graph can establish a relationship with any other node, and each node can set multiple properties [25]. Each relationship in the graph must have a start node and an end node, and each relationship can also set multiple properties [25].

Our research used the Cypher algorithm of Neo4j version 3.4.15 to construct an association graph among the high-risk population, the general population, vehicles, public places, and other key information; reveal the hidden network of relationships; and identify the risks in the relationships to identify and trace the contacts of COVID-19 cases.

Figure 1. Graph database model based on Neo4j.



Data Visualization and Interactive Operating System Based on ECharts

ECharts (Apache Software Foundation) is an open-source visualization tool based on JavaScript that can run smoothly on personal computers and mobile devices and is compatible with most current web browsers. The bottom layer relies on a vector graphics tool, ZRender, to provide intuitive, interactive, and highly personalized data visualization [26]. ECharts uses incremental rendering technology (4.0+) with various detailed optimizations to display tens of millions of data points, and it can still perform smooth zoom and pan interactions at this data level [26]. ECharts provides legends, visual mapping, data area zoom, tooltips, and data-filtering interactive components, which can perform multidimensional filtering and view zooming and display details [26].

Our research used ECharts to visualize basic data and the results of the graph database algorithms as well as to develop web portals and interactive operating systems to intuitively and dynamically query and analyze epidemic data.

Results

Organizational Process of Digital Contact Tracing During the COVID-19 Epidemic in Hainan Province

Based on the data planning, data collection, data storage, data analysis, and data application processes involved in the data life cycle, we used interview data to summarize the activities performed by various participants in the digital contact tracing process of the COVID-19 epidemic in Hainan Province. The specific organizational process is as follows.

Data Planning Process

Data planning includes organization, assessment of the situation and demand, and the formulation of strategic goals. This project was initiated by the Hainan Provincial Command of COVID-19 Epidemic Prevention and Control on January 29, 2020. The governor of Hainan Province, as the person in charge, instructed various government departments to cooperate with the Big Data Administration to apply big data technology to COVID-19 epidemic emergency management–related work. That is, the Hainan Provincial Command of COVID-19 Epidemic Prevention and Control played a leading and coordinating role, the Hainan Big Data Administration played a leading role in implementation, and other government departments played cooperative roles. Thus, an integrated and flat organizational structure and coordination mechanism were formed.

The Hainan Big Data Administration immediately assessed the COVID-19 epidemic situation and emergency management measures, coordinated communication with the Command of COVID-19 Epidemic Prevention and Control and other relevant government departments, and organized epidemiologists and big data technicians to formulate an implementation plan. The plan determined the specific implementation details of using multisource epidemic big data for digital technology tracing and submitted achievable data requirements to the Command of COVID-19 Epidemic Prevention and Control. Then, the COVID-19 Epidemic Prevention and Control headquarters coordinated the cooperation of work by relevant departments and provided the specified data.

Data Collection Process

Hainan Provincial Big Data Management collected the first batch of data in Excel format (Microsoft Corporation) from the Command of COVID-19 Epidemic Prevention and Control,

Public Security Department, Health Commission, communication operators, and other departments on February 10, 2020.

Data Storage Process

The technicians of the Hainan Provincial Big Data Administration cleaned and merged the data under the guidance of officials and epidemiologists, uploaded all databases to Hainan's big data public service platform, and used the hive data warehouse tool to manage, extract, query, and analyze the data.

Data Analysis Process

The data technicians of the Hainan Provincial Big Data Administration retrieved data from Hainan's big data public service platform; used the graph database Neo4j algorithm to perform association analysis on key populations, vehicles, and public places; and then used ECharts to visualize the data analysis results and determine contacts and high-risk public places.

Data Application Process

The Hainan Provincial Big Data Administration regularly writes COVID-19 epidemic contact tracing reports based on the results of data analysis and reports them to the Command of COVID-19 Epidemic Prevention and Control. The Command of COVID-19 Epidemic Prevention and Control promptly releases early warning information to the public; strengthens cross-departmental sharing of information; and guides the Health Commission, CDC, medical and health institutions, and grassroots residents' committees in the implementation of measures, such as isolation, health monitoring, and nucleic acid testing of contacts. Measures are also implemented to limit the flow of people and close high-risk infected public places.

Technical Process of Digital Contact Tracing During the COVID-19 Epidemic in Hainan Province

Application of the Graph Database Algorithm to Determine and Trace Contacts

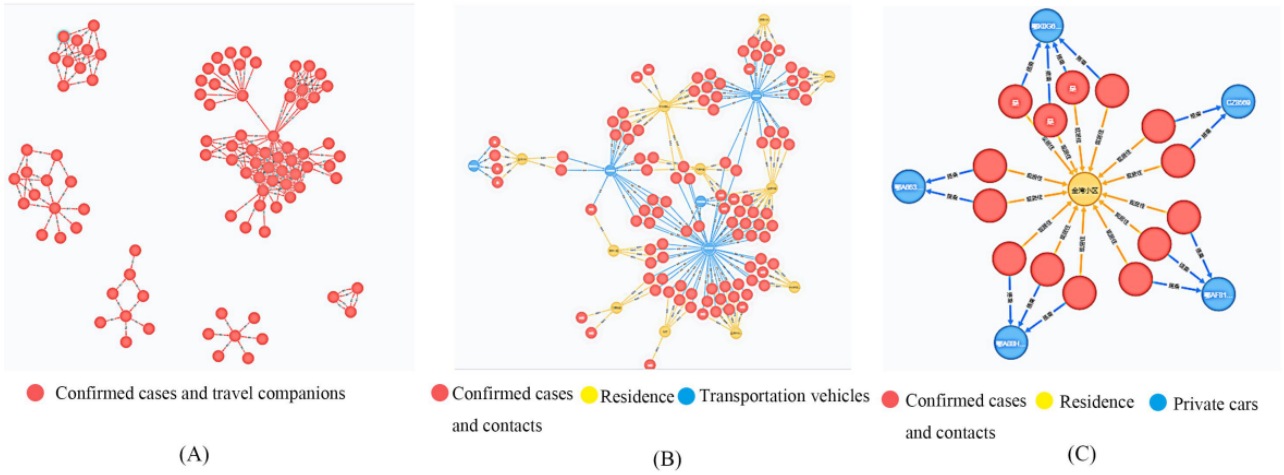
The technicians of the Hainan Provincial Big Data Administration took confirmed, suspected, and asymptomatic infected individuals, as well as individuals with a history of residence in Hubei Province, as the nodes of population; private cars, trains, and flights as the nodes of transportation; and communities and shopping malls as the nodes of public places. The technicians defined a relationship as appearance in the same public place or vehicle at the same time. Based on these nodes and relationships, the Neo4j Cypher algorithm graph database was applied to build an association graph to analyze and display the associations among key populations, vehicles, public places, and other key pieces of information. The core algorithm of the graph database used in our research is shown in Table 2.

Table 2. Hainan provincial digital contact tracing core algorithm based on a graph database.

Algorithm function	Specific algorithm	Algorithm description
Tracing the travel companions of confirmed, suspected, and asymptomatic infected individuals	MATCH p=(n:Individual)-[r:'sameTransportation']-(n1)-[rr:'sameTransportation']-(n2) where n.name<>n2.name and upper(n.ID) in ['ID1','ID2','ID3','ID4','ID...'] RETURN distinct n.name,upper(n.ID),n.phoneNo,n1.name,upper(n1.ID),n1.phoneNo,n2.name,upper(n2.ID),n2.phoneNo;	"Individual" means the individuals included in the analysis; "sameTransportation" means taking one mode of transportation at the same time; "Name" means the name of the analyzed objects; "ID" is the ID number, where "ID/ID1/ID2/ID3/ID4/ID..." are the ID numbers of confirmed, suspected and asymptomatic infected individuals;"phoneNo" is the mobile phone number.
Tracing individuals who have had contact with >2 confirmed infected individuals	MATCH p=(n:'Individual')-[r1]->()<-[r2]-(nm)-[r21]->()<-[r22]-(m:'Individual') where n.is_confirmedindividual in ['yes','sameID'] and m.is_confirmedindividual in ['yes','sameID'] and EXISTS(nm.is_confirmedindividual)=false and n.ID<>m.ID return p;	"confirmedindividual" means a confirmed individual; "sameID" means the same ID number.
Tracing contacts through transportation modes with high risk of infection	match p=(n:Transportation)-[*..4]-() where n.name='CarNo1' return p;	"Transportation" means the transportation included in the analysis (private car, train, airplane); "CarNo1" is the license plate number of a certain private car.

The representative analysis results of the graph database algorithm are shown in Figure 2. Figure 2A shows the close contacts of confirmed cases who rode in the same vehicle and the second-degree and higher-degree contacts associated with the abovementioned close contacts in other vehicles. Figure 2B shows the close, second-degree, and higher-degree contacts associated with >2 confirmed cases through any means of transportation or public place. Figure 2C shows the close contacts associated with confirmed cases through private cars and communities. The Neo4j software package supports the export of the information on the close, second-degree, and higher-degree contacts in Excel format so that the authorities can implement quarantine measures on the contacts.

Figure 2. Contact tracing based on the algorithm of the Neo4j software package for the graph database: (A) contacts associated with travel, (B) contacts associated with >2 confirmed cases, and (C) contacts associated with private cars.

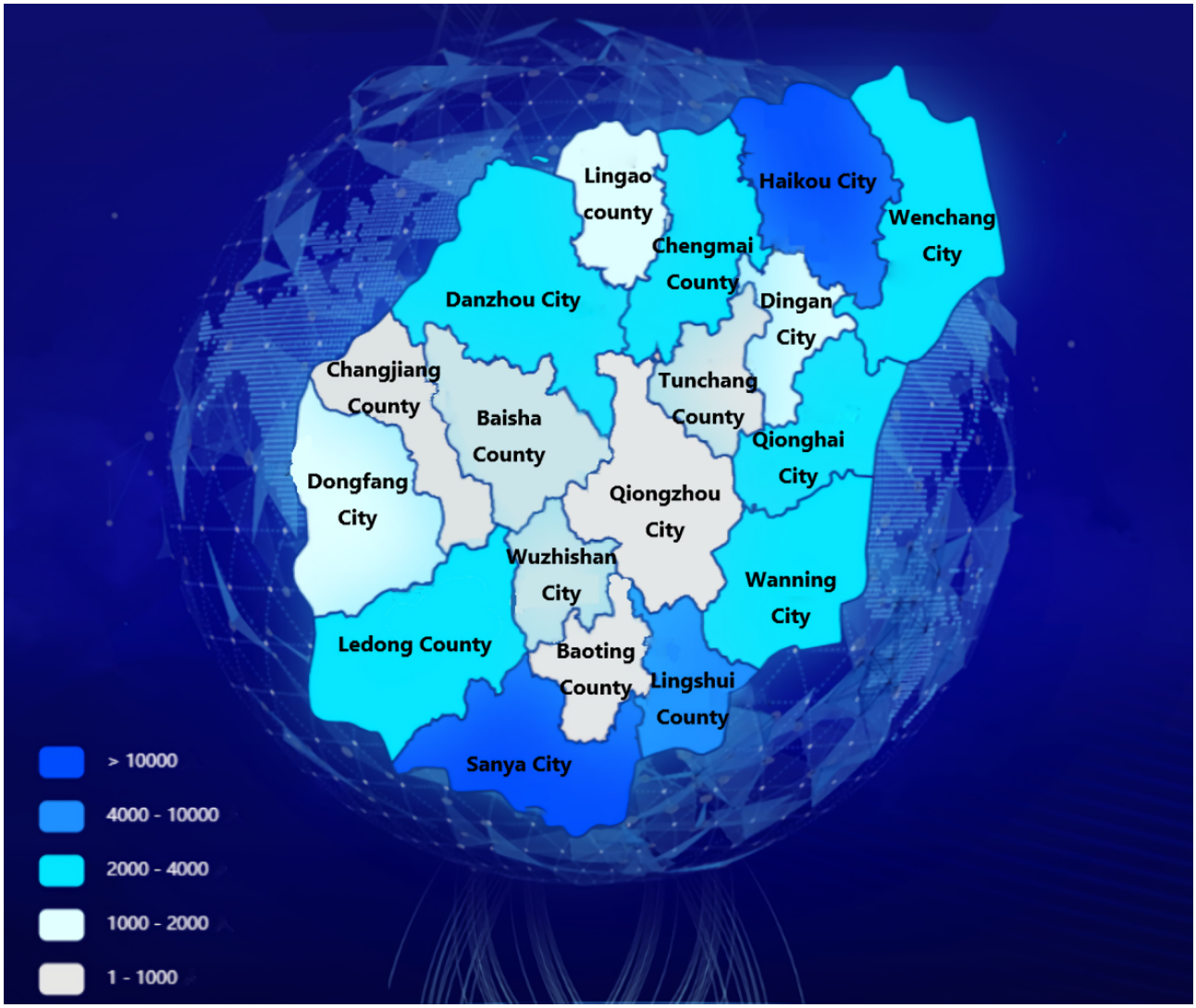


Application of the Graph Database Algorithm to Determine Public Places With High Risk of Infection

The Hainan Provincial Big Data Administration applied ECharts data visualization tools to compute and visualize the distribution of the population with a residence history in Hubei Province in all cities of Hainan Province. As shown in Figure 3, Haikou, the administrative center of Hainan Province, and the well-known tourist city, Sanya, have the largest numbers of imported residents with a history of residence in Hubei Province; both populations were greater than 10,000. Overall, the population of individuals in Hainan Province with residence history in Hubei Province was more concentrated in coastal cities than in central cities.

The Hainan Provincial Big Data Administration applied the graph database algorithm to rank the frequency of confirmed, suspected, and asymptomatic infected individuals and their contacts appearing in shopping malls and communities. The results were used to infer and predict public places with high risk of infection.

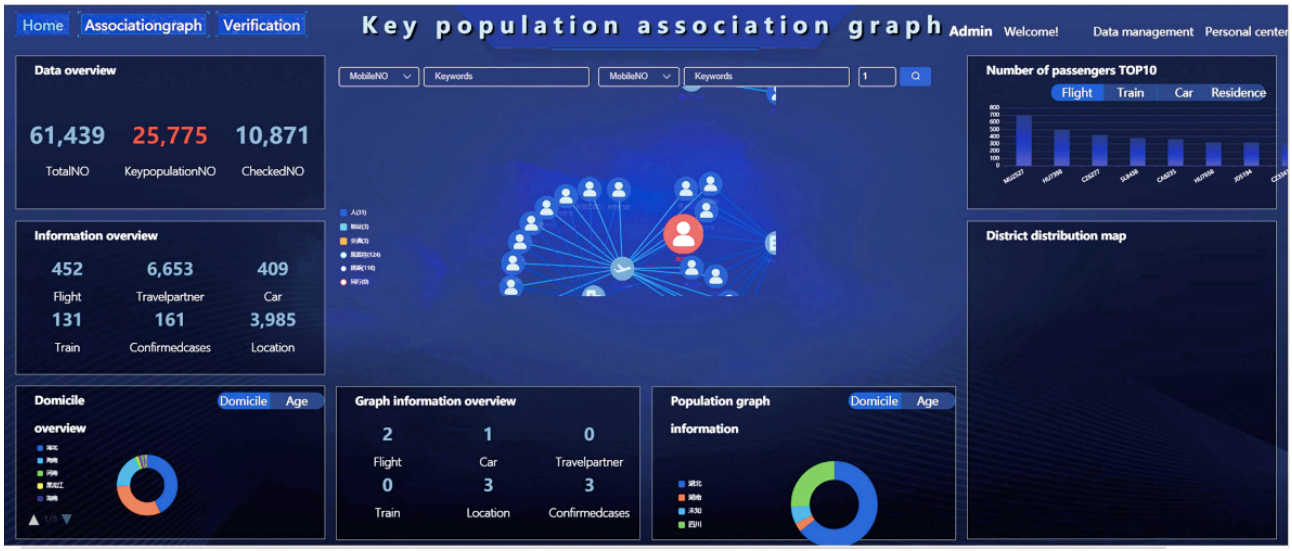
Figure 3. Distribution of the population of residents of Hainan Province who have a history of residence in Hubei Province.



Application of ECharts to Design Data Visualization and an Interactive Operating System

The Hainan Provincial Big Data Administration used ECharts to visualize the basic data and analysis results and also designed an interactive operating system that can be accessed through a web page on a browser. The system interface is shown in Figure 4. The interactive operating system displays the total amount of data processed, the overview of epidemic prevention and control information, the statistics of the permanent residence of the imported persons from Hubei Province, the overview of the graph information, the information of the persons involved in the graph, and the distribution of high-risk infected individuals in transportation vehicles and public places. In addition, the system provides a query function of the graph analysis results so that administrative officials and disease prevention and control staff can directly view the graph results by keyword query.

Figure 4. Data visualization and interactive system interface designed with ECharts.



Main Achievements of Digital Contact Tracing During the COVID-19 Epidemic in Hainan Province

The Hainan Provincial Big Data Administration identified 61,439 analysis objects from hundreds of thousands of records and identified 10,871 contacts. The authorities took measures to isolate, transport, and monitor these contacts. Hainan provincial digital contact tracing identified 378 individuals with the highest infection risk (that is, the highest numbers of contacts and exposures), including 106 close contacts and 154 second-degree contacts who traveled on the same vehicles, 110 contacts within the same communities, and 8 contacts within the same malls. The Hainan Provincial Big Data Administration identified 6 high-risk communities and a number of high-risk shopping malls. According to the list of the highest-risk infected individuals and high-risk shopping malls and communities, the Command of Hainan Provincial COVID-19 Epidemic Prevention and Control directed health committees, the CDC, medical and health institutions, grassroot governments, and community residents' committees to take mandatory measures (such as nucleic acid testing, isolation, and intensive treatment) for the highest-risk infected individuals and to implement emergency shutdown measures for high-risk public places. As a result, a patient who was not detected by traditional contact tracing was discovered through digital contact tracing, which provided vital information support to comprehensively curb the spread of the COVID-19 epidemic from the source.

Discussion

Main Findings

We summarized the organizational process, technical process, and main achievements of the graph database algorithm in tracing the contacts of COVID-19 cases in Hainan Province, China. This approach has practical importance in overcoming the limitations of traditional emergency management measures and existing digital contact tracing methods. From the perspective of the organizational process, our research found that Hainan Province formed a scientific and effective organizational structure and operating mechanism for digital contact tracing during the COVID-19 epidemic. The success of the organizational process is attributed to (1) the establishment of a special high-level administrative leadership agency to coordinate the entire process, namely, the Hainan Provincial Command of COVID-19 Epidemic Prevention and Control; (2) the clarification of the rights and obligations of relevant emergency management agencies in the digital contact tracing of the COVID-19 epidemic; (3) the formation of a flat and integrated organizational structure and operating mechanism for multiple agencies to communicate effectively in a timely manner and achieve collaborative governance; (4) the establishment of a dedicated big data management department to provide technical support required for data lifecycle management; and (5) the establishment of a government big data public service platform that supports the storage, recall, and analysis of multisource data according to the centralized model, and the application of the platform in digital contact tracing during the COVID-19 epidemic.

From the perspective of the technical process, our research is based on the use of the Neo4j graph database algorithm and ECharts data visualization tool to mine and analyze multisource COVID-19 epidemic big data. Hidden contacts and public places associated with confirmed patients, suspected patients, and asymptomatic infections were discovered and traced so that contacts and high-risk public places could be accurately identified. The results show that the digital contact tracing in Hainan Province is compatible with multisource heterogeneous epidemic big data; it can quickly and accurately find close, second-degree, or third-degree contacts, and it can identify public places with high risk of infection. Visualization technology is of great importance in optimizing public decision-making [27]. In our research, the interactive operating system for data visualization developed based on ECharts helps COVID-19 epidemic emergency management officials to simply and directly query populations and public places with high risk of infection, and it can dynamically monitor the epidemic situation to provide decision support for evidence-based emergency management strategies for the COVID-19 epidemic. The digital contact tracing system in Hainan Province can efficiently detect a large number of people and accurately locate

the contacts of individuals with COVID-19, which overcomes the inefficiency and low completeness of traditional manual contact tracing and helps comprehensively curb the spread of the COVID-19 epidemic from the source.

Application and Practice Prospects

Our research is expected to overcome the main difficulties faced in emergency management of the COVID-19 epidemic in different countries and regions due to their unique political, cultural, and civic concepts [18,19,28-31]. Our case study in Hainan Province found that digital contact tracing based on a graph database algorithm has high flexibility and agility in responding to the COVID-19 epidemic and may have greater application prospects in the aforementioned countries and regions. The majority of countries and regions, especially countries with more liberal policies, require more rapid and accurate strategies to identify and trace populations and public places with high risk of infection to adopt precise quarantine and emergency measures for these people and places instead of adopting mandatory measures for all citizens. Because of their agility and responsiveness, collaboration, communication, and evidence-based digital contact tracing have important applications in the early, middle, and late stages of the COVID-19 epidemic and in the small-scale rebound after confirmed cases are cleared. In short, our research results have broad application prospects for different countries and regions and in different stages of the COVID-19 epidemic.

Limitations and Improvements

The main limitation of our research is the small coverage and insufficient information of COVID-19 epidemic data. Thus, tracing high-risk infected populations among the entire population of Hainan Province and their contacts is impossible. The participants in digital contact tracing in Hainan Province are limited to people with a history of residence in Hubei Province. This selection greatly reduces the scope of digital contact tracing. In addition, WeChat positioning–based GPS data, UnionPay consumption data, and railway and flight passenger information, which can accurately reflect the trajectory of residents, are owned by enterprises. These enterprises did not agree to provide users' spatiotemporal trajectory data to the Hainan Provincial Big Data Administration to protect the privacy of their users.

Hainan Province needs to further improve the accuracy of data contact tracing using a graph database algorithm. Authorities need to adopt more comprehensive, accurate, and dynamic population spatiotemporal trajectory data to combat the epidemic. These data are the key to improve the accuracy of digital-based tracing. In addition, when confirmed, suspected, and asymptomatic infected individuals come into contact with the general population, the distance and duration of exposure are the main indicators for assessing the risk of cross-infection [10]. Hainan Province treats people who appeared in the same mode of transportation, shopping malls, or communities with confirmed, suspected, or asymptomatic infected individuals at the same time as contacts. This consideration lacks a threshold for exposure distance and duration and has a negative impact on the accuracy of contact tracing. Therefore, the digital contact tracing based on the graph database algorithm in Hainan Province needs to determine the thresholds of exposure distance and duration in different scenarios based on the professional opinions of experts in disease prevention and control and of medical staff. Digital technicians need to optimize the algorithm to improve the precision of digital contact tracing based on the thresholds of exposure distance and duration.

Furthermore, security issues in the process of digital contact tracing of epidemics based on a digital government platform under the centralized model need to be properly resolved in two aspects: governance and technology. In terms of governance, the law, system, and data management structure must be improved for the government to obtain citizens' personal information in public health emergencies. Data sharing standards, authority management, and data security management in the process of epidemic data analysis and application, as the bases for the government to adopt useful data to respond to epidemics under the premise of ensuring the privacy of citizens, should be clarified. In addition, the government's use of data from enterprises, social organizations, and other nongovernmental organizations to respond to the epidemic needs to be further clarified and improved. In terms of technology, data security technologies, such as cryptography methods, data encryption technologies, system vulnerability monitoring and repair technologies, virus-killing technologies, and automatic data deletion technologies, must be comprehensively used in the digital platform for epidemic emergency management.

Finally, the methods of collecting epidemic data were not sufficiently intelligent. For example, epidemiological investigation data were mainly collected by health personnel through face-to-face interviews with the cooperation of community resident committees. These data were manually entered into the information system by the staff of the CDC or medical institutions. This data collection method was relatively inefficient and requires substantial staffing. Furthermore, onsite investigation by staff increases the risk of cross-infection. Therefore, for digital contact tracing, it is necessary to develop more intelligent means of completion. Therefore, authorities should follow the development trend of governance in the digital age; develop an epidemic emergency management big data platform with the aid of governmental big data platforms; and fully use cloud transmission, cloud storage, and cloud computing technology to realize the real-time and dynamic collection of epidemic data from the transaction information systems of public and private agencies through a network interface to replace the traditional methods of collecting multisource epidemic data.

Conclusion

Hainan Province, China, was selected as the case for this study. The use of a graph database algorithm and ECharts visualization tools in Hainan Province to trace contacts and identify high-risk public places was summarized in a centralized model during the process of emergency management of the COVID-19 epidemic. Moreover, the organizational and technical processes of the case we studied can help government agencies to implement precise emergency management measures and provide agile, evidence-based decision support to improve the responsiveness of the public health emergency response system. The organizational arrangements and technology involved in

our research can be applied in different countries and regions to respond to the COVID-19 epidemic and can be applied at different stages of the COVID-19 epidemic. Future research should focus on solving the problems of insufficient data coverage information, insufficient contact tracing accuracy, insufficient intelligence of data collection methods, security concerns, and insufficient real-time performance of data sharing caused by the tradeoff between citizen privacy and public health security to optimize digital contact tracing.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China, the National Social Science Fund of China and the Fundamental Research Funds for the Central Universities for funding this study; we also thank all the officers and professionals from the Hainan Provincial Government who were involved in this study. We also acknowledge the reviewers and editors of the Journal of Medical Internet Research for improving this study. This work was supported by the National Natural Science Foundation of China (grant numbers: 71734002, 72042016), the National Social Science Fund of China (grant number: 20ZDA038), and the Fundamental Research Funds for the Central Universities, HUST (grant numbers: 2020kfyXGYJ022, 2020JYCXJJ035). The funders had no role in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript.

Authors' Contributions

ZM contributed to the conceptualization, methodology, and writing of the original draft. HY contributed to the data curation, investigation, visualization, and writing of the original draft. QZ was involved in the investigation, methodology, writing, reviewing, and editing. WZ was involved in the writing, reviewing, and project administration. YD contributed to the writing and editing and supervised the study. All authors revised the manuscript.

Conflicts of Interest

None declared.

References

1. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491) [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
2. Cohen IG, Gostin LO, Weitzner DJ. Digital Smartphone Tracking for COVID-19: Public Health and Civil Liberties in Tension. JAMA 2020 Jun 16;323(23):2371-2372. [doi: 10.1001/jama.2020.8570] [Medline: 32459289]
3. Tibbetts JH. Researchers Continue Quest to Contain Spread of COVID-19. BioScience 2020 Aug 11;70(8):633-639. [doi: 10.1093/biosci/biaa071]
4. Huang Z, Guo H, Lee Y, Ho EC, Ang H, Chow A. Performance of Digital Contact Tracing Tools for COVID-19 Response in Singapore: Cross-Sectional Study. JMIR mHealth uHealth 2020 Oct 29;8(10):e23148 [FREE Full text] [doi: 10.2196/23148] [Medline: 33006944]
5. Martin T, Karopoulos G, Hernández-Ramos J, Kambourakis G, Nai Fovino I. Demystifying COVID-19 Digital Contact Tracing: A Survey on Frameworks and Mobile Apps. Wirel Commun Mob Comput 2020 Oct 17;2020:1-29. [doi: 10.1155/2020/8851429]
6. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: 10.1056/NEJMe2009758] [Medline: 32329972]
7. Wang S, Ding S, Xiong L. A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. JMIR Mhealth Uhealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]
8. Hargreaves S, Kumar BN, McKee M, Jones L, Veizis A. Europe's migrant containment policies threaten the response to covid-19. BMJ 2020 Mar 26;368:m1213. [doi: 10.1136/bmj.m1213] [Medline: 32217531]
9. Moon MJ. Fighting COVID-19 with Agility, Transparency, and Participation: Wicked Policy Problems and New Governance Challenges. Public Adm Rev 2020 May 20;80(4):651-656 [FREE Full text] [doi: 10.1111/puar.13214] [Medline: 32836434]
10. Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ 2020 Jun 15;192(24):E653-E656 [FREE Full text] [doi: 10.1503/cmaj.200922] [Medline: 32461324]
11. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
12. Nuzzo A, Tan CO, Raskar R, DeSimone DC, Kapa S, Gupta R. Universal Shelter-in-Place Versus Advanced Automated Contact Tracing and Targeted Isolation: A Case for 21st-Century Technologies for SARS-CoV-2 and Future Pandemics. Mayo Clin Proc 2020 Sep;95(9):1898-1905 [FREE Full text] [doi: 10.1016/j.mayocp.2020.06.027] [Medline: 32861334]

13. Wirth FN, Johns M, Meurers T, Prasser F. Citizen-Centered Mobile Health Apps Collecting Individual-Level Spatial Data for Infectious Disease Management: Scoping Review. JMIR mHealth uHealth 2020 Nov 10;8(11):e22594 [FREE Full text] [doi: 10.2196/22594] [Medline: 33074833]
14. Hernández-Orallo E, Calafate CT, Cano J, Manzoni P. Evaluating the Effectiveness of COVID-19 Bluetooth-Based Smartphone Contact Tracing Applications. Appl Sci 2020 Oct 13;10(20):7113. [doi: 10.3390/app10207113]
15. Nakamoto I, Jiang M, Zhang J, Zhuang W, Guo Y, Jin M, et al. Evaluation of the Design and Implementation of a Peer-To-Peer COVID-19 Contact Tracing Mobile App (COCOA) in Japan. JMIR mHealth uHealth 2020 Dec 01;8(12):e22098 [FREE Full text] [doi: 10.2196/22098] [Medline: 33170801]
16. Lin L, Hou Z. Combat COVID-19 with artificial intelligence and big data. J Travel Med 2020 Aug 20;27(5) [FREE Full text] [doi: 10.1093/jtm/taaa080] [Medline: 32437541]
17. Cohen IG, Gostin LO, Weitzner DJ. Digital Smartphone Tracking for COVID-19: Public Health and Civil Liberties in Tension. JAMA 2020 Jun 16;323(23):2371-2372. [doi: 10.1001/jama.2020.8570] [Medline: 32459289]
18. Primc K, Slabe-Erker R. The Success of Public Health Measures in Europe during the COVID-19 Pandemic. Sustainability 2020 May 25;12(10):4321. [doi: 10.3390/su12104321]
19. Gerstein D. Assessing the US government response to the coronavirus. Bull At Sci 2020 Jul 20;76(4):166-174. [doi: 10.1080/00963402.2020.1778356]
20. Real-time status of the novel coronavirus pneumonia in Hainan Province. Webpage in Chinese. The Health Commission Of Hainan Province. 2020 Aug 09. URL: http://wst.hainan.gov.cn/yqfk/ [accessed 2021-01-19]
21. Regulations on the Development and Application of Big Data in Hainan Province. Webpage in Chinese. Hainan Provincial People's Congress. 2019 Oct 5. URL: http://www.hainan.gov.cn/hainan/dfxfg/201910/9c18c64f369c4faab6f0dded3c04adea.shtml [accessed 2021-01-19]
22. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. Int J Environ Res Public Health 2020 May 02;17(9) [FREE Full text] [doi: 10.3390/ijerph17093176] [Medline: 32370204]
23. Castelltort A, Laurent A. Fuzzy Queries over NoSQL Graph Databases: Perspectives for Extending the Cypher Language. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2014. Communications in Computer and Information Science, vol 444. 2014 Presented at: 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU); July 15-19, 2014; Montpelier, France p. 384-395. [doi: 10.1007/978-3-319-08852-5_40]
24. Giabelli A, Malandri L, Mercorio F, Mezzanzanica M. GraphLMI: A data driven system for exploring labor market information through graph databases. Multimed Tools Appl 2020 Jun 29. [doi: 10.1007/s11042-020-09115-x]
25. Tang X, Ma C, Yu M, Liu C. A Visualization Method Based on Graph Database in Security Logs Analysis. 2017 Presented at: 6th International Conference on Advanced Materials and Computer Science (ICAMCS); April 29-30, 2017; Zhengzhou, China p. 82-89. [doi: 10.23977/icamcs.2017.1012]
26. ECharts. Apache. 2020 Aug 05. URL: https://echarts.apache.org/zh/index.html [accessed 2021-01-19]
27. Moody R, Bekkers V. Visualizations, technology and the power to influence policy. Gov Inf Q 2018 Sep;35(3):437-444. [doi: 10.1016/j.giq.2018.06.004]
28. Li C, Romagnani P, von Brunn A, Anders H. SARS-CoV-2 and Europe: timing of containment measures for outbreak control. Infection 2020 Jun;48(3):483-486 [FREE Full text] [doi: 10.1007/s15010-020-01420-9] [Medline: 32274649]
29. Tanne J. Covid-19: US cases soar as Trump pushes for schools to open. BMJ 2020 Jul 10;370:m2803. [doi: 10.1136/bmj.m2803] [Medline: 32651239]
30. Habayeb S. Epidemics and Pandemics: Coronavirus Disease in the United States, Lessons Learned and Way Forward. World Med Health Policy 2020 Jul 29;12(3):218-222. [doi: 10.1002/wmh3.354]
31. Nakamoto I, Wang S, Guo Y, Zhuang W. A QR Code-Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity. JMIR mHealth uHealth 2020 Sep 07;8(9):e22321 [FREE Full text] [doi: 10.2196/22321] [Medline: 32841151]

Abbreviations

QR: quick response
Wi-Fi: Wireless Fidelity
CDC: Center for Disease Control and Prevention

Edited by G Eysenbach; submitted 30.12.20; peer-reviewed by S Pesälä, A Benis; comments to author 09.01.21; revised version received 14.01.21; accepted 14.01.21; published 22.01.21.

Please cite as:
Mao Z, Yao H, Zou Q, Zhang W, Dong Y
Digital Contact Tracing Based on a Graph Database Algorithm for Emergency Management During the COVID-19 Epidemic: Case Study
JMIR Mhealth Uhealth 2021;9(1):e26836
URL: http://mhealth.jmir.org/2021/1/e26836/
doi:10.2196/26836
PMID:33460389

Original Paper

The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study

Michael Sobolev[1,2], PhD; Rachel Vitale[3], BA; Hongyi Wen[1], BSc, MSc; James Kizer[1], BSc, MSc; Robert Leeman[4], PhD; J P Pollak[1], PhD; Amit Baumel[5], PhD; Nehal P Vadhan[2], PhD; Deborah Estrin[1], PhD; Frederick Muench[3], PhD

[1]Cornell Tech, Cornell University, New York City, NY, United States

[2]Feinstein Institute for Medical Research, Northwell Health, Great Neck, NY, United States

[3]The Partnership to End Addiction, New York, NY, United States

[4]College of Health and Human Performance, Department of Health Education and Behavior, University of Florida, Gainsville, FL, United States

[5]University of Haifa, Haifa, Israel

Corresponding Author:
Frederick Muench, PhD
The Partnership to End Addiction
485 Lexington Avenue
3rd Floor
New York, NY, 10017
United States
Phone: 1 9175320623
Email: fmuench@toendaddiction.org

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2021/1/e27439/

Abstract

Background: The classic Marshmallow Test, where children were offered a choice between one small but immediate reward (eg, one marshmallow) or a larger reward (eg, two marshmallows) if they waited for a period of time, instigated a wealth of research on the relationships among impulsive responding, self-regulation, and clinical and life outcomes. Impulsivity is a hallmark feature of self-regulation failures that lead to poor health decisions and outcomes, making understanding and treating impulsivity one of the most important constructs to tackle in building a culture of health. Despite a large literature base, impulsivity measurement remains difficult due to the multidimensional nature of the construct and limited methods of assessment in daily life. Mobile devices and the rise of mobile health (mHealth) have changed our ability to assess and intervene with individuals remotely, providing an avenue for ambulatory diagnostic testing and interventions. Longitudinal studies with mobile devices can further help to understand impulsive behaviors and variation in state impulsivity in daily life.

Objective: The aim of this study was to develop and validate an impulsivity mHealth diagnostics and monitoring app called Digital Marshmallow Test (DMT) using both the Apple and Android platforms for widespread dissemination to researchers, clinicians, and the general public.

Methods: The DMT app was developed using Apple's ResearchKit (iOS) and Android's ResearchStack open source frameworks for developing health research study apps. The DMT app consists of three main modules: self-report, ecological momentary assessment, and active behavioral and cognitive tasks. We conducted a study with a 21-day assessment period (N=116 participants) to validate the novel measures of the DMT app.

Results: We used a semantic differential scale to develop self-report trait and momentary state measures of impulsivity as part of the DMT app. We identified three state factors (inefficient, thrill seeking, and intentional) that correlated highly with established measures of impulsivity. We further leveraged momentary semantic differential questions to examine intraindividual variability, the effect of daily life, and the contextual effect of mood on state impulsivity and daily impulsive behaviors. Our results indicated validation of the self-report sematic differential and related results, and of the mobile behavioral tasks, including the Balloon

Analogue Risk Task and Go-No-Go task, with relatively low validity of the mobile Delay Discounting task. We discuss the design implications of these results to mHealth research.

Conclusions: This study demonstrates the potential for assessing different facets of trait and state impulsivity during everyday life and in clinical settings using the DMT mobile app. The DMT app can be further used to enhance our understanding of the individual facets that underlie impulsive behaviors, as well as providing a promising avenue for digital interventions.

Trial Registration: ClinicalTrials.gov NCT03006653; https://www.clinicaltrials.gov/ct2/show/NCT03006653

(JMIR Mhealth Uhealth 2021;9(1):e25018) doi:10.2196/25018

KEYWORDS

impulse control; impulsivity; self-regulation; self-control; mobile health; mHealth; ecological momentary assessment; active task; ResearchKit

Introduction

Background

The classic Marshmallow Test performed by Mischel and colleagues [1] determined that the inability to inhibit short-term responding in childhood was predictive of lower educational attainment, lower stress resilience, and higher drug use and BMI in adulthood. In this test, children were offered a choice between one small but immediate reward (eg, one marshmallow) or a larger reward (eg, two marshmallows) if they waited for a period of time. Despite future research suggesting that multiple factors such as socioeconomic status mediated the relationship between delayed gratification and life outcomes [2], the study instigated a wealth of research on the relationships among impulsive responding, self-regulation, and clinical and life outcomes [3-11].

Impulsivity is a multidimensional construct characterized primarily by the inability to inhibit responding for short-term rewards despite long-term negative consequences or loss of potential gains [12-14]. Impulsivity is a common transdiagnostic feature of many disorders in the Diagnostic and Statistical Manual [15]. A plethora of psychological and medical studies have demonstrated the relationship of impulsivity traits to a variety of physical and mental health outcomes [14,16]. Across studies and subtypes, highly impulsive individuals were found to be significantly more likely to suffer from obesity, type II diabetes, substance use disorder, attention deficit/hyperactivity disorder (ADHD), gambling problems, bipolar disorder, borderline personality disorder, and suicidal behaviors, among others [14,16,17]. Levels of impulsivity not only predict the onset of numerous conditions but also the likelihood of successful intervention outcomes [18-20].

Measurement of Impulsive Behavior

Measurement of impulsivity has long been considered challenging in psychological and medical research due to the multidimensional nature and heterogeneous manifestations of the construct [13,14]. Impulsive behavior includes a number of related but distinct types of traits such as positive and negative urgency, lack of planning or premeditation, lack of perseverance, inattention, present and future discounting, response inhibition, and sensation seeking [13]. Evidence suggests that each of the subtypes of impulsivity manifests itself in different ways on self-report and neurobiological and cognitive measurements, and that different types of measurements have strengths and weaknesses in identifying underlying components of impulsive behaviors [21-23].

Consequently, relations between self-report and performance-based assessment are consistently of low magnitude, but are independently associated with cognitions and behaviors [21,24]. For example, a meta-analysis of the relationship between impulsivity and BMI found that performance-based behavioral measures of impulsivity yielded significantly larger effect sizes than questionnaires, and that different domains of impulsivity were independently associated with BMI [25]. Because these measures are not highly correlated but do predict different facets of impulsivity and clinical outcomes, assessment paradigms should include a wide range of assessments with the ability to personalize to the specific clinical context. This assessment methodology will increase diagnostic accuracy by predicting specific underlying facets to advance the science rather than focusing on a single construct of impulsivity [26].

A distinction between impulsivity as a personality type or trait exhibited over time and across contexts versus a temporary state influenced by substances and other stimuli also warrants examination [10]. In general, trait-based personality models of impulsive behavior reveal robust relationships with life outcomes [27] and symptomatology [16]. At the same time, trait-based studies can be confounded by other factors, including environment, mood, cognition, and social setting [2,28,29], and are heavily influenced by current state and context. Consequently, it is important to measure both trait and state impulsivity via self-report and behavioral measures over time to better understand the relationship to clinical outcomes in real-world settings [29]. The majority of trait and behavioral measures of impulsivity were not designed or validated as state measures or for use as part of a frequent monitoring assessment paradigm; however, several initial studies have revealed that impulsive behaviors can be reliably measured in real-world settings using ecological momentary assessment (EMA) and experience sampling [21,29-33].

Mobile Health

Mobile health (mHealth) technology has demonstrated the ability of smartphone apps and sensors to collect data pertaining to individual activity, behavior, symptoms, cognition, and context [34-37]. mHealth research platforms and frameworks, including Apple's ResearchKit (iOS) [38] and Android's ResearchStack [39], provide the opportunity to develop novel and scalable

mHealth studies utilizing a variety of patient-reported and generated data [40,41]. mHealth studies demonstrated the potential of collecting personalized and frequent multimodal data in the lived experience of individuals to enhance the assessment, monitoring, and diagnostics of medical conditions, and to reveal symptom clusters [42,43].

mHealth technology can further advance the science of impulsivity by increasing the accuracy with which impulsivity as a whole can predict negative outcomes such as onset or exacerbation of psychiatric or medical conditions and treatment failure. mHealth apps can greatly facilitate intensive longitudinal studies [44,45] to understand within-subject differences in impulsive behaviors in everyday contexts. Multimodal methods for studying the underlying constructs of impulsivity separately combine behavioral and self-report measures, and include both trait- and state-based methods to enable a more comprehensive and frequent assessment of the facets of impulsivity. Each of these trait and state measures of impulsive behavior can be further personalized and adapted to individuals and to the context of the study. This personalized and modular approach is particularly useful for the study of impulsive behaviors as they are common in clinical trials of physical, medical, and psychological conditions.

To expand the measure of impulsivity, we developed mobile versions of validated laboratory assessments of impulsivity to be performed on a mobile phone along with daily and momentary self-report measures using Apple's ResearchKit (iOS) [38] and Android's ResearchStack [39] mHealth platforms. We combined these measures with traditional self-report and laboratory measures of impulsivity in a comprehensive study called the Digital Marshmallow Test (DMT).

Objective

The primary aim of this study was to advance the science of impulsivity and the study of impulsive behavior by developing and refining a mobile diagnostic and monitoring app using trait- and state-based self-report and performance measurements of the underlying facets of impulsivity. To achieve this goal, we conducted a 21-day intensive longitudinal study measuring facets of impulsivity using the mobile DMT app.

Methods

DMT App

Overall App Design

We developed a mobile monitoring app for remote assessment and monitoring of impulsive behavior called the DMT. The DMT app was developed based on Apple's ResearchKit (iOS) and Android's ResearchStack open source frameworks for developing health research study apps (Figure 1), which allow for researchers to easily develop intuitive and standardized data-collecting mobile apps. The DMT app consists of three main modules: baseline self-report, EMA of the current state, and active behavioral performance tasks (Figure 2).

Figure 1. Digital Marshmallow Test (DMT) mobile apps for Apple (iOS) and Android.

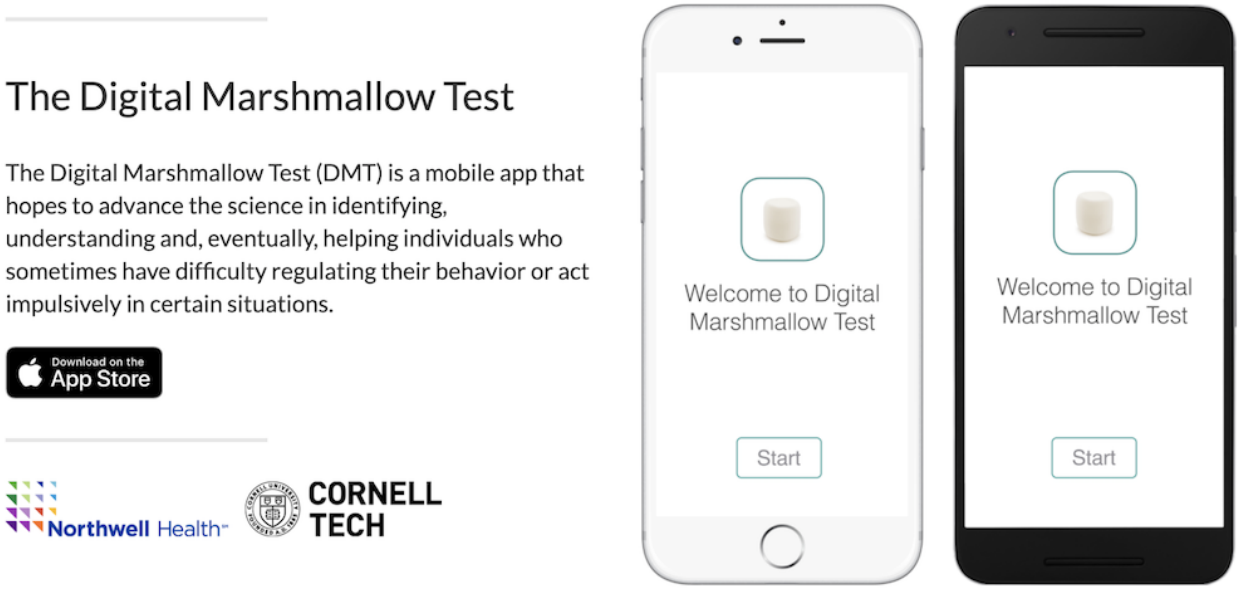
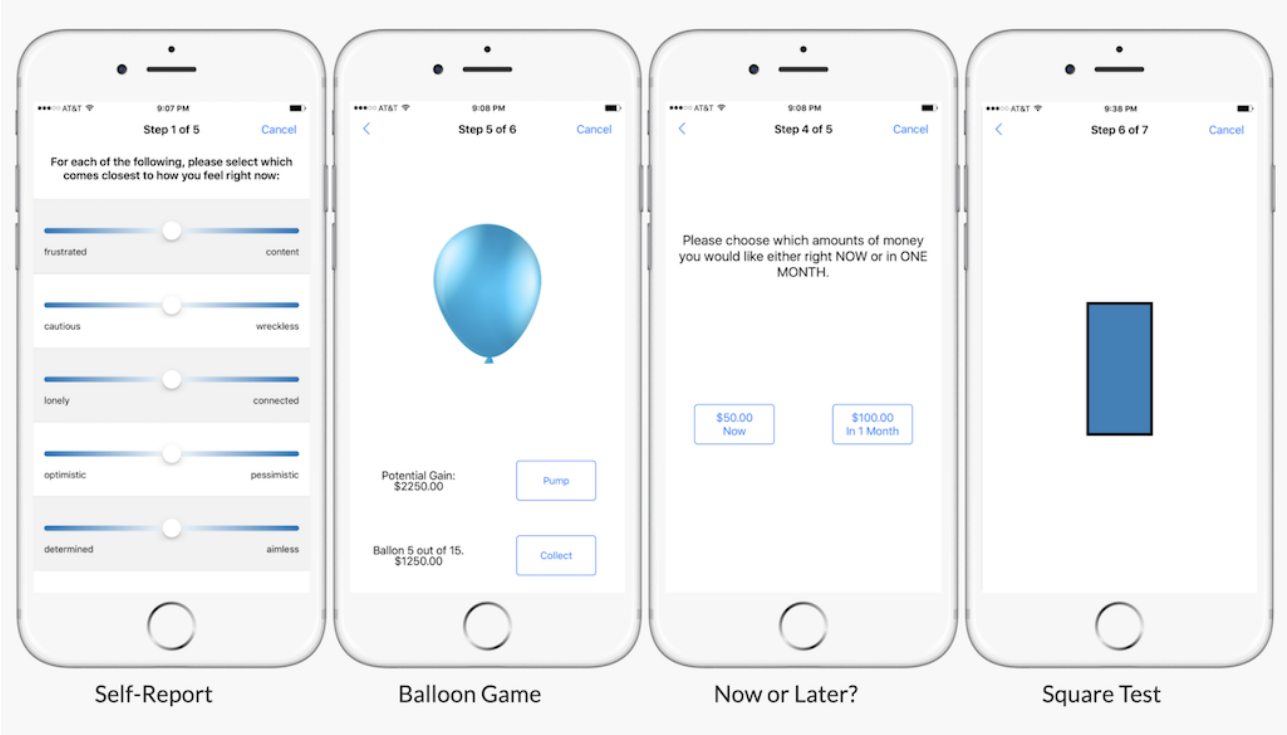


Figure 2. Active performance tasks and self-report in the Digital Marshmallow Test (DMT) app.



Self-Report

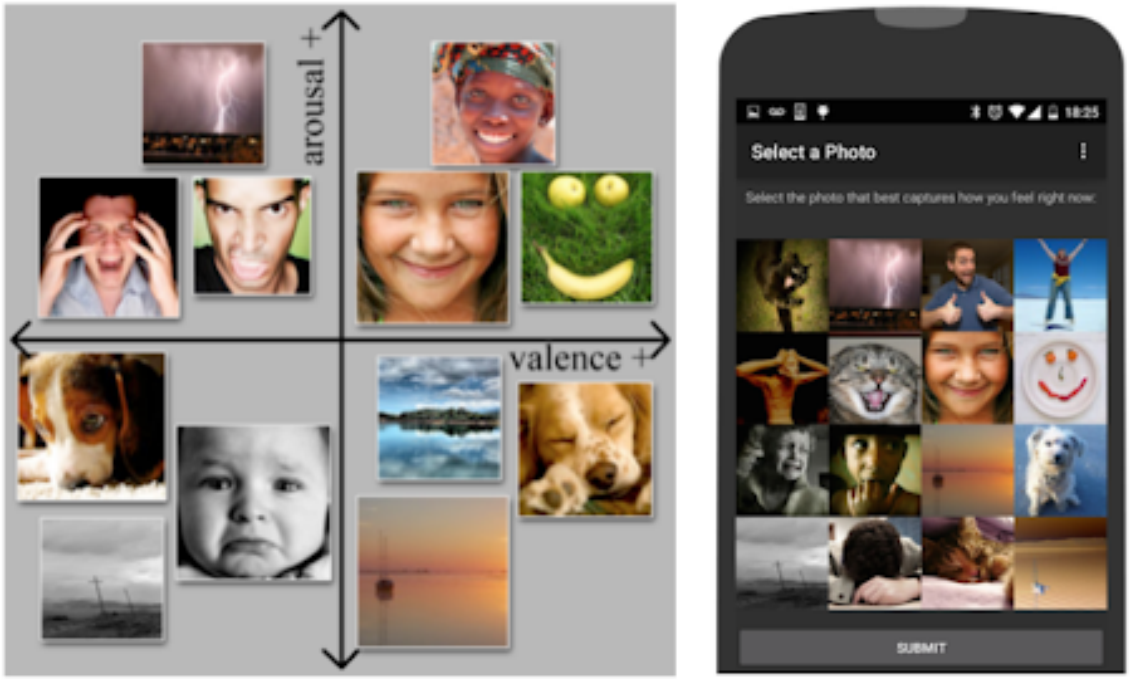
Self-report data and patient-reported outcomes are ubiquitous in behavioral and medical research. Self-report measures of personality and traits are common in assessments of impulsivity in clinical trials and practice [29]. We collected a variety of clinically relevant self-report measures and outcomes using semantic differentials [46], general trait measures of impulsivity [13,47], and daily measures of impulsivity [29] as described below.

EMA

EMA methods involve repeated sampling of subjects' current behaviors and experiences [48]. EMA measures are commonly used in clinical trials and mHealth research. For impulsivity, EMA methodology can be used to understand intraindividual variability and the situational factors of impulsive behavior [29,49]. Our DMT app includes a variety of EMA questions based on the semantic differential scale [46,50] that are prompted in the morning and the evening every day.

We also implemented the Photographic Affect Meter (PAM; Figure 3) to measure emotional state and affect. The PAM is designed for assessing momentary response in which users choose an image that best represents their emotion at a given time [51]. We used the positive and negative affect scores from the PAM that had been validated to correspond to the Positive and Negative Affect Schedule (PANAS) [52].

Figure 3. Photographic Affect Meter (PAM) for ecological momentary assessment.



Active Performance Tasks

Active performance tasks are some of the more innovative parts of Apple's ResearchKit (iOS) [38] and Android's ResearchStack [39] open source frameworks. These tasks invite users to perform activities under partially controlled conditions while phone sensors are used to collect data. ResearchKit [38] includes several predefined documented tasks developed by Apple and the research community [53], which fall into categories such as motor activities, fitness, cognition, and speech. ResearchStack supports a wide variety of community-contributed apps, although at the time of writing there is no centralized listing or repository of these tasks. In the cognition category, one relevant example is the adaptation of the classic Stroop Color and Word Test that is widely used in clinical practice and psychological research [54].

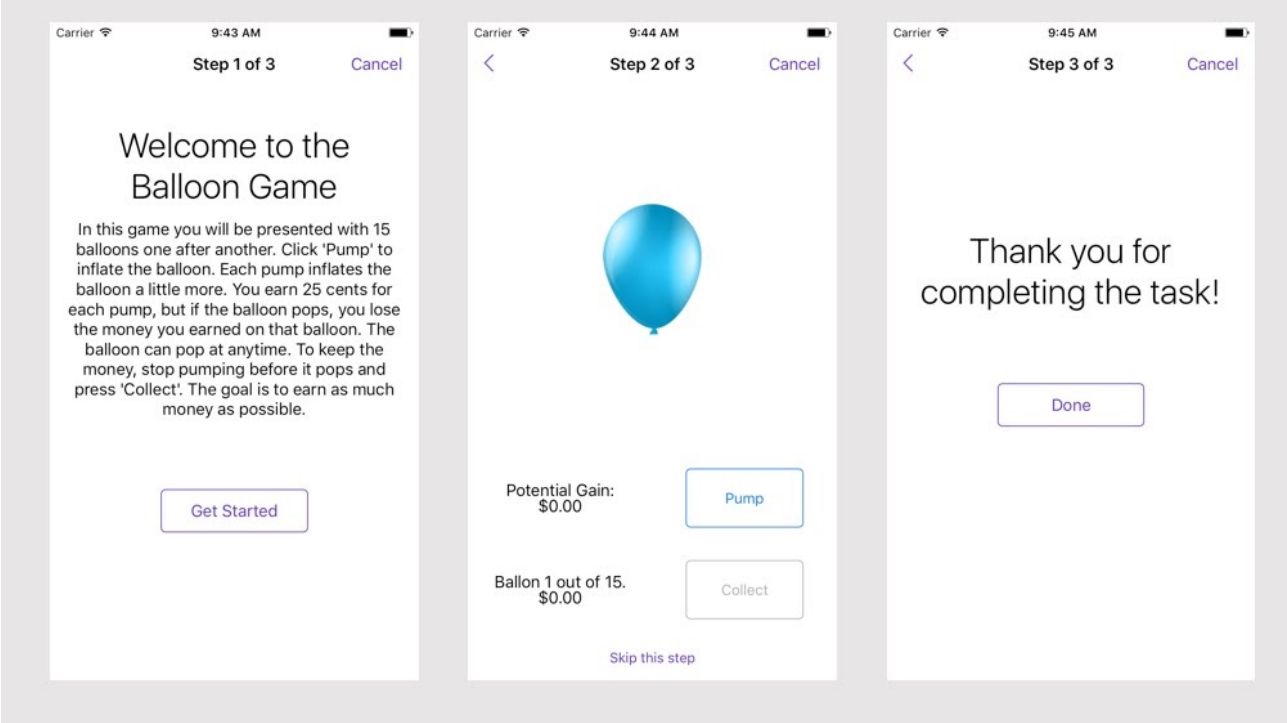
As part of the DMT app and study, we adapted three relevant behavioral and cognitive performance tasks to mobile devices. Specifically, we adapted three laboratory behavioral measures related to impulse control: mobile Balloon Analogue Risk Task (mBART), mobile Go-No-Go (mGNG), and mobile Delay Discounting (mDD). These tasks were modified visually to conform to mobile phone specifications, and were adapted to be used daily to measure behavioral manifestations of impulse control and behavior. For example, the mBART presented users with 15 balloons in each trial and took about 2 minutes to complete (Figure 3). Additional details on the development of active tasks can be found in the DMT project folder at the OSF [55].

ResearchKit and ResearchStack

We developed DMT using an extension of Apple's ResearchKit (iOS) and Android's ResearchStack open source frameworks for developing research study apps, which allow researchers to easily develop intuitive and standardized data-collecting mobile apps. These platforms are designed to meet the requirements of most scientific research, including capturing participant consent, extensible input tasks, and the security and privacy needs necessary for Institutional Review Board approval. The extension was built on top of ResearchKit and ResearchStack, and extends the available surveys to include adaptable visual assessments and custom performance tasks such as the BART, and supports integration of the Ohmage-Omh [56,57] backend out of the box. Other server integrations can easily be created, such as the Sage Bionetworks Bridge Platform [40].

The structure of an app is defined by a JavaScript Object Notation (JSON) file, which specifies the survey or active task steps to be instantiated by the app. The JSON file is converted into an array of Step objects, which the app uses to create a task using the Task Builder that is then presented to the research participant. The results of the task are handled by the Results Processor, which includes modules for storing the results locally and emailing them to the researcher, sending them to the Ohmage-Omh study manager, or sending them to a custom server. For example, to create the mBART for DMT (Figure 4), a researcher would need to create a JSON file [55]. The mBART consists of three steps: (1) an instruction step introducing the study, (2) the mBART active step, and (3) a final instruction step thanking the participant.

Figure 4. Example of an active task: mobile Balloon Analogue Risk Task (mBART).



Testing

Ten beta users tested app functionality between August and November 2016. Both iOS and Android platforms were tested, feedback was provided to the developers, and a second version of the app was released in December 2016. The final version used for the study was released in January 2017.

DMT Study

Participants

Participants were recruited from the Genotype and Phenotype Registry [58], a genetics bank initiated and managed by the Feinstein Institute for Medical Research at Northwell Health [59]. Participants completed a brief anonymous online screening assessment, which indicated whether or not they were eligible to participate in the study. Eligible participants then completed a phone call with a researcher, which involved a general overview of informed consent and scheduling a one-time in-person appointment at the Northwell Health lab. All study data were sent to a HIPAA (Health Insurance Portability and Accountability Act)-compliant database server provided by Sage Bionetworks. This study was approved by the Feinstein Institute of Medical Research within the Northwell Health Institutional Review Board [55].

Eligibility criteria for this study consisted of being fluent in, and able to read, English at the eighth grade level, being between the ages of 18 and 75 years, and owning a smartphone. Individuals who reported serious mental or physical health concerns as evidenced by current treatment or threshold symptoms over the past year were not included in outreach emails. Mental health concerns included any form of psychosis or psychotic disorder, bipolar disorder, and major depression. Participants received US $50 for their baseline interview, US $25 for completing at least 80% of their morning and evening assessments, and US $25 for the day-21 survey.

The total sample size was 116, with 63.8% (n=74) women and a mean age of 44.7 (SD 13.92) years. Overall, 70.7% (82/116) of the participants identified as White, 10.3% (12/116) as Black/African American, 7.8% (9/116) as Hispanic/Latinx, and 11.2% (13/116) as other. The sample was mixed in terms of education, with 36.2% (42/116) having less than a college degree, 27.6% (32/116) having a college degree, and 36.2% (42/116) having a graduate degree. Among the 116 participants, 85 (73.3%) were employed and 58 (50.0%) were married. The average BMI was 28.1 (SD 6.86) kg/m^2. Attrition was relatively low compared to other mHealth studies [60]. Of the 116 participants recruited, 104 (89.7%) completed the mobile baseline assessment, 100 (86.2%) completed at least one morning and one evening assessment, and 93 (80.2%) completed the day-21 assessment.

Procedure

During the in-person appointment at the Northwell Health lab, participants were able to address any concerns pertaining to the study, including smartphone usage and privacy. The appointment was then divided into three parts. First, subjects completed the standard self-report and behavioral measures on a computer (see below). In the second part of the appointment, participants were instructed on how to download the DMT app onto their smartphones and were shown a 5-minute training video on how to use the app, as well as what was expected of their participation throughout the 21-day study. After participants watched the training video and had the opportunity to ask any questions, they completed the baseline assessment on the DMT app. Additional details on the procedure and materials of the study can be found in the DMT project folder at the OSF [55].

Laboratory Assessments

Trait Self-Report

We used two of the most common generalized impulsivity questionnaires: the Barratt Impulsiveness Scale (BIS) and the Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency Impulsive Behavior Scale (UPPS).

The BIS is the most widely cited instrument for the assessment of impulsiveness, and has been used to advance understanding of this construct and its relationship to other clinical phenomena [47,61]. We used a shorter version called BIS-15 [62]. The BIS-15 scale measures three aspects of impulsivity: attention (inability to focus attention or concentrate), motor (to act without thinking), and nonplanning (lack of future orientation or forethought).

The UPPS Impulsive Behavior Scale [13] assesses impulsivity on the subscales of urgency (acting rashly under conditions of negative affect), lack of premeditation (difficulty in thinking and reflecting on consequences of an act), lack of perseverance (inability to remain focused on a task), and sensation seeking (tendency and openness to try and enjoy exciting or dangerous activities). These subscales have a heterogeneous relationship with psychopathology [16]

Behavioral and Cognitive Active Performance Tasks

We used validated standard versions of three behavioral measures commonly used to measure impulse control and related constructs: BART, GNG, and DD. These measures are computerized and were performed at the Northwell Health lab.

BART is a measure of risk-taking that requires individuals to balance the potential for reward and loss via repeated opportunities to earn virtual money by pumping a balloon [63]. The standard BART has been found to predict risk-taking behavior, substance misuse, gambling, and unhealthy eating [63,64]. We used Inquisit software [65] with a script to measure impulsivity and risk aversion based on Lejuez et al [63], which has been validated in previous studies. Participants were presented with 30 balloons, one at a time. For each balloon they had the opportunity to repeatedly pump up the balloon to increase their potential hypothetical earnings, or to stop pumping and collect their accumulated earnings. However, if the balloon pops, the participant loses all of their potential winnings for the current balloon. The average number of pumps for unexploded balloons is the main dependent variable in this paradigm, with higher numbers indicating increased risk-taking. The standard laboratory BART task takes approximately 7 minutes to complete. We recorded the average number of pumps across all trials as a measure of risk-taking proclivity [63].

GNG is a measure of behavioral inhibition and cognitive control. Studies have found that individuals with ADHD display worse inhibitory control compared to controls [66]. GNG performance also differs between healthy controls and substance users or individuals with disordered eating [67,68]. We used Inquisit software [65] with an implemented procedure based on Fillmore et al [69]. Participants were asked to press the spacebar when they see a green rectangle (=go) but to refrain from pressing the spacebar when they see a blue rectangle (=no-go). The blue and green rectangles can be vertical or horizontal. The vertical rectangle has a high probability of being green (a “go” trial) and the horizontal rectangle has a high probability of being blue (a “no-go” trial). Participants receive information about the orientation of the rectangle (=cue) shortly before the color of the rectangle is revealed. Activation and inhibitory tendencies develop cue-dependence consistently [70]. The task included 250 cues and took approximately 10 minutes to complete. We recorded inhibition commission and omission errors jointly, and reaction time for responses to the targets across all trials.

DD is a measure of the ability to delay immediate smaller virtual rewards for delayed larger rewards. DD is a transdiagnostic process in psychiatric disorders [71,72]. We used the laboratory-based Inquisit software with an implemented procedure based on Richards et al [73]. Participants were asked to choose between either a standard amount of hypothetical money (US $10) with different time or probability delays or a variable amount with no delay until an indifference point for each delay is found, or until 30 trials have been run for each delay. This script establishes (1) the hypothetical payoffs at which participants start to discount higher monetary rewards in favor of shorter wait periods (delay discounting), and (2) the fictional payoffs at which participants start to discount high monetary rewards of unsure events in favor of lower monetary rewards of sure wins (probability discounting). The task takes approximately 15 minutes to complete. We recorded the cumulative probability of choosing the smaller reward across all 5 trials for each task administration [74,75].

Mobile Assessments (DMT App)

Schedule

We conducted a mobile study with a 21-day assessment period using the DMT app. The schedule of mobile assessments using the DMT app is summarized in Table 1.

Table 1. Digital Marshmallow Test (DMT) app assessment schedule during the 21-day study.

Assessment item	Baseline	Morning	Evening	Day 21
Self-report: feel in general (semantic differential items 1-20)	X			X
Ecological momentary assessment				
Feel right now (semantic differential items 4, 5, 7, 8, 11, 12, 15, 17, 19, 20)		X	X	
Photographic Affect Meter	X	X	X	
Active Task				
mBART[a]	X	X[b]	X[b]	X
mGNG[c]	X	X[b]	X[b]	X
mDD[d]	X	X[b]	X[b]	X

[a]mBART: mobile Balloon Analogue Risk Task.

[b]Randomly display one out of mBART, mGNG, mDD.

[c]mGNG: mobile Go-No-Go.

[d]mDD: mobile Delay Discounting.

DMT Self-Report and EMA

The DMT app primary measures of impulsivity were assessed via self-report semantic differentials [46,50]. We selected 20 items from the semantic differential scale and used a selection of items at different time points during the study as described in Table 1. All of the items were measured both at baseline and at the end of the study. Participants were asked to either report semantic differentials based on trait (feeling in general) or current state (feel right now).

The DMT app also prompted the PAM (Figure 3) to measure emotional state and mood. The PAM is designed for momentary response where users choose an image that best represents their emotion at a given time [51]. The PAM was prompted at baseline, every morning and evening, and at the end of study (Table 1).

DMT Active Performance Tasks

The DMT app version of the BART (mBART) was similar to the laboratory version except that it was shorter (15 trials; about 2 minutes long). Participants were instructed to earn as much money as possible during the 15 trials (Figure 4). For participants, the task was named the “Balloon Game” (Figure 2).

The DMT app version of the cued GNG (mGNG) included 75 trials, each of which had the following sequence: fixation cross (250 milliseconds); blank screen (250 milliseconds), vertical or horizontal cue (white rectangle) for 1 of 6 stimulus onset asynchronies (100, 200, 300, 400, 500, 750 milliseconds); go or no-go target (green or blue rectangle, respectively) until the participant responds or for 500 milliseconds; and an intertrial interval (250 milliseconds). Participants were instructed to respond by pressing the screen as fast as possible to green, but not to blue, targets. Cues signal a target at 70% probability (horizontal: go, vertical: no-go). For participants, the task was named the “Square Test” (Figure 2).

In the DMT app version of the DD task (mDD), participants were given 5 choices between a smaller hypothetical monetary or time-based reward that varied from trial to trial based on the previous response and a larger fixed reward that remained the same throughout all trials. For participants, the task was named “Now or Later?” (Figure 2).

Results

Approach and Descriptive Statistics

In this study, we validated self-report, EMA, and behavioral measures of impulsive behavior on mobile devices. Overall, we validated our mobile assessments against previously validated clinical measures of impulsivity such as the BIS-15 and UPPS. We also examined the psychometric properties of our novel measures. Descriptive statistics and analyses are available in the DMT project folder at the OSF [55].

Semantic Differentials

We performed a principal component analysis with varimax rotation of the 20-item semantic differential scale that was measured at baseline (Table 1). Our exploratory analysis yielded a solution with 6 factors of traits we called inefficient, negative, calm, unhealthy, thrill-seeking, and intentional. Combined, these components explained 74% of the variance in the scale. Full results of the principal components analysis are shown in Table 2.

Table 2. Factor loadings of principal component analysis of the semantic differential scale at baseline.

Semantic differentials	Inefficient	Negative	Calm	Unhealthy	Thrill	Intentional
1. Efficient-Inefficient	*0.849*[a]	0.146	–0.055	0.129	–0.054	–0.009
2. Organized-Unorganized	*0.821*	–0.026	–0.044	–0.073	0.172	–0.080
3. Productive-Unproductive	*0.765*	0.289	0.017	0.318	–0.031	0.062
4. Focused-Distracted	*0.721*	0.209	–0.337	0.135	–0.052	–0.229
5. Determined-Aimless	*0.580*	0.534	–0.063	0.176	–0.091	0.069
6. Clear headed-Confused	*0.580*	0.210	–0.348	0.128	–0.149	–0.375
7. Bored-Engaged	*–0.539*	–0.184	0.392	–0.229	0.015	0.272
8. Optimistic-Pessimistic	0.133	*0.867*	–0.122	0.016	–0.072	0.075
9. Positive-Negative	0.249	*0.837*	–0.156	–0.056	0.040	0.090
10. Sad-Happy	–0.052	*–0.682*	0.527	–0.245	0.063	0.087
11. Lonely-Connected	–0.015	*–0.654*	0.311	–0.280	–0.145	0.135
12. Proud-Ashamed	0.460	*0.646*	–0.063	0.113	–0.001	–0.035
13. Calm-Anxious	0.227	0.094	*–0.836*	–0.106	–0.037	0.013
14. Stressed-Relaxed	–0.071	–0.239	*0.835*	–0.216	–0.141	–0.071
15. Frustrated-Content	–0.073	–0.444	*0.597*	–0.125	–0.132	0.138
16. Healthy-Unhealthy	0.139	0.118	–0.028	*0.882*	0.002	–0.047
17. Energetic-Tired	0.373	0.122	–0.235	*0.708*	–0.139	0.163
18. Conservative-Progressive	0.024	–0.006	–0.166	–0.147	*0.795*	0.239
19. Cautious-Thrill seeking	–0.039	0.029	0.045	0.065	*0.784*	–0.341
20. Impulsive-Intentional	–0.178	0.070	0.011	0.077	–0.059	*0.908*
Explained variance (%) (α)	20 (.88)	18 (.87)	13 (.81)	9 (.71)	7 (.46)	7 (N/A[b])

[a]Values in italics indicate factors corresponding to the component.

[b]N/A: not applicable.

Correlations with BIS-15 and UPPS

We examined correlations between validated measures (BIS-15 and UPPS) and our 6 factors. The inefficient and thrill-seeking factors were highly correlated with various trait measures of impulsivity and impulsive behavior. In contrast, the negative, calm, and unhealthy factors showed only minimal or nonsignificant correlations with trait measures of impulsivity and impulsive behavior. The *impulsive-intentional* factor, which consists of only one item, was significantly correlated with 7 out of 9 trait-based measures. Full results of correlations are shown in Table 3.

Table 3. Correlations between semantic differential factors and the Barratt Impulsiveness Scale (BIS-15)/Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency Impulsive Behavior Scale (UPPS).

Variable	Inefficient	Negative	Calm	Unhealthy	Thrill	Intentional
Motor						
r	0.320	0.008	–0.216	–0.068	0.106	–0.534
P value	.001	.94	.03	.50	.29	<.001
Nonplanning						
r	0.352	–0.012	0.097	0.059	–0.098	–0.234
P value	<.001	.90	.33	.55	.33	.02
Attention						
r	0.423	0.209	–0.307	0.067	–0.055	–0.150
P value	<.001	.03	.002	.50	.58	.13
BIS-15						
r	0.516	0.097	–0.188	0.033	–0.031	–0.416
P value	<.001	.33	.06	.74	.76	<.001
Urgency						
r	0.376	0.115	–0.146	0.105	–0.023	–0.419
P value	<.001	.25	.14	.29	.82	<.001
Premeditation						
r	0.278	–0.035	–0.050	0.058	0.324	–0.304
P value	.004	.72	.61	.56	.001	.002
Perseverance						
r	0.575	0.150	–0.206	0.094	0.019	–0.062
P value	<.001	.13	.04	.34	.85	.53
Sensation seeking						
r	–0.082	–0.136	0.068	0.026	0.542	–0.361
P value	.41	.17	.49	.79	<.001	<.001
UPPS						
r	0.346	0.008	–0.094	0.101	0.359	–0.477
P value	<.001	.94	.34	.31	<.001	<.001

Intraindividual Variability

We further examined the intraindividual variability in self-reported semantic differentials between baseline and morning and evening measures. We compared how individuals' "feeling in general" self-reports correlated with average daily reports of the same semantic differentials over 21 days.

As shown in Table 4, correlations between baseline and morning measures ranged from moderate (r=0.4) to high (r=0.7) with *lonely-connected*, *optimistic-pessimistic*, and *determined-aimless* having the highest correlations, and *focused-distracted*, *energetic-tired*, and *bored-engaged* having the lowest correlations. Correlations between baseline and evening measures ranged from low (r=0.2) to high (r=0.7) with *lonely-connected*, *cautious-thrill seeking*, and *impulsive-intentional* having the highest correlations, and *focused-distracted*, *energetic-tired*, and *bored-engaged* having the lowest correlations. Overall, the deviations and variability from baseline were similar across morning and evening momentary measures, with evening demonstrating lower calibration with baseline measures.

Correlations between morning and evening measures were very high (r=0.8-0.9). We found that individuals were more impulsive, distracted, aimless, tired, pessimistic, and thrill-seeking in the evening compared to the morning. Otherwise, we found that individuals reported similar levels of boredom, loneliness, ashamedness, and frustration in the mornings and evenings.

Table 4. Correlations and paired t test results between baseline, morning, and evening with semantic differentials.

Semantic differentials	Baseline vs morning			Baseline vs evening			Morning vs evening		
	r	*t* (df=97)	*P* value	*r*	*t* (df=98)	*P* value	*r*	*t* (df=97)	*P* value
Focused-Distracted	0.355	–2.313	.02	0.177	–3.864	<.001	0.820	–4.744	<.001
Determined-Aimless	0.548	–6.357	<.001	0.422	–8.308	<.001	0.836	–5.447	<.001
Bored-Engaged	0.391	0.582	.56	0.319	1.298	.20	0.808	1.741	.09
Optimistic-Pessimistic	0.574	–1.089	.28	0.481	–1.950	.05	0.916	–2.752	.007
Lonely-Connected	0.701	0.000	>.99	0.693	–0.111	.91	0.902	–0.702	.48
Proud-Ashamed	0.529	–3.526	.001	0.459	–3.007	.003	0.911	1.150	.25
Frustrated-Content	0.549	0.604	.55	0.514	0.081	.94	0.875	–1.201	.23
Energetic-Tired	0.492	–2.747	.007	0.273	–5.584	<.001	0.590	–5.254	<.001
Cautious-Thrill seeking	0.534	–0.350,	.73	0.539	–1.096	.28	0.911	–2.255	.03
Impulsive-Intentional	0.531	–0.279	.78	0.531	1.424	.16	0.872	4.545	<.001

Effect of Emotional State and Affect

We examined four metrics from the PAM task (valence, arousal, positive, and negative) as they related to momentary semantic differentials in the morning and evening (Table 5). Across the four PAM metrics, positive affect generally correlated higher than others with various semantic differentials. Across the 10 semantic differentials examined, *energetic-tired* and *frustrated-content* showed the highest correlations with PAM metrics. However, both *impulsive-intentional* and *cautious-thrill seeking* semantic differentials did not generally correlate with any of the PAM metrics (Table 5).

Table 5. Correlations between semantic differential factors and Photographic Affect Meter measures.

Semantic differentials	Valence		Arousal		Positive		Negative	
	Morning	Evening	Morning	Evening	Morning	Evening	Morning	Evening
Focused-Distracted								
r	–0.392	–0.406	–0.448	–0.333	–0.475	–0.461	0.244	0.283
P value	<.001	<.001	<.001	.001	<.001	<.001	.02	.004
Determined-Aimless								
r	–0.365	–0.452	–0.500	–0.371	–0.468	–0.513	0.197	0.314
P value	<.001	<.001	<.001	<.001	<.001	<.001	.05	.001
Bored-Engaged								
r	0.317	0.425	0.551	0.425	0.442	0.508	–0.128	–0.264
P value	.001	<.001	<.001	<.001	<.001	<.001	.21	.008
Optimistic-Pessimistic								
r	0.434	–0.422	–0.413	–0.236	–0.525	–0.442	0.331	0.337
P value	<.001	<.001	<.001	.02	<.001	<.001	.001	.001
Lonely-Connected								
r	0.434	0.445	0.483	0.361	0.522	0.504	–0.275	–0.311
P value	<.001	<.001	<.001	<.001	<.001	<.001	.006	.002
Proud-Ashamed								
r	–0.449	–0.585	–0.453	–0.347	–0.526	–0.620	0.302	0.460
P value	<.001	<.001	<.001	<.001	<.001	<.001	.002	<.001
Frustrated-Content								
r	0.578	0.607	0.390	0.361	0.618	0.644	–0.461	–0.477
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Energetic-Tired								
r	–0.479	–0.461	–0.706	–0.722	–0.630	–0.639	0.239	0.182
P value	<.001	<.001	<.001	<.001	<.001	<.001	.02	.07
Cautious-Thrill seeking								
r	–0.012	–0.065	0.079	0.104	0.014	–0.021	0.043	0.108
P value	.90	.52	.44	.31	.89	.84	.68	.29
Impulsive-Intentional								
r	0.190	0.322	0.008	0.078	0.167	0.303	–0.197	–0.299
P value	.06	.001	.94	.44	.10	.002	.05	.003

Active Performance Tasks

mBART

To validate the mBART, we assessed the correlation between behavior in the validated laboratory measure of BART and the exploratory mBART active task. The number of explosions in the lab (N=114; mean 6.55, SD 5.25) was highly correlated (r=0.658, P<.001) with the number of explosions in the mBART (N=102; mean 5.62, SD 2.65) at baseline. We also estimated the test-retest reliability of the number of explosions in the mBART and found high correlations between baseline and morning (r=0.663, P<.001), evening (r=.0673, P<.001), and day 21 (r=0.451, P<.001). Results for the number of pumps were almost identical to the results for the number of explosions, as these measures are highly correlated (r=0.643, P<.001). The number of explosions on mBART moderately correlated with the sensation-seeking trait (r=0.216, P=.03). Both the number of explosions (r=0.30, P=.002) and the number of pumps (r=0.268, P=.006) on the mBART correlated with the thrill-seeking factor from semantic differentials.

mGNG

To validate the mGNG, we tested the correlation between behavior in the validated laboratory measure of GNG and the exploratory mGNG active task. Response time (in milliseconds) in the lab (N=109; mean 353, SD 43) was highly correlated (r=0.467, P<.001) with response time in the mGNG (N=97;

mean 430, SD 80) at baseline. We also estimated the test-retest reliability of response time in the mGNG and found high correlations between baseline and morning (r=0.88, P<.001), evening (r=0.862, P<.001), and day 21 (r=0.789, P<.001). Error rates between the lab and mobile version were not correlated due to the low overall error rate in the lab task (mean 0.00765, SD 0.014569) and the high overall rate of error in the mGNG (mean 0.39, SD 0.74). Notably, average error rates on the mGNG at baseline did not correlate with morning, evening, and day-21 error rates. The test-retest reliability changed during the study since morning correlated with evening (r=0.477, P<.001) and day 21 (r=0.454, P<.001), which also correlated with evening (r=0.461, P<.001). This is consistent with the participants' reported frustration with mGNG during the study, which might have led to poorer performance. The response rate on the mGNG task negatively correlated with the sensation-seeking trait (r=–0.310, P=.002). The error rate on mGNG marginally negatively correlated with the organization factors from semantic differentials (r=–0.194, P=.06) and response time marginally negatively correlated with the cautious factor from semantic differentials (r=–0.196, P=.05).

mDD

We had trouble validating the mDD active task with the equivalent lab version, as we used a shortened exploratory version of the DD [74]. However, our results yielded moderate test-retest reliability and convergent validity. We examined the propensity of choosing the later reward with respect to both money and time. The propensity to choose the later reward (money) in 6 months correlated highly with the propensity of choosing the later reward (money) at 1 month (r=0.489, P=.002) and the later reward (time) in 1 year (r=0.396, P<.001). The propensity to choose the later reward (time) in 1 year highly correlated with the propensity to choose the later time reward in 6 months (r=0.523, P=.001). We also estimated the test-retest reliability of the propensity to choose later in the mDD and found high correlations. Propensity to choose the later reward (money) in 6 months at baseline correlated highly with the propensity of choosing the later reward (money) in 6 months at day 21 (r=0.414, P<.001). The propensity to choose the later reward (time) in 12 months at baseline correlated highly with the propensity of choosing the later reward (time) in 12 months at day 21 (r=0.394, P<.001). The propensity to choose the later reward (money) in 6 months at day 21 correlated highly with the propensity of choosing the later reward (time) in 12 months at day 21 (r=0.411, P<.001). There was no association between the mDD and any self-report measure.

Discussion

Principal Results

Overall, the present study demonstrated the potential for assessing different facets of trait and state impulsivity during everyday life using the DMT mobile app. Similar to previous research, the results suggest varying levels of concurrent and predictive validity between existing self-report measures and computer performance tasks, and mobile state and trait versions of these tasks measured over a 21-day period.

Trait and State Self-Report Measures

We built on the semantic differential scale to develop self-report trait and state measures of impulsivity in the DMT app. Our exploratory principal component analysis of the baseline semantic differentials yielded six factors of trait impulsivity: inefficient, negative, calm, unhealthy, thrill-seeking, and intentional. We found that inefficient, intentional, and thrill-seeking factors were highly correlated with various facets of trait impulsivity, whereas the negative, calm, and unhealthy factors only slightly correlated with trait-based measures. Notably, the *impulsive-intentional* factor, which consists of only one item, significantly correlated with 7 out of 9 trait-based measures (BIS-15/UPPS) and can be potentially used as a parsimonious single-item measure of trait impulsivity.

To enhance understanding of state impulsivity and intraindividual variability, we examined the differences between general self-reports and momentary measures of semantic differentials in the morning and evening over the duration of the DMT study. Correlations between baseline and morning measures ranged from moderate (r=0.40) to high (r=0.70), with *lonely-connected*, *optimistic-pessimistic*, and *determined-aimless* showing the highest correlations, and *focused-distracted*, *energetic-tired*, and *bored-engaged* showing the lowest correlations. Correlations between baseline and evening measures ranged from low (r=0.20) to high (r=0.70), with *lonely-connected*, *cautious-thrill seeking*, and *impulsive-intentional* showing the highest correlations, and *focused-distracted*, *energetic-tired*, and *bored-engaged* showing the lowest correlations. Overall, the deviations and variability from baseline were similar across morning and evening momentary measures, with evening responses demonstrating lower calibration with baseline measures.

Our study design also allowed us to investigate these constructs in the context of daily life by comparing morning and evening momentary self-reports. Correlations between morning and evening measures were very high (r=0.80-0.90). We found that individuals were more impulsive, distracted, aimless, tired, pessimistic, and thrill-seeking in the evening compared to the morning. Otherwise, we found that individuals report similar levels of boredom, loneliness, ashamedness, and frustration in mornings and evenings. These results help highlight variations in the facets of impulsivity across the day. Measures that can be attributed to physical and mental depletion [76] had the most variability from morning to evening, whereas those that assess trait-based characteristics were more stable. It is important to recognize that we used a nonclinical sample. Previous studies (eg, Tomko et al [29]) reported that daily impulsivity may vary more in clinical samples than nonclinical samples, suggesting the need for further study in clinical populations.

The results also suggest that some momentary state assessments are highly related over time and day such as *focused-distracted* and *determined-aimless*, whereas others such as *lonely-connected* and *frustrated-content* revealed no significant relationships across all assessment periods. It is also noteworthy that some items, including *impulsive-intentional* and *thrill seeking-cautious*, were only correlated in the morning and evening versus from baseline to morning or evening, suggesting

that although the means may vary from morning to evening, there is a relative intraday association and stability versus over time.

Finally, we examined the role of emotional state, including valence, arousal, positive, and negative, using the PAM [51]. We observed that positive affect generally correlated higher than valance, arousal, and negative affect with various momentary semantic differentials. Specifically, *energetic-tired* and *frustrated-content* showed the highest correlations with PAM metrics. However, both *impulsive-intentional* and *cautious-thrill seeking* semantic differentials did not generally correlate with any valence, arousal, positive, or negative measures of emotional state. These high correlations suggest that using self-report photos instead of or combined with text-based self-report items may enable expanding momentary state assessment to wider audiences regardless of language or education [51]. Further research is warranted to test the PAM in clinical samples across various cohorts.

Active Performance Tasks

One of the primary goals of the DMT study was to validate the behavioral and cognitive active performance tasks in the DMT app. Previous research has highlighted the transdiagnostic potential of behavioral tasks [71,72] but has also identified challenges in test-retest reliability compared to self-reports [77]. In the DMT app, we modeled the design of the DMT active performance tasks (mBART, mGNG, and mDD) based on validated computerized versions of these tasks [63,69,74]. Despite the effort to match the mobile tasks to laboratory tasks, we found only moderate success in validation of these tasks. In this study, mBART demonstrated the highest validity, followed by mGNG and then mDD with the lowest validity.

Specifically, the mBART active task showed high correlations with the lab BART task, high test-retest reliability, and convergent validity with self-report measures. Risk taking in the mBART task correlated with self-reported sensation seeking, which corresponds to prior research with the lab-based BART [78]. Our results also correspond to those of MacLean and colleagues [79] who revealed that a different mobile version of the BART demonstrated good concurrent and predictive validity with the lab computer version. Unlike our results, which were mostly stable across administrations both regarding time of day and over time, there were some differences in BART indices over time in their sample of nondaily smokers. When the studies are combined, it appears that the BART can be translated to a mobile phone to reliably assess risk taking in real-world settings. Nevertheless, the weak correlation between self-report and behavioral measures of risk, as found in other studies [78,80], warrants future investigation of domain-specific or more general measures of risk [80] in the context of impulsive behavior.

GNG is a common behavioral measure of inhibition and cognitive control in clinical trials [66-69]. In these trials, two primary outcomes are usually used to measure cognitive control: error rate and response time. Error rate is particularly important in addiction and substance use studies. On mobile devices, and when performed in a natural setting and outside of the lab (ie, mGNG), the distribution of these metrics is expected to change dramatically. Our results suggest that reaction time is stable across time and contexts with correlations across baseline, morning, evening, and day 21 ranging from 0.79 to 0.86 with no significant mean differences. Error rates had less robust associations across time points, but overall means were relatively stable. This finding may be due to a floor effect in the computerized lab version of the GNG task, which is common in healthy samples and in clinical samples at baseline and without experimental manipulation (eg, alcohol administration) [81-83].

DD is used to measure the ability to delay immediate, smaller, shorter rewards for longer, time-lapsed, but larger rewards. In this study, we did not manage to obtain concurrent validity of the novel mDD and the lab DD task. Nevertheless, individual choices in the mDD during the DMT study showed moderate convergent validity via the correlation between DD money and time versions. Choices in the mDD also showed moderate test-retest reliability from baseline to day 21. The null findings might also be due to the hypothetical, as opposed to incentive-compatible, structure of the mDD task, which decreases validity [84,85], or use of the brief version of the task consisting of only 5 decision points.

When taken together, our results highlight that laboratory mobile assessments can be reliably collected in the field. Although there are some concerns over the relatively weak relationships with self-report impulsivity measures, except for the mBART, similar results have been found with previously validated computer versions of these tasks performed at baseline, suggesting more systemic problems in the objective measurement of impulsivity [23,26]. At the moment, these problems do not appear to be solved through the mobile versions of these tasks. We plan to refine and further validate the mBART, mGNG, and mDD tasks in future studies of the DMT app.

Comparison With Other mHealth Apps and Related Studies

One common clinical use of mHealth apps is remote diagnosis [86]. A systematic review of direct-to-consumer apps identified lack of sufficient clinical evidence for many symptom checkers and diagnostic apps [87]. Our study, which combined validated assessments and novel measures, generated evidence to support the diagnostic capabilities of the DMT app. In the future, the DMT app can be used as a remote patient-facing mHealth app to diagnose and monitor impulsive behaviors.

Our goal was to develop and validate the DMT app for both researchers and clinicians. We used Apple's ResearchKit [38] and Android's ResearchStack [39] open source frameworks for developing health research study apps (Figure 1), which allow researchers to easily develop intuitive and standardized data-collecting mobile apps. The DMT app and measures we developed are cross-platform, open source, and standardized. Our mBART and mGNG tasks, for example, can be easily adapted by other researchers across a variety of psychological, behavioral, and clinical studies that use mobile devices.

Our study also suggests broader design implications for behavioral and cognitive active performance tasks in mHealth studies and apps. In these tasks, users perform activities under

partially controlled conditions while phone sensors are used to collect data. User interfaces and user experience on mobile devices and apps are dramatically different from validated laboratory behavioral and cognitive tasks. Mobile performance tasks are often performed, as intended, in the lived experience of individuals with limited attention and ample distractions. Some tasks such as the mGNG and mDD in this study require more sustained concentration and information processing, while others such as the mBART are more engaging and gamified. The effect of user experience provides a challenge to validation studies, and requires more careful design of behavioral and cognitive active performance tasks in mHealth studies.

Future of Impulsivity Assessments and Interventions

Despite the predictive power of laboratory and self-report measures on trait impulsivity, more research is needed with different samples to disentangle the relationship between impulsivity and health outcomes [2,28,29]. Our study revealed similar results to previous studies that impulsivity is not a unitary construct but is rather composed of qualitatively different constructs, which may or may not have some overlap [22]. Moreover, the complex relationship between impulsive behaviors and health outcomes within each individual might require an n-of-1 approach to prediction and control of impulsive behaviors [37,88,89]. New mHealth methods such as the DMT using multimodal assessment strategies that take trait and state impulsivity into account with contextual variables are needed to further our understanding of how to predict impulsive behavior. Future studies should account for contextual factors such as setting, mood, and intentionality to further disentangle the relationship between trait and state impulsivity, and the different dimensions measured by these tools. Contextual factors can also be used to design more precise behavior change and digital health interventions with mobile technology [88,90-93].

Our ultimate goal is to move from measurement of trait and state impulsivity toward implementation and evaluation of interventions for impulse control and behavior. Despite the overwhelming research on the impact of impulsivity on mental and physical health outcomes, it has been largely ignored as a target of intervention in its own right. Our mobile-based measures can be used to design personalized and adaptive interventions on the same mobile devices and app. Just-in-time adaptive interventions (JITAI) can be designed to provide the right type/amount of support, at the right time, by adapting to an individual's changing internal and contextual state [90,94].

Research performed with daily self-report measures revealed that fluctuations in certain state impulsivity domains (eg, lack of planning, negative urgency) predict heavy drinking, highlighting the opportunity to trigger intervention based on day-to-day fluctuations [95]. Similarly, behavioral active tasks can detect deterioration in inhibitory control during the day [96]. To design JITAI, the combination of the single-item intentionality measurement with results from the mBART could potentially predict a vulnerable state of reduced intentionality and more risk taking on a particular morning compared to other days, or data trends that reveal slow changes in these variables over time. Consequently, with more research, DMT can potentially serve as a just-in-time intervention system for people who are prone to impulsivity and could be made available to people around the world.

Subsequently, understanding the user's state using a game-like component can inform the design of new digital psychological interventions, since the same component could be used both for assessment and intervention seamlessly. For example, upon failure during the mBART, the user can potentially be directed to interact with a new balloon in a way that may help them reassess the number of pumps that may result in an explosion, in the same fashion that new health video games assess and adapt to the user state in an ongoing manner for the enhancement of therapeutic impact [97]. Similarly, research on interventions that manipulate discounting identified learning-based interventions as the most effective [98]. Using DMT to combine both assessment and intervention within one component opens up room for digital microinterventions that focus on very small and beneficial steps that people can take in their daily life [99], which may be far more acceptable than traditional long-term interventions. We plan to introduce and study different personalized and adaptive digital interventions [91,92,99,100] to reduce impulsive response in future studies of the DMT app.

Limitations

The design and implementation of the current DMT study was not sufficient to fully validate the behavioral and cognitive active performance tasks we developed for the DMT app. Similar to other studies that have created mobile versions of impulsivity assessments, this is another step in the right direction despite limitations. In particular, we emphasize our limitation in validating the mDD against equivalent objective laboratory tasks due to challenges with both the lab task we selected and the mobile task we developed. We will continue in our effort to further refine and validate the mobile self-reports and active performance tasks against clinical symptom profiles, diagnoses, contextual factors, and behaviors to generate data on how these constructs are related to mental health and everyday life interactions.

This study was also substantially burdensome for participants due to the sheer number and frequency of required daily assessments. Our results will help to design a lean and personalized version of the DMT app for future studies as we attempt to replicate and further refine our measures. Finally, we plan to validate the DMT app in clinical samples in the context of obesity, addiction, and mental health.

Conclusions

The DMT app can be used to enhance our understanding of impulsivity, impulsive behavior, and failure in self-regulation. Impulsivity measurement is a complex undertaking because of the multidimensionality of the construct as highlighted by the range of measures that assess multiple distinct components. Adding to this problem of construct validity are the various modes of assessment (eg, self-report versus behavioral active performance tasks) and the increased use of daily and momentary assessments. These challenges also present an opportunity to hone our assessment strategies.

Eventually, the goal is to use trait- and state-based self-report and behavioral measures to predict global and momentary

clinical outcomes that can trigger personalized and adaptive digital interventions. These interventions can be targeted and tailored to reduce the various underlying triggers of impulsive responding and enhance self-regulation. Only through rigorous innovation and testing can we begin to build these timely interventions.

Acknowledgments

This study was supported by a Robert Wood Johnson Pioneer Portfolio grant entitled “The Digital Marshmallow Test” (principal investigators: FM and DE).

Authors' Contributions

MS and FM wrote the manuscript. FM, DE, and JP designed the study. JK and HW implemented the mobile app for the study under the supervision of JP and DE. RV recruited and assessed the participants under the supervision of FM and NV. MS, RV, FM, and RL conducted all statistical analyses. All authors reviewed the final manuscript.

Conflicts of Interest

AB has received payment for consulting, from Pro-Change Behavior Systems. All other authors have no conflicts to declare.

References

1. Mischel W, Shoda Y, Rodriguez MI. Delay of gratification in children. Science 1989 May 26;244(4907):933-938. [doi: 10.1126/science.2658056] [Medline: 2658056]
2. Watts TW, Duncan GJ, Quan H. Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes. Psychol Sci 2018 Jul;29(7):1159-1177 [FREE Full text] [doi: 10.1177/0956797618761661] [Medline: 29799765]
3. Casey BJ, Somerville LH, Gotlib IH, Ayduk O, Franklin NT, Askren MK, et al. Behavioral and neural correlates of delay of gratification 40 years later. Proc Natl Acad Sci USA 2011 Oct 06;108(36):14998-15003 [FREE Full text] [doi: 10.1073/pnas.1108561108] [Medline: 21876169]
4. De Ridder RD, De Wit WJ. Concepts, theories, and central issues. In: Self-regulation in health behavior. Chichester: John Wiley & Sons, Ltd; 2006:1-23.
5. Hennessy EA, Johnson BT, Acabchuk RL, McCloskey K, Stewart-James J. Self-regulation mechanisms in health behavior change: a systematic meta-review of meta-analyses, 2006-2017. Health Psychol Rev 2020 Mar;14(1):6-42. [doi: 10.1080/17437199.2019.1679654] [Medline: 31662031]
6. Robson DA, Allen MS, Howard SJ. Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. Psychol Bull 2020 Apr;146(4):324-354. [doi: 10.1037/bul0000227] [Medline: 31904248]
7. Grant JE, Potenza MN, Weinstein A, Gorelick DA. Introduction to behavioral addictions. Am J Drug Alcohol Abuse 2010 Oct;36(5):233-241 [FREE Full text] [doi: 10.3109/00952990.2010.491884] [Medline: 20560821]
8. Weintraub D, Koester J, Potenza MN, Siderowf AD, Stacy M, Voon V, et al. Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients. Arch Neurol 2010 May;67(5):589-595. [doi: 10.1001/archneurol.2010.65] [Medline: 20457959]
9. Duckworth AL, Seligman MEP. Self-discipline outdoes IQ in predicting academic performance of adolescents. Psychol Sci 2005 Dec;16(12):939-944. [doi: 10.1111/j.1467-9280.2005.01641.x] [Medline: 16313657]
10. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: perseverance and passion for long-term goals. J Pers Soc Psychol 2007 Jul;92(6):1087-1101. [doi: 10.1037/0022-3514.92.6.1087] [Medline: 17547490]
11. de Ridder DTD, Lensvelt-Mulders G, Finkenauer C, Stok FM, Baumeister RF. Taking stock of self-control: a meta-analysis of how trait self-control relates to a wide range of behaviors. Pers Soc Psychol Rev 2012 Mar;16(1):76-99. [doi: 10.1177/1088868311418749] [Medline: 21878607]
12. Fishbach A, Converse BA. Identifying and battling temptation. In: Handbook of self-regulation: Research, theory and applications. New York: Guilford Press; 2010:244-260.
13. Whiteside S, Lynam D, Miller J, Reynolds S. Validation of the UPPS impulsive behaviour scale: a four ‐ factor model of impulsivity. Eur J Pers 2020 Dec 02;19(7):559-574. [doi: 10.1002/per.556]
14. Grant J, Potenza M. The Oxford handbook of impulse control disorders. Oxford, England: Oxford University Press; 2011.
15. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Washington, DC: American Psychiatric Pub; 2013.
16. Berg JM, Latzman RD, Bliwise NG, Lilienfeld SO. Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the UPPS for psychopathology. Psychol Assess 2015 Dec;27(4):1129-1146. [doi: 10.1037/pas0000111] [Medline: 25822833]
17. Moeller FG, Barratt ES, Dougherty DM, Schmitz JM, Swann AC. Psychiatric aspects of impulsivity. Am J Psychiatry 2001 Dec;158(11):1783-1793. [doi: 10.1176/appi.ajp.158.11.1783] [Medline: 11691682]

18. Nederkoorn C, Jansen E, Mulkens S, Jansen A. Impulsivity predicts treatment outcome in obese children. Behav Res Ther 2007 May;45(5):1071-1075. [doi: 10.1016/j.brat.2006.05.009] [Medline: 16828053]
19. Hershberger AR, Um M, Cyders MA. The relationship between the UPPS-P impulsive personality traits and substance use psychotherapy outcomes: A meta-analysis. Drug Alcohol Depend 2017 Sep 01;178:408-416 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.032] [Medline: 28709080]
20. Churchill S, Jessop DC. Too impulsive for implementation intentions? Evidence that impulsivity moderates the effectiveness of an implementation intention intervention. Psychol Health 2011 May;26(5):517-530. [doi: 10.1080/08870441003611536] [Medline: 20945252]
21. Sharma L, Markon KE, Clark LA. Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. Psychol Bull 2014 Mar;140(2):374-408. [doi: 10.1037/a0034418] [Medline: 24099400]
22. Dalley JW, Robbins TW. Fractionating impulsivity: neuropsychiatric implications. Nat Rev Neurosci 2017 Feb 17;18(3):158-171. [doi: 10.1038/nrn.2017.8] [Medline: 28209979]
23. Stahl C, Voss A, Schmitz F, Nuszbaum M, Tüscher O, Lieb K, et al. Behavioral components of impulsivity. J Exp Psychol Gen 2014 May;143(2):850-886. [doi: 10.1037/a0033981] [Medline: 23957282]
24. King KM, Patock-Peckham JA, Dager AD, Thimm K, Gates JR. On the Mismeasurement of Impulsivity: Trait, Behavioral, and Neural Models in Alcohol Research among Adolescents and Young Adults. Curr Addict Rep 2014 Jan 4;1(1):19-32. [doi: 10.1007/s40429-013-0005-4]
25. Emery RL, Levine MD. Questionnaire and behavioral task measures of impulsivity are differentially associated with body mass index: A comprehensive meta-analysis. Psychol Bull 2017 Aug;143(8):868-902. [doi: 10.1037/bul0000105] [Medline: 28493725]
26. Strickland JC, Johnson MW. Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. Psychol Rev 2020 Oct 24. [doi: 10.1037/rev0000263] [Medline: 32969672]
27. Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. Perspect Psychol Sci 2007 Dec;2(4):313-345 [FREE Full text] [doi: 10.1111/j.1745-6916.2007.00047.x] [Medline: 26151971]
28. Carlson SM, Shoda Y, Ayduk O, Aber L, Schaefer C, Sethi A, et al. Cohort effects in children's delay of gratification. Dev Psychol 2018 Aug;54(8):1395-1407. [doi: 10.1037/dev0000533] [Medline: 29939038]
29. Tomko RL, Solhan MB, Carpenter RW, Brown WC, Jahng S, Wood PK, et al. Measuring impulsivity in daily life: the momentary impulsivity scale. Psychol Assess 2014 Jul;26(2):339-349 [FREE Full text] [doi: 10.1037/a0035083] [Medline: 24274047]
30. Halvorson MA, Pedersen SL, Feil MC, Lengua LJ, Molina BSG, King KM. Impulsive States and Impulsive Traits: A Study of the Multilevel Structure and Validity of a Multifaceted Measure of Impulsive States. Assessment 2020 Jul 17:1073191120939161. [doi: 10.1177/1073191120939161] [Medline: 32680433]
31. Sperry SH, Lynam DR, Walsh MA, Horton LE, Kwapil TR. Examining the multidimensional structure of impulsivity in daily life. Pers Individ Differ 2016 May;94:153-158. [doi: 10.1016/j.paid.2016.01.018]
32. Sperry SH, Lynam DR, Kwapil TR. The convergence and divergence of impulsivity facets in daily life. J Pers 2018 Oct;86(5):841-852. [doi: 10.1111/jopy.12359] [Medline: 29125631]
33. Griffin SA, Trull TJ. Alcohol use in daily life: Examining the role of trait and state impulsivity facets. Psychol Addict Behav 2020 Oct 10:Advance online. [doi: 10.1037/adb0000679] [Medline: 32914990]
34. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
35. Kumar S, Nilsen W, Pavel M, Srivastava M. Mobile Health: Revolutionizing Healthcare Through Transdisciplinary Research. Computer 2013 Jan;46(1):28-35. [doi: 10.1109/MC.2012.392]
36. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
37. Estrin D. Small data, where n = me. Commun ACM 2014 Apr;57(4):32-34. [doi: 10.1145/2580944]
38. ResearchKit. Apple Developer. URL: https://developer.apple.com/researchkit/ [accessed 2020-11-04]
39. ResearchStack. URL: http://researchstack.org/ [accessed 2020-11-04]
40. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011. [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
41. Jardine J, Fisher J, Carrick B. Apple's ResearchKit: smart data collection for the smartphone era? J R Soc Med 2015 Aug;108(8):294-296 [FREE Full text] [doi: 10.1177/0141076815600673] [Medline: 26268915]
42. Sequeira L, Perrotta S, LaGrassa J, Merikangas K, Kreindler D, Kundur D, et al. Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: A scoping review. J Affect Disord 2020 Mar 15;265:314-324. [doi: 10.1016/j.jad.2019.11.156] [Medline: 32090755]
43. Wang R, Aung M, Abdullah S, Brian R, Campbell A, Choudhury T, et al. CrossCheck: Toward Passive Sensing Detection of Mental Health Changes in People with Schizophrenia. USA: ACM; 2016 Presented at: Proceedings of the 2016 ACM

International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12, 2016; Heidelberg, Germany p. 886-897. [doi: 10.1145/2971648.2971740]
44. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York City: Guilford Press; 2013.
45. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. Annu Rev Psychol 2003;54:579-616. [doi: 10.1146/annurev.psych.54.101601.145030] [Medline: 12499517]
46. Lorr M, Wunderlich RA. A semantic differential mood scale. J Clin Psychol 1988 Jan;44(1):33-36. [doi: 10.1002/1097-4679(198801)44:1<33::aid-jclp2270440106>3.0.co;2-n] [Medline: 3343359]
47. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995 Dec;51(6):768-774. [doi: 10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1] [Medline: 8778124]
48. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
49. Fleeson W. Situation-based contingencies underlying trait-content manifestation in behavior. J Pers 2007 Aug;75(4):825-861. [doi: 10.1111/j.1467-6494.2007.00458.x] [Medline: 17576360]
50. Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. J Behav Ther Exp Psych 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9]
51. Pollak JP, Adams P, Gay G. PAM: A Photographic Affect Meter for Frequent, in Situ Measurement of Affect. USA: ACM; 2011 Presented at: Proceedings of the SIGCHI conference on Human factors in computing systems; May 7, 2011; Vancouver, BC p. 725-734. [doi: 10.1145/1978942.1979047]
52. Thompson ER. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). J Cross Cult Psychol 2016 Jul 26;38(2):227-242. [doi: 10.1177/0022022106297301]
53. Active Tasks. ResearchKit. URL: http://researchkit.org/docs/docs/ActiveTasks/ActiveTasks.html [accessed 2020-11-04]
54. Scarpina F, Tagini S. The Stroop Color and Word Test. Front Psychol 2017;8:557. [doi: 10.3389/fpsyg.2017.00557] [Medline: 28446889]
55. Digital Marshmallow Test (DMT). OSF. URL: https://osf.io/mnxfh/ [accessed 2020-12-28]
56. Ramanathan N, Alquaddoomi F, Falaki H, George D, Hsieh C, Jenkins J, et al. ohmage: An open mobile system for activity and experience sampling. 2012 Presented at: 6th International Conference on Pervasive Computer Technology in Healthcare; Pervasive Health Workshop; 2012; San Diego, California. [doi: 10.4108/icst.pervasivehealth.2012.248705]
57. Tangmunarunkit H, Hsieh CK, Longstaff B, Nolen S, Jenkins J, Ketcham C, et al. Ohmage: a general and extensible end-to-end participatory sensing platform. ACM Trans Intell Syst Technol 2015 May 20;6(3):1-21. [doi: 10.1145/2717318]
58. Gregersen PK, Klein G, Keogh M, Kern M, DeFranco M, Simpfendorfer KR, et al. The Genotype and Phenotype (GaP) registry: a living biobank for the analysis of quantitative traits. Immunol Res 2015 Dec;63(1-3):107-112. [doi: 10.1007/s12026-015-8711-8] [Medline: 26467974]
59. Feinstein Institutes for Medical Research. URL: https://feinstein.northwell.edu/ [accessed 2020-11-15]
60. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
61. Stanford MS, Mathias CW, Dougherty DM, Lake SL, Anderson NE, Patton JH. Fifty years of the Barratt Impulsiveness Scale: An update and review. Person Ind Differ 2009 Oct;47(5):385-395. [doi: 10.1016/j.paid.2009.04.008]
62. Spinella M. Normative data and a short form of the Barratt Impulsiveness Scale. Int J Neurosci 2007 Mar;117(3):359-368. [doi: 10.1080/00207450600588881] [Medline: 17365120]
63. Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, et al. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). J Exp Psychol Appl 2002 Jul;8(2):75-84. [doi: 10.1037//1076-898x.8.2.75] [Medline: 12075692]
64. Lejuez CW, Aklin WM, Zvolensky MJ, Pedulla CM. Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. J Adolesc 2003 Aug;26(4):475-479. [doi: 10.1016/s0140-1971(03)00036-8] [Medline: 12887935]
65. Inquisit. Millisecond. URL: https://www.millisecond.com/ [accessed 2020-11-04]
66. Weafer J, Fillmore MT, Milich R. Increased sensitivity to the disinhibiting effects of alcohol in adults with ADHD. Exp Clin Psychopharmacol 2009 May;17(2):113-121 [FREE Full text] [doi: 10.1037/a0015418] [Medline: 19331488]
67. Kaufman JN, Ross TJ, Stein EA, Garavan H. Cingulate hypoactivity in cocaine users during a GO-NOGO task as revealed by event-related functional magnetic resonance imaging. J Neurosci 2003 Aug 27;23(21):7839-7843 [FREE Full text] [Medline: 12944513]
68. Hester R, Garavan H. Executive dysfunction in cocaine addiction: evidence for discordant frontal, cingulate, and cerebellar activity. J Neurosci 2004 Dec 08;24(49):11017-11022 [FREE Full text] [doi: 10.1523/JNEUROSCI.3321-04.2004] [Medline: 15590917]
69. Fillmore MT, Rush CR, Hays L. Acute effects of cocaine in two models of inhibitory control: implications of non-linear dose effects. Addiction 2006 Oct;101(9):1323-1332. [doi: 10.1111/j.1360-0443.2006.01522.x] [Medline: 16911732]

70. Posner MI. Orienting of attention. Q J Exp Psychol 1980 Mar 01;32(1):3-25. [doi: 10.1080/00335558008248231] [Medline: 7367577]
71. Amlung M, Marsden E, Holshausen K, Morris V, Patel H, Vedelago L, et al. Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. JAMA Psychiatry 2019 Dec 01;76(11):1176-1186 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.2102] [Medline: 31461131]
72. Lempert KM, Steinglass JE, Pinto A, Kable JW, Simpson HB. Can delay discounting deliver on the promise of RDoC? Psychol Med 2019 Jan;49(2):190-199. [doi: 10.1017/S0033291718001770] [Medline: 30070191]
73. Richards JB, Zhang L, Mitchell SH, de Wit H. Delay or probability discounting in a model of impulsive behavior: effect of alcohol. J Exp Anal Behav 1999 Mar;71(2):121-143 [FREE Full text] [doi: 10.1901/jeab.1999.71-121] [Medline: 10220927]
74. Frye CCJ, Galizio A, Friedel JE, DeHart WB, Odum AL. Measuring Delay Discounting in Humans Using an Adjusting Amount Task. J Vis Exp 2016 Jan 09(107):53584 [FREE Full text] [doi: 10.3791/53584] [Medline: 26779747]
75. Myerson J, Baumann AA, Green L. Discounting of delayed rewards: (A)theoretical interpretation of the Kirby questionnaire. Behav Processes 2014 Oct;107:99-105 [FREE Full text] [doi: 10.1016/j.beproc.2014.07.021] [Medline: 25139835]
76. Baumeister RF, Vohs KD. Self-Regulation, Ego Depletion, and Motivation. Social Pers Psych Compass 2007 Nov;1(1):115-128. [doi: 10.1111/j.1751-9004.2007.00001.x]
77. Enkavi A, Eisenberg I, Bissett P, Mazza G, MacKinnon D, Marsch L, et al. Large-scale analysis of test-retest reliabilities of self-regulation measures. Proc Natl Acad Sci USA 2019 Mar 19;116(12):5472-5477 [FREE Full text] [doi: 10.1073/pnas.1818430116] [Medline: 30842284]
78. Lauriola M, Panno A, Levin I, Lejuez C. Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. J Behav Dec Making 2013 Apr 22;27(1):20-36. [doi: 10.1002/bdm.1784]
79. MacLean RR, Pincus AL, Smyth JM, Geier CF, Wilson SJ. Extending the Balloon Analogue Risk Task to Assess Naturalistic Risk Taking via a Mobile Platform. J Psychopathol Behav Assess 2018 Mar;40(1):107-116 [FREE Full text] [doi: 10.1007/s10862-017-9628-4] [Medline: 30505069]
80. Frey R, Pedroni A, Mata R, Rieskamp J, Hertwig R. Risk preference shares the psychometric structure of major psychological traits. Sci Adv 2017 Oct 04;3(10):e1701381. [doi: 10.1126/sciadv.1701381] [Medline: 28983511]
81. Paz AL, Rosselli M, Conniff J. Identifying Inhibitory Subcomponents Associated with Changes in Binge Drinking Behavior: A 6-Month Longitudinal Design. Alcohol Clin Exp Res 2018 Sep;42(9):1815-1822. [doi: 10.1111/acer.13830] [Medline: 29969149]
82. Eigsti I, Zayas V, Mischel W, Shoda Y, Ayduk O, Dadlani MB, et al. Predicting cognitive control from preschool to late adolescence and young adulthood. Psychol Sci 2006 Jul;17(6):478-484. [doi: 10.1111/j.1467-9280.2006.01732.x] [Medline: 16771797]
83. Durston S, Thomas KM, Yang Y, Ulug AM, Zimmerman RD, Casey B. A neural basis for the development of inhibitory control. Develop Sci 2002 Nov;5(4):F9-F16. [doi: 10.1111/1467-7687.00235]
84. Tang J, Chrzanowski-Smith OJ, Hutchinson G, Kee F, Hunter RF. Relationship between monetary delay discounting and obesity: a systematic review and meta-regression. Int J Obes (Lond) 2019 Jun;43(6):1135-1146. [doi: 10.1038/s41366-018-0265-0] [Medline: 30482934]
85. Strickland JC, Lee DC, Vandrey R, Johnson MW. A systematic review and meta-analysis of delay discounting and cannabis use. Exp Clin Psychopharmacol 2020 May 20. [doi: 10.1037/pha0000378] [Medline: 32309968]
86. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
87. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018 Sep 25;5(3):95-105 [FREE Full text] [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
88. Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision Health: The Role of the Social and Behavioral Sciences in Advancing the Vision. Ann Behav Med 2020 Dec 01;54(11):805-826 [FREE Full text] [doi: 10.1093/abm/kaaa018] [Medline: 32338719]
89. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. BMC Med 2019 Jul 17;17(1):133 [FREE Full text] [doi: 10.1186/s12916-019-1366-x] [Medline: 31311528]
90. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
91. Okeke F, Sobolev M, Estrin D. Towards A Framework for Mobile Behavior Change Research. : ACM; 2018 Presented at: Proceedings of the Technology, Mind, and Society; April 5, 2018; Washington, DC p. 1-6. [doi: 10.1145/3183654.3183706]
92. 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]
93. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The Pace of Technologic Change: Implications for Digital Health Behavior Intervention Research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]

94. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
95. Pedersen SL, King KM, Louie KA, Fournier JC, Molina BSG. Momentary fluctuations in impulsivity domains: Associations with a history of childhood ADHD, heavy alcohol use, and alcohol problems. Drug Alcohol Depend 2019 Dec 01;205:107683. [doi: 10.1016/j.drugalcdep.2019.107683] [Medline: 31704385]
96. Jones A, Tiplady B, Houben K, Nederkoorn C, Field M. Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. Psychopharmacology (Berl) 2018 May;235(5):1487-1496 [FREE Full text] [doi: 10.1007/s00213-018-4860-5] [Medline: 29497782]
97. Anguera JA, Gazzaley A. Video games, cognitive exercises, and the enhancement of cognitive abilities. Current Opinion in Behavioral Sciences 2015 Aug;4:160-165. [doi: 10.1016/j.cobeha.2015.06.002]
98. Rung JM, Madden GJ. Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. J Exp Psychol Gen 2018 Oct;147(9):1349-1381 [FREE Full text] [doi: 10.1037/xge0000462] [Medline: 30148386]
99. Baumel A, Fleming T, Schueller SM. Digital Micro Interventions for Behavioral and Mental Health Gains: Core Components and Conceptualization of Digital Micro Intervention Care. J Med Internet Res 2020 Oct 29;22(10):e20631 [FREE Full text] [doi: 10.2196/20631] [Medline: 33118946]
100. Muench F, Vitale R, Potenza M. Digital Technologies in the Assessment and Treatment of Impulsivity and Problematic Alcohol and Drug Use. In: The Oxford Handbook of Digital Technologies and Mental Health. Oxford, England: Oxford University Press; 2020.

Abbreviations

ADHD: attention deficit/hyperactivity disorder
BART: Balloon Analogue Risk Task
BIS: Barratt Impulsiveness Scale
DD: Delay Discounting
DMT: Digital Marshmallow Test
EMA: ecological momentary assessment
GNG: Go-No-Go
JITAI: just-in-time adaptive interventions
JSON: JavaScript Object Notation
mBART: mobile Balloon Analogue Risk Taker
mDD: mobile Delay Discounting
mGNG: mobile Go-No-Go
mHealth: mobile health
PAM: Photographic Affect Meter
UPPS: Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency Impulsive Behavior Scale

Edited by G Eysenbach; submitted 15.10.20; peer-reviewed by E Airola, P Chow; comments to author 27.10.20; revised version received 29.11.20; accepted 07.12.20; published 22.01.21.

Please cite as:
Sobolev M, Vitale R, Wen H, Kizer J, Leeman R, Pollak JP, Baumel A, Vadhan NP, Estrin D, Muench F
The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study
JMIR Mhealth Uhealth 2021;9(1):e25018
URL: http://mhealth.jmir.org/2021/1/e25018/
doi:10.2196/25018
PMID:33480854

Original Paper

Exploring How Older Adults Use a Smart Speaker–Based Voice Assistant in Their First Interactions: Qualitative Study

Sunyoung Kim[1], PhD

School of Communication and Information, Rutgers University, New Brunswick, NJ, United States

Corresponding Author:
Sunyoung Kim, PhD
School of Communication and Information
Rutgers University
4 Huntington St
New Brunswick, NJ, 08901-1071
United States
Phone: 1 848 932 7585
Email: sunyoung.kim@rutgers.edu

Abstract

Background: Smart speaker–based voice assistants promise support for the aging population, with the advantages of hands-free and eyes-free interaction modalities to handle requests. However, little is known about how older adults perceive the benefits of this type of device.

Objective: This study investigates how older adults experience and respond to a voice assistant when they first interact with it. Because first impressions act as strong predictors of the overall attitude and acceptability of new technologies, it is important to understand the user experiences of first exposure.

Methods: We conducted semistructured interviews with 18 people 74 years and older who had never used a smart speaker before, investigating the patterns of use, usability issues, and perspectives that older adults have when using a voice assistant for the first time.

Results: The overall first response to a voice assistant was positive, thanks to the simplicity of a speech-based interaction. In particular, a positive and polite response to complete the interaction with a voice assistant was prevalent, such as expressing gratitude or giving feedback about the quality of answers. Two predominant topics of commands made in the first interaction include asking health care–related questions and streaming music. However, most of the follow-up reactions were unfavorable because of the difficulty in constructing a structured sentence for a command; misperceptions about how a voice assistant operates; and concerns about privacy, security, and financial burdens. Overall, a speech-based interaction was perceived to be beneficial owing to its efficiency and convenience, but no other benefits were perceived.

Conclusions: On the basis of the findings, we discuss design implications that can positively influence older adults' first experiences with a voice assistant, including helping better understand how a voice assistant works, incorporating mistakes and common interaction patterns into its design, and providing features tailored to the needs of older adults.

(JMIR Mhealth Uhealth 2021;9(1):e20427) doi:10.2196/20427

KEYWORDS

older adults; voice assistant; smart speaker; technology acceptance; quality of life

Introduction

Background

Recent advances in speech technology and artificial intelligence have made speech a promising form of input modality to interact with personal computing technology. Consequently, a smart speaker with an integrated voice assistant is increasingly available in the market to function as a virtual assistant to perform everyday tasks (eg, Amazon Echo, Google Home). The first commercial version of a smart speaker, Amazon Echo, with an integrated voice assistant, Alexa, was launched in 2015. As of 2019, more than 60 million people in the United States own smart speaker devices, with a 48% annual growth [1]. A voice assistant allows users to perform a range of basic everyday tasks through voice commands, including but not limited to searching for information, streaming music, getting weather and news

updates, ordering groceries, and sending and receiving text messages.

As the worldwide population is aging and countries are facing ongoing challenges in caring for their aging population, there has been increased awareness and interest in the potential of smart speaker–based voice assistants to support older adults for their health and independence [2]. Although a voice assistant holds great promise to support older adults through its simple speech-based interaction modality, little is known about how older adults perceive and respond to the idea of talking to a device that does not have a graphical user interface. Researchers have recently begun to investigate older adults' use of a voice assistant, but with few exceptions [3], they produce preliminary or interim reports [2,4,5].

As a first step toward gaining insights into the perspective and use of voice assistants among older adults, this paper investigates novice older users' first impressions of voice assistants. Since first impressions act as strong predictors of overall attitudes toward new technologies [6], it is important to understand how older adults perceive and respond to a voice assistant when first exposed to it. Through interviews with 18 people aged 74 years and above who had never used a smart speaker, we investigated how older adults interact with a voice assistant in their first use. Specifically, we aimed to answer the following research questions:

1. What usage patterns do older adults have in their first interactions with a voice assistant?
2. What challenges do older adults face when using a voice assistant for the first time?
3. What first impressions do older adults have with a voice assistant?

By answering these questions, we aim to provide insights into the design of a voice assistant that can form a positive first impression among older adults to improve its acceptance and adoption as well as meeting the needs for quality of later life.

The notion of a voice-controlled system has been framed in many different ways, such as a virtual assistant, embodied conversational agent, intelligent personal agent, autonomous agent, or avatar [7]. Throughout this paper, we use the term *voice assistant* to encompass the above terminology and to specifically refer to voice assistants on smart speakers.

Older Adults and Technology Acceptance

Although new technologies are increasingly being introduced to revolutionize aging in place, the actual acceptance of these technologies is still low [8]. Thus, much effort has been devoted to identifying factors that influence older adults' use of technology [4,9,10]. The most commonly identified barriers come in the form of aging-related declines [11]. For instance, the small size and low contrast of buttons on a mobile display have a significant negative influence on interaction performance [5]. Another common barrier is related to psychological and mental obstacles, such as negative cognitive perception [3], technophobia [12], and lack of self-efficacy [4]. Older adults especially tend to refuse a new technology due to perceived effort associated with learning [13] or lack of perceived usefulness [14]. As such, several physical and psychological factors were found to have a significant influence on older adults' technology acceptance.

Research on Voice Assistants

In recent years, there has been a growing interest within the human-computer interaction community in understanding people's experiences with voice assistants. Among several, one stream of research has focused on exploring the use of voice assistants in various contexts, such as different locations [15,16] or tasks [17,18]. Another stream of research has focused on investigating factors that constitute an effective conversation with voice assistants, either through personification [8] or using conversational cues [18]. Some researchers, however, argued that user experiences with voice assistants remain disappointing due to a lack of human-like conversational capabilities [19]. Although existing voice assistants are being called *conversational* agents, promising to enable human-like conversation with a device, they are in fact not truly conversational in nature. Instead, simple and constrained request-response structures are the norm, rarely including a realistic dialog [20]. Lastly, research has explored the utility and usability of voice assistants as an assistive technology [21,22]. In particular, researchers are increasingly recognizing the potential that voice assistants can offer in the aging society [5,23], exploring application areas to facilitate voice assistants to support older adults [24,25]. This paper contributes to this emerging body of literature by specifically investigating the first reactions that older adults have to voice assistants.

Benefits and Challenges of Voice Assistant Use in Older Adult

Voice assistants allow users to interact with it in a universally understood form of interaction modality, speech. Thus, they are deemed to be simple and easy for older adults to use [26]. However, several challenges exist that prevent older adults from interacting with voice assistants [5]. One problem is associated with hearing loss, a common physical complaint in older adults [27]. Because older adults using hearing aids often cannot cope with high levels of ambient noise or have difficulty processing a dialog without contextual information, hearing loss imposes a significant challenge to the use of voice assistants for many of them. Another problem stems from a lack of understanding of the actual expectations and needs of older adults in using voice assistants. Designers are usually considerably younger and may not know about the physical and psychological aspects of aging and have grown up using more advanced technologies than older adults [28]. These problems are important factors to investigate before designing a new system meant to be adaptive and responsive to the perspectives and expectations of older adults. The crucial question is whether a voice assistant is designed to be suitable for older adults. Although researchers have increasingly focused on various usability aspects of voice assistants, relatively less effort has been made to understand them from the perspectives of older adults.

Methods

Participants

For participant recruitment, we first contacted a local assisted-living facility located in the greater New York area. We visited the facility and explained the purpose of the study to the manager. Upon their approval, we posted a recruitment flyer in the lobby. Two recruitment criteria were age over 65 years and had no prior experience with a voice assistant. In total, we recruited 18 participants (n=11, 61% females and n=7, 39% males; mean age 79, range 74-91, SD 4.5 years; Table 1). A total of 2 (11%) participants were wearing a hearing aid but did not have any problems with having a conversation. Other than this, no other specific health concerns were reported. Almost all participants (n=16, 89%) were widows or widowers, living in one-person rooms. The other 2 participants were a couple living in a 2-person room. The average length of residency in the facility was 2.3 years (Min=9 months, max=4 years, SD 1.2). All participants said that they were familiar with personal computers, tablets, and smartphones, and 17% (n=3) participants said that they had seen a smart speaker in their children's homes but had never used it. Seven participants owned a tablet, and all participants reported regularly using a computer for information search and email. We recruited participants from an assisted-living facility for convenience of recruitment. The study protocol was reviewed and approved by the institutional review board.

Table 1. Participants' demographics.

Participant ID	Age (years)	Gender
P1	74	Female
P2	87	Male
P3	76	Female
P4	75	Male
P5	78	Female
P6	83	Male
P7	76	Female
P8	80	Male
P9	75	Female
P10	84	Female
P11	77	Female
P12	78	Female
P13	79	Male
P14	76	Male
P15	79	Female
P16	80	Female
P17	91	Female
P18	82	Male

Data Collection

We constructed the interview protocol to investigate older adults' first experiences of and perspectives on the use of voice assistants to answer our research questions. To that end, we created a set of open-ended interview questions with 3 themes: (1) examining first impressions and perceived utility of a voice assistant, (2) identifying common usage patterns of a voice assistant during their first interactions with it, and (3) finding difficulties and challenges participants experience when using a voice assistant. In addition, we collected participants' basic demographic information, including age, gender, health concerns, and experience with technology.

Participants were randomly assigned to 1 of 3 groups. We conducted an interview with 3 participants in a group so that we could investigate not only participants' experiences with a voice assistant but also their reactions to speech-based interactions initiated and performed by other participants. All participants reported that they knew other participants in a group as they were living in the same facility for years.

To facilitate participants' interaction with voice assistants, we created 7 decks (categories) of cards, on each of which a voice-command query was written to evoke various functionalities. The categories include getting weather and daily news updates, asking general questions, listening to music, reading a book, playing games, texting and communications, and setting up reminders and alarms. Each deck consisted of 4 to 7 cards depending on the topic of the category. For instance, the deck for weather and news update had 5 cards, which queries include:

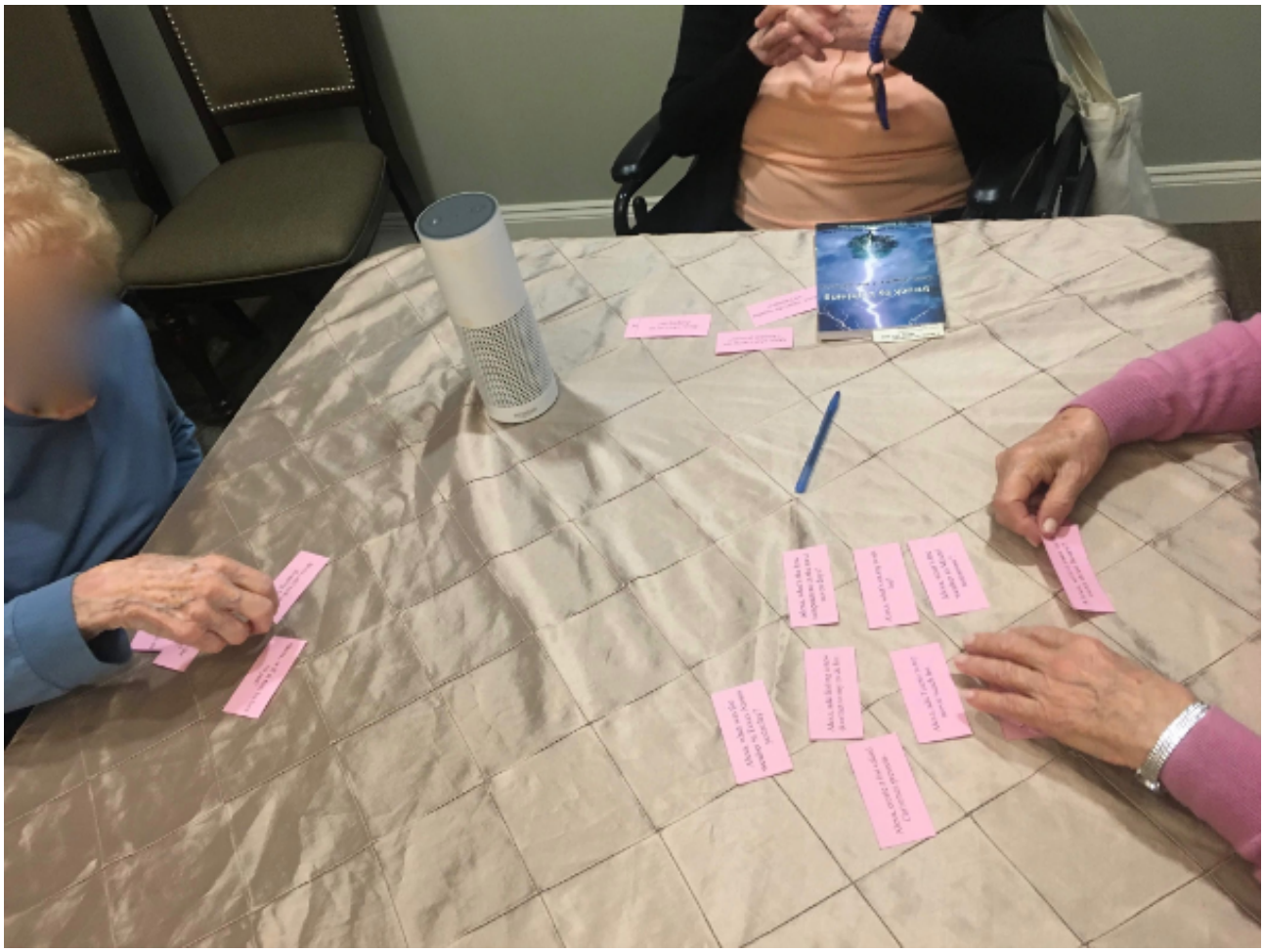
- Alexa, what is the weather like today?
- Alexa, will it rain this weekend?

- Alexa, give me a 7-day weather forecast.
- Alexa, what is a headline in the news today?
- Alexa, give me news updates today?

We also created a simple story for each category that illustrates a real-life situation in which the voice-command queries on cards can be used. For instance, for a deck of cards for weather and news updates, we told a story of leaving for a vacation in Milan, when you might want to check the traffic conditions to the airport, the current weather in Milan, and the weather forecast of the day of return.

The interviews were conducted in a meeting room of the facility where a smart speaker (Amazon Echo) was set up (Figure 1). In the interview, we first introduced a smart speaker to participants as "a device that follows your voice command, providing you answers about news, music, weather, and more." We then demonstrated how to use a voice assistant by asking basic questions about weather and time and by executing simple commands such as streaming music and making a phone call. After this simple introduction, we asked participants to interact freely with the voice assistant as much as they wanted without further training. When participants had no more ideas of what to do with the voice assistant, we then narrated a scenario to inform them of potential use, gave out 1 deck of cards, and asked participants to try out queries written on a card.

Figure 1. The setting of a semistructured interview with a group of participants.



Once all participants had enough interaction with a voice assistant and had no more ideas for interaction, they were asked to freely discuss their perspectives on it and its potential utilities for older adults. After completing the interview, participants were compensated monetarily for their participation. Each interview lasted between one and a half and two hours. All interviews were audio-recorded and transcribed.

Data Analysis

We analyzed the interview data using thematic analysis to reveal patterns across datasets, through open coding, axial coding, and selective coding [29]. The emerged themes were continuously discussed with another author until data were saturated with recurring themes and no new information was anticipated.

First, we conducted open coding to identify and code concepts significant in the data as abstract representations of events, objects, happenings, actions, etc. The example below explains one participant's concern about the use of a voice assistant with respect to cost. This response is coded as *Financial concern*:

> *How much does it cost a month?* [P5, Financial concern]

Next, we categorized the related concepts created by open coding into conceptual phenomena using axial coding. Phenomena refers to repeated patterns of events, happenings, actions, and interactions that represent people's responses to problems and situations. For instance, *Misperception* is a phenomenon that represents a participant's incorrect

understanding or interpretation of the operation of voice assistants. During axial coding, the open code *Financial concern* in the example above was categorized as *Misperception* since using a voice assistant does not require a monthly fee unless you subscribe to charged services and as long as you have a wireless internet at home. Lastly, we followed the selective coding process to assemble our conceptual phenomena extracted from axial coding. The goal of this step is to integrate all concepts by building relationships across phenomena.

Results

Overview

The overall first response to a voice assistant among our participants was positive, showing a favorable attitude toward the simplicity of a speech-based interaction. However, shortly after interacting with it, they started experiencing difficulty with it, asking a lot of questions and raising concerns about its use. In what follows, we report in detail the findings of our participants' first interactions with a voice assistant by answering our research questions. The main categories of the findings are summarized in Textbox 1.

Textbox 1. Research questions and summary of results.

R1. What usage patterns do older adults have in their first interactions with a voice assistant?

- Topics of commands made in the first interaction include:
  - Asking health care-related questions (91/234, 38.9% interactions);
  - Streaming classical music and songs from old days (66/234, 28.2% interactions);
  - Asking for directions to or a location of a place (30/234, 12.8% interactions);
- A positive response to complete the interaction with a voice assistant was prevalent, such as expressing gratitude or giving feedback.

R2. What challenges do older adults face when using a voice assistant for the first time?

- Difficulty in constructing a structured sentence for a command
- Misperceptions about how a voice assistant operates
- Concerns about privacy, security, and financial burdens

R3. What first impression do older adults have with a voice assistant?

- A speech-based interaction was perceived beneficial thanks to its efficiency and convenience
- No other benefits were perceived: “It’s not for me.”

The Usage Patterns of a Voice Assistant

In total, participants made 234 conversational interactions with a voice assistant throughout the interviews. From these interactions, prominent patterns in the initial use of a voice assistant among our participants have emerged in 2 categories: common topics of commands and positive responses to a voice assistant’s answers.

Topics of Commands

Since participants had no prior experience using voice assistants, they mostly relied on the cards we provided for the first few interactions. They then came up with their own commands by making variations or extensions of the queries written on cards. Among the varied topics of commands that participants asked a voice assistant, 3 topics were asked most frequently. The topic of the most frequently executed commands was health care-related (91/234, 38.9% interactions), such as body condition, health supplements, and medications (In the excerpts, P# refers to the #th interviewee, VA refers to a voice assistant, Amazon Echo’s Alexa, and *I* refers to an interviewer). Example queries include the following:

> *Alexa, I got a flu. What kind of medicine should I take?* [P4]
>
> *Alexa, what can I do for arthritis?* [P16]
>
> *Alexa, what does magnesium do for your body?* [P18]

The next most frequent topic of commands was to stream songs from old days and classical music reminiscent of old times (66/234, 28.2% interactions). The third topic of commands was asking for directions to or a location of a place (30/234, 12.8% interactions). Example queries include the following:

> *Alexa, play popular songs from the 70s.* [P10]
>
> *Alexa, where is the Veteran's Cemetery in New Jersey?* [P2]

Other topics of commands include the search for general information, playing audiobooks, and setting up timers and reminders. We expected that participants would ask a voice assistant many questions about the weather forecast and news updates because previous research showed that the most popular smartphone applications among older adults were weather forecasts and news updates [30]. Contrary to our expectations, however, our participants asked a few questions about weather and news updates, except when using queries written on a card, which they expressed little interest in asking. This might be because weather forecasts and news updates are what they would normally check in the morning [30] and thus did not ask in the afternoon when the interviews were held. Alternatively, this

might be because they already have their own ways of checking such information (eg, via television news) and thus are less inclined to use a voice assistant to retrieve it. We will discuss the perceived utility of voice assistants in the following section.

Positive Response to a Voice Assistant

Interestingly, our participants completed most interactions with a voice assistant by responding back to the answer, either expressing gratitude or giving feedback about the quality of a voice assistant's answer.

> *P1: Who is the, Alexa, who is the father of our country?*
>
> *VA: George Washington was one of the founding fathers of the United States of America and served as the nation's first president.*
>
> *P1: Good. Thank you.*
>
> *P9: Alexa, what was the weather in Texas, Austin, yesterday?*
>
> *VA: Currently in Austin, Texas, it is 73 degrees with clear skies and the sun. You can expect more of the same today with a high of 84 degrees and a low of 61 degrees.*
>
> *P9: Why do we not move there? I wish we had that weather.*
>
> *P10: Yeah, I wish that, too.*

Lopatovska and Williams [8] have previously framed this type of interaction as personification with mindless politeness that "humans say to each other without meaning anything." Our observation was different from theirs in that most responses made by our participants had distinct functional purposes. Some responding remarks were used as a cue to finish interacting with a voice assistant as they came back to speak with the interviewer or other participants. Some responses were used as triggers to initiate a follow-up human-human conversation. All these patterns, however, might have resulted from the novelty of using a voice assistant. Because all participants had never used a voice assistant before, they might have expressed their feelings and first impressions through the responding remarks in more extreme attitudes than they would normally do. Alternatively, these patterns might have been shown since the interview was helped as a group with other participants. Thus, further research is needed to investigate how older adults would respond to a voice assistant in an everyday use setting.

Challenges in Using Voice Assistants

Unlike a common belief, or marketing hype, that it is easy to interact with a voice assistant thanks to its conversational capabilities, most participants experienced difficulty having a *conversation* with it. Prior research has indicated several reasons why people found it difficult to interact with a voice assistant, including the goal of a conversation needing to be highly functional and task-oriented, lack of social aspects of conversation, and a sequential dialog structure of request-response [16,31,32]. Whereas, we found that the challenges our participants experienced were more elementary. Two predominant challenges that emerged from our study include the difficulty in constructing a command sentence and misunderstanding how a voice assistant operates.

Difficulty in Constructing a Structured Sentence

To use a voice assistant, a user should first speak a wake word (a word to activate a voice assistant, such as *Alexa* for Amazon Echo or *Hey Google for Google Home*), followed by a concise and definitive sentence for a task. However, many participants kept forgetting to start a command with a wake word or confused the wake word with other similar words throughout the interview.

> *P15: What would you, what, Alexa, what is a good mystery book to read today? (A voice assistant did not activate and P15 paused for a few seconds...)*
>
> *P15: Hmm? What's going on?*
>
> *P6: What time does the Alaska show on the...*
>
> *I: You need to start a sentence with "Alexa"*
>
> *P6: Alaska. No wait. Alexa, what time does the late-night show on the TV tonight start?*
>
> *VA: Tonight, the late-night talk show starts at...*

Even when given a command with a wake word, the voice assistant sometimes did not activate because the command did not have enough pause between a wake word and a following sentence for the voice assistant to capture the wake word or because a wake word was not clearly pronounced. However, using the wrong or no wake word would be an easy fix as users would make fewer mistakes as they continue using a voice assistant. A bigger issue was that many participants struggled to compose a concise command sentence for a task. They often spoke lengthy, unstructured, and descriptive sentences, some of which even included another question within a question. A command with a long sentence resulted in the voice assistant losing track of the voice command and returning to a deactivation mode in the middle of the user speaking. When this happened, participants were not aware that it happened, waited for a response for a few seconds, and became puzzled by why it did not respond to their command. In some other cases, the voice assistant picked up only the first few words of a user's command due to a pause between words and provided a wrong answer.

> *P11: Alexa, let me see, I am trying to think of the book. What hat was the title of the book? Ah, can you read the first page of... (Alexa did not activate, and a user paused for a few seconds)*
>
> *P11: Why is it not answering?*
>
> *P5: Alexa, what is the status of... the status of the new garbage collection?*
>
> *VA: Status is a relative position or standing on things, especially persons in the society or a state at a particular time.*
>
> *P5: It's giving me a wrong answer.*

Misperceptions: Operational Inquiries and Concerns

Since participants had no prior experience with a voice assistant, they asked a lot of questions about its use during the interview. Although the interviewer answered all questions, the topics of

the questions provided us with insights into how older adults might perceive and expect voice assistants to function. In addition, we found that the underlying reasoning of many such questions stemmed from the misperceptions of how a voice assistant would operate, which resulted in forming negative first impressions.

Inquiries About Voice Assistant Operation

The most frequently asked question about voice assistants was how it spontaneously responds to random commands and requests. Most participants did not recognize that a voice assistant retrieves information from the internet. Instead, they supposed that it would retrieve relevant information from the stored local database. This presumption made them expect that they would have to store all the information in a voice assistant before its use. Because of this perception, participants expressed their strong unwillingness to use it or lack of interest in using it due to the perceived effort to store data. In addition, the expectation of having to store all information in a voice assistant aroused privacy and security concerns, which we will discuss in the next section:

> *[After demonstrating music playing and reading of a voice assistant] Who put the music in it? Where does it get music from?... Does this have a text for a whole book? Where does it get that from?* [P13]

> *I think I know how it works or how you make it go. The storage of whatever is in there. That would be a bit of a time-consuming effort to put all of that information on, and whether or not I would want to put certain information on a machine like that. I do not think I would, but you never know.* [P4]

Once participants were informed that a voice assistant does not require storing data but retrieves information from the internet, they started to explore its capability in various capacities. First, participants interacted with the voice assistant by asking simple factual questions (eg, “who is the 7th president of the United States?”). After receiving a satisfactory answer to such questions, they jumped into asking complicated, and probably impossible to answer, questions (eg, “who will be the next president of the United States?”), to which they received a nonsatisfactory answer. After hopping between asking very simple questions and impossible-to-answer questions several times, participants were dissatisfied with the voice assistants’ capabilities and rejected its adoption.

> *P11: Alexa, do you think Trump is gonna be re-elected for President?*
>
> *VA: Sorry, I am not sure how to help with that.*
>
> *P3: Alexa, can you give me the winning lottery numbers?*
>
> *VA: Sorry, I am not sure how to help with that.*
>
> *P3: Of course not.*

Concerns About Privacy, Security, and Financial Burden

The biggest concern participants expressed was the potential risk associated with privacy and security. Because of the privacy-intrusive potential of a voice assistant’s *always-on* ability to continuously listen to voices in intimate spaces such as the home, privacy concerns about using a voice assistant have been subject to much research over recent years [33,34]. We then found that privacy concerns that our participants had were not only about its capability of *always listening* but also related to their misperception that a voice assistant would store a user’s personal information. Because a voice assistant can respond to a verbal command, participants were concerned that other people could easily retrieve their personal information by talking to a voice assistant:

> *I think too much information goes into that (a voice assistant) and it worries me. All the information that you put into that machine goes all over the world. That's my concern.* [P5]

> *What happens if somebody else asks it (a voice assistant) what's the balance in my checking account? Can you set it up to use a password or a security question so that only you can ask questions to it?* [P10]

Another prevalent concern was the cost of using a voice assistant. Apart from the cost of the device itself, several participants expected that they would have to pay a monthly service fee to use a voice assistant. Because a voice assistant responds to a user’s question or inquiry, participants considered it as a service to be paid, which they did not want:

> *All these questions that we asked, you are going to have to pay for, right? How much do you pay? How much does it cost a month?* [P3]

Perceived (No) Usefulness of Voice Assistants

Overall, the speech-based interaction modality was well received by all participants. They appreciated the efficiency and convenience of using speech to receive information. After a few interactions, participants started to talk about the potential benefits of voice assistants to help interact with a device without aging-related physical constraints such as vision or mobility:

> *I used to do a lot more with my eyesight, but now it is the most important thing I have to preserve. I cannot read the screen that is well. Down the road when my eyes fail me as they are slowly doing, I may resort to one of those (a voice assistant) to read things. One of my favorite things is reading, and I would eventually...You know if my eyes go, that's where I might use it (a voice assistant).* [P3]

> *I’m not very good at spelling and with computer. It's kind of hard. And I am always afraid that I am going to lose everything by pressing the wrong button. Usually, when I text, I make mistakes. My fingers are too big for the little things, so I hit the wrong buttons usually, so this (a voice assistant) is much better, much faster. I can go about it easier than texting because you do not have to do the typing.* [P11]

However, most of the rest of the follow-up reactions were not positive, and no other potential benefits were discussed. The most prevalent response to voice assistants was *it’s good, but not for me*. Participants mentioned that a voice assistant might be useful for people other than themselves, such as younger populations with children:

> *My kids have two children who are really into this (a voice assistant), but it's not for me. I think it's a good thing for somebody who is into various things but for old folks like me we are content with what we are and sometimes ignorance is bliss.* [P8]

> *Probably when I was younger with my children and my husband and I had to do everything, you know, when my husband and I were with the children to help their homework and all, that (a voice assistant) would have come in very handy. But at this age there is nothing much to ask.* [P17]

The fact that many queries written on the cards were simple questions and commands might have influenced this perspective. Answering simple questions and executing simple commands that come up during day-to-day activities in the house is one primary feature of voice assistants. However, we found that this is not what our participants perceived to be the most useful:

> *I think this device is good in that it can give you the answers to general questions right away. But I'd say there is limited use for elderly people because you really do not get involved in things that you need to ask a bunch of questions about.* [P14]

It was then not just a first impression that older adults did not see direct benefits of using a voice assistant to them. A recent study showed that the attrition rate of voice assistants among older adults is high primarily due to lack of beneficial uses [3]. This implies that more features tailored to the needs of older adults are required to better assist them with voice assistants. In fact, several features designed specifically for older adults are already available, such as reminding about medications, sending alert messages verbally to their loved ones, and making emergency calls. Although these features were also included in the cards, participants did not find them useful either. Participants said that they were set in their own ways of doing these things and so did not need a new gadget to perform those tasks. For instance, they kept paper diaries for their schedules and used a pillbox to keep track of taking medications. Thus, participants did not perceive the features that are supposed to support older adults useful to them:

> *I don't have any interest in it (a voice assistant) because I keep my diaries to keep track of things and use a pill box to take medication. And I am content with it. I do not need anything else.* [P2]

Participants considered a voice assistant to be useful to older adults with aging-related physical constraints who can benefit from hands- and eyes-free interaction. This perspective led them to associate the use of a voice assistant with negative aspects of aging. Several participants mentioned that using voice assistants might make other people think that they were not capable of doing things on their own and needed support. Because aging-related changes are often associated with negative aspects, such as disability, stigma, and dependence, older adults tend to avoid supporting aids and assistive technologies, even though these can be beneficial to them [17,35]. We found a similar pattern in that some participants perceived voice assistants as yet another aid for aging-related declines, associated its use with negative stereotypes of aging, and rejected its adoption. This illustrates that voice assistants may fail regardless of the useful features it offers unless such negative stereotypes are mitigated:

> *I am healthy enough. My memory is still good enough to remember things. I do not need a device that tells me to remind me of things to do. I do not need a device that tells me to do things. I can still do things on my own.* [P10]

> *I think it (a voice assistant) can be useful for those with Alzheimer or old people who cannot move around or cannot do things on their own. I am not like that. I have no problem moving around and doing things. I may use it years later, but not yet.* [P15]

Discussion

Our findings revealed the patterns of use, difficulties, and perspectives that older adults might have when they first interact with a voice assistant. On the basis of these findings, we discuss design strategies for voice assistants that would allow older adults to have a positive first impression and to better leverage the capabilities of this technology. The design strategies include helping to better understand how a voice assistant works, incorporating mistakes and common interaction patterns into its design, and providing features tailored to the needs of older adults.

Help Understand How a Voice Assistant Operates

We found that older adults might have several misperceptions about how a voice assistant operates, which negatively contribute to their perspectives on using it. Three primary misperceptions include perceived efforts to store information before its use, privacy concerns associated with data storage and retrieval, and the cost of its use. In fact, these misperceptions might be universal for all first-time users. However, further considerations should be made to support older adults as they tend to lack self-efficacy about technology and thus experience much more difficulty understanding even basic concepts of new technologies compared with their younger counterparts [14]. Removing these misperceptions and helping them easily understand the basic concepts of how a voice assistant operates would be the first step toward lowering barriers to entry and helping novice older users better explore and facilitate its capabilities.

An immediate and straightforward solution is to provide a voice-based tutorial or educational application to address common misperceptions. Although a user can find out how a voice assistant operates by simply asking relevant questions to it, our participants did not know even what they did not know or what they misunderstood. Therefore, an introductory discourse-based tutorial that explains the basic concepts of how a voice assistant operates might be useful. However, providing a tutorial should be considered a temporary remedy because a well-designed user interface must be intuitive enough for novice users without needing any manuals or instruction.

A long-term solution is to incorporate the basic concepts of how a voice assistant operates into a responding answer, such as the source of information or how to handle personal information,

at least in the first few interactions. Although this might increase the number of responses and the total interaction time, successively providing relevant content upon request can reduce mental and temporal burdens on users. In addition, it could be more important for novice users to have an appropriate mental model and understand the system rather than an efficient user experience [32]. For instance, the system can contextualize basic instruction about its operation in the responding answer. The system then completes the response by asking if a user wants to hear more about how it operates. Upon a user's request, the system can provide more detailed information about the operational mechanism. The system can automatically and gradually reduce the contents related to operating instructions with more usage. The example conversational structure is as follows:

User: Alexa, what is tomorrow's weather going to be like?

VA: Let me check weather information online. According to AccuWeather, tomorrow's temperature will be...Do you want to hear more about how I instantly retrieved this information from AccuWeather?

User: Yes, please.

VA: I am connected to the Internet to search for information...

Incorporate Common Mistakes and Usage Patterns into Voice Interaction

Even though speech is supposed to be easier than typing or clicking, our participants still had difficulties conversing with a voice assistant. We found that a command query should be structured in a particular way that a voice assistant can comprehend (eg, a wake word followed by a concise and definitive sentence for a task after a brief pause) was a primary challenge, making the interaction not truly conversational but rather a series of one-directional comments. Participants especially expressed significant frustration when they realized only after completing a lengthy command that a voice assistant did not activate. One solution to this problem is to separate a wake word and a content sentence in a command structure. That is, instead of speaking the entire sentence for a task at once, a user would first speak a wake word to activate a voice assistant, just like calling someone's name to draw attention. A voice assistant would then make a simple greeting comment to indicate that it is ready to take commands. This will prevent users from mistake in speaking to a voice assistant when it is not activated. The example conversational structure is as follows:

User: Alexa

VA: Hi John, is there anything that I can do for you?

User: Yes, what is tomorrow's weather going to be like?

VA: The weather tomorrow will be...

One prominent usage pattern we found was that participants completed most interactions with a voice assistant by expressing gratitude to or giving feedback about the quality of a voice assistant's answer. Although this pattern might be due to the novelty effect, it is still worth considering this usage pattern in the human-agent conversation, since it can provide positive and more conversation-like experiences to novice users. For instance, instead of letting a user complete the interaction, a voice assistant would respond back to the user's comment and offer suggestions for more features. The example conversational structure is as follows:

User: Alexa, what is tomorrow's weather going to be like?

VA: The weather tomorrow will be...

User: That's great. Thanks.

VA: It's my pleasure. Do you want to check the weather in the next four days?

Explore Features Tailored to the Needs of Older Adults

The topics of frequently asked commands provide us with a clue to ideas for new features that older adults might find useful. Two topics of primary interest that emerged from our participants' interactions with a voice assistant include seeking information about health conditions and medication and streaming music from the old days. Thus, enriching the responding contents when executing these commands might give older adults a chance to find more features and functionalities, such as providing relevant or personalized extra information (eg, suggesting more songs to play) or suggesting other features relating to the command (eg, offering local contacts for health care services, providing alternative supplement or medication information).

Although this solution can expose novice older users to a range of new features, a more fundamental question lies in how to positively frame the perspectives that older adults might have about a voice assistant. A positive frame would be to take it from a tool to answer innocuous questions to a useful device that could address the more crucial needs of older adults. Older adults are more likely to adopt a new technology that helps them remain independent, allows them to have control and authority over its features and functions, and does not show signs of aging or frailty [36]. Thus, one solution is to associate the utility of voice assistants with positive aspects and the assets of aging, such as older adults' skills, knowledge, and resources. For instance, a voice assistant can be a gateway to connect peers within a community (eg, residents of an assisted-living facility) for labor- or information-sharing and social engagement. In this way, the role of a user can be reframed from a passive recipient of services to a proactive actor to provide information and services to others in a community. The mental model of a voice assistant can then be reshaped from a device to support the negative aspects of aging to a pathway to constructively engage in the community. The example conversational structure can be expressed as follows:

User1: Alexa, I am going grocery shopping at 11AM this Saturday and have two seats available to give a ride.

VA: Okay, I will let you know if anyone needs a ride.

User2: Alexa, is anyone going grocery shopping this Saturday? I need a ride.

VA: I found one person. You can get a ride from Jason at 11AM in front of the main entrance…

This solution and all of the previously suggested strategies are just a starting point for further investigation. More research is essential to gain a better understanding of how older adults would naturally interact with a voice assistant and build more naturalistic conversational interfaces to support better human-agent conversations.

Limitations

Our findings must be evaluated within the context of several limitations. First, our sample size was small, and thus our participant pool may not be representative of a general population. Second, we used convenience sampling for recruitment by recruiting participants from an assisted-living facility, which also runs the risk of compromising generalizability. Selection bias or unmeasured factors (eg, the homogeneity of participant characteristics by living in the same facility) could have influenced the responses during the interviews. Third, our study only investigated the first interactions and experiences that a potential user had when introduced to a voice assistant. Such behaviors might be different from those of users who own a device and use it for their real daily needs. However, we believe that it is important to identify common difficulties that older novice users have in their first interactions with a voice assistant, as it can provide new insights about the design of a device.

Conclusions

Personal technologies have been considered a breakthrough to tackle challenges associated with aging, and efforts have been made to develop design strategies that meet the needs of the aging population. As part of this effort, this paper explored how older adults would perceive and experience a voice assistant, one fast-growing type of personal technology, when they first interact with it. From interviews with 18 people aged 74 years or above who had never used a voice assistant, we investigated the patterns of use, difficulties, and perspectives that novice older users have when they use a voice assistant for the first time. On the basis of these findings, we discuss design implications that can positively influence older adults' first experiences with a voice assistant, including helping older adults better understand how a voice assistant works, incorporating mistakes and common interaction patterns into its design, and providing features tailored to the needs of older adults. We are hopeful that these findings can be used to expand our knowledge and practices for leveraging emerging personal technologies, a smart speaker–based voice assistant, to support the aging society.

Conflicts of Interest

None declared.

References

1. Richter F. Smart speaker adoption continues to rise. Statista. 2020. URL: https://www.statista.com/chart/16597/smart-speaker-ownership-in-the-united-states/ [accessed 2020-12-22]
2. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2011 Oct 2;17(1):127-144 [FREE Full text] [doi: 10.1007/s00779-011-0470-5]
3. Trajkova M, Martin-Hammond A. Alexa is a toy?: exploring older adults' reasons for using, limiting, and abandoning echo. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: The ACM Conference on Human Factors in Computing Systems; April, 2020; Honolulu p. 1-13 URL: https://doi.org/10.1145/3313831.3376760 [doi: 10.1145/3313831.3376760]
4. Kowalski J, Jaskulska A, Skorupska K, Abramczuk K, Biele C, Kope W, et al. Older adults and voice interaction: a pilot study with Google Home. 2019 Presented at: The ACM conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-6 URL: https://doi.org/10.1145/3290607.3312973 [doi: 10.1145/3290607.3312973]
5. Sayago S, Neves B, Cowan B. Voice assistants and older people: some open issues. In: Proceedings of the 1st International Conference on Conversational User Interfaces. 2019 Presented at: The International Conference on Conversational User Interfaces; August, 2019; Dublin p. 1-3 URL: https://doi.org/10.1145/3342775.3342803 [doi: 10.1145/3342775.3342803]
6. Charness N, Best R, Evans J. Supportive home health care technology for older adults: Attitudes and implementation. Gerontechnology 2016 Mar 23;15(4):233-242 [FREE Full text] [doi: 10.4017/gt.2016.15.4.006.00] [Medline: 29033700]
7. von der Pütten AM, Krämer NC, Gratch J, Kang SH. 'It doesn't matter what you are!' Explaining social effects of agents and avatars. Comp in Hum Beh 2010 Nov;26(6):1641-1650 [FREE Full text] [doi: 10.1016/j.chb.2010.06.012]
8. Lopatovska I, Oropeza H. User interactions with 'Alexa' in public academic space. 2019 Feb Presented at: Proceedings of the Association for Information Science and Technology; 2018; Vancouver p. 309-308 URL: https://doi.org/10.1002/pra2.2018.14505501034 [doi: 10.1002/pra2.2018.14505501034]
9. Chen K, Chan AH. A review of technology acceptance by older adults. Gerontechnology 2011;10(1):1-12 [FREE Full text] [doi: 10.4017/gt.2011.10.01.006.00]
10. Cotten S, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-761 [FREE Full text] [doi: 10.1093/geronb/gbu018] [Medline: 24671896]

11. Price M, Pak R, Müller H, Stronge A. Older adults' perceptions of usefulness of personal health records. Univ Access Inf Soc 2013;12(2):191-104 [FREE Full text] [doi: 10.1007/s10209-012-0275-y]
12. Dogruel L, Joeckel S, Bowman N. The use and acceptance of new media entertainment technology by elderly users: development of an expanded technology acceptance model. Behaviour & Information Technology 2015 Sep 07;34(11):1052-1063. [doi: 10.1080/0144929X.2015.1077890]
13. Barnard Y, Bradley M, Hodgson F, Lloyd A. Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Comp in Hum Beh 2013 Jul;29(4):1715-1714 [FREE Full text] [doi: 10.1016/j.chb.2013.02.006]
14. Kim S, Gajos K, Muller M, Grosz B. Acceptance of mobile technology by older adults: a preliminary study. 2016 Presented at: The International Conference on Human-Computer Interaction with Mobile Devices and Services; September, 2016; Florence, Italy p. 147-157 URL: https://doi.org/10.1145/2935334.2935380 [doi: 10.1145/2935334.2935380]
15. Candello H, Pinhanez C, Pichiliani M, Cavalin P, Figueiredo F, Vasconcelos M, et al. The effect of audiences on the user experience with conversational interfaces in physical spaces. 2019 Presented at: The ACM Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-13 URL: https://doi.org/10.1145/3290605.3300320
16. Porcheron M, Fischer JE, Reeves S, Sharples S. Voice interfaces in everyday life. 2018 Presented at: The ACM Conference on Human Factors in Computing Systems; April, 2018; Montreal p. 1-12 URL: https://doi.org/10.1145/3173574.3174214 [doi: 10.1145/3173574.3174214]
17. Birken T, Pelizäus-Hoffmeister H, Schweiger P. Judging the desirability and acceptance of assistance systems for the elderly--lessons learned with a fieldwork approach. 2016 Presented at: 49th Hawaii International Conference on System Sciences; 2016; Hawaii p. 579-578 URL: https://10.1109/HICSS.2016.78
18. Vtyurina A, Fourney A. Exploring the role of conversational cues in guided task support with virtual assistants. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: The ACM Conference on Human Factors in Computing Systems; April, 2018; Montreal p. 1-7 URL: https://doi.org/10.1145/3173574.3173782 [doi: 10.1145/3173574.3173782]
19. Murad C, Munteanu C. "I don't know what you're talking about, HALexa": the case for voice user interface guidelines. In: Proceedings of the 1st International Conference on Conversational User Interfaces. 2019 Presented at: The International Conference on Conversational User Interfaces; August, 2019; Dublin p. 1-3 URL: https://doi.org/10.1145/3342775.3342795 [doi: 10.1145/3342775.3342795]
20. Fischer J, Reeves S, Porcheron M, Sikveland R. Progressivity for voice interface design. 2019 Presented at: The International Conference on Conversational User Interfaces; August, 2019; Dublin Ireland p. 1-8 URL: https://doi.org/10.1145/3342775.3342788 [doi: 10.1145/3342775.3342788]
21. Metatla O, Oldfield A, Ahmed T, Vafeas A, Miglani S. Voice user interfaces in schools: co-designing for inclusion with visually-impaired and sighted pupils. 2019 Presented at: The ACM Conference on Human Factors in Computing Systems; May, 2019; New York p. 1-15 URL: https://doi.org/10.1145/3290605.3300608 [doi: 10.1145/3290605.3300608]
22. Pradhan A, Mehta K, Findlater L. 'Accessibility came by accident': use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: The ACM Conference on Human Factors in Computing Systems; April, 2018; Montreal p. 1-13 URL: https://doi.org/10.1145/3173574.3174033 [doi: 10.1145/3173574.3174033]
23. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput-Hum Interact 2020 Sep 25;27(4):1-27 [FREE Full text] [doi: 10.1145/3373759]
24. Sanders J, Martin-Hammond A. Exploring autonomy in the design of an intelligent health assistant for older adults. In: Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion. 2019 Presented at: 24th International Conference on Intelligent User Interfaces: Companion; March, 2019; California p. 95-96 URL: https://doi.org/10.1145/3308557.3308713 [doi: 10.1145/3308557.3308713]
25. Shade M, Rector K, Kupzyk K. Verification of pain medication adherence in older adults using interactive voice reminders. Innov Aging 2019;3(Supplement_1):S926 [FREE Full text] [doi: 10.1093/geroni/igz038.3371]
26. Morris ME, Adair B, Miller K, Ozanne E, Hansen R, Pearce AJ, et al. Smart-home technologies to assist older people to live well at home. J of Aging Sci 2013;1(1):1-9. [doi: 10.4172/2329-8847.1000101]
27. Blair J, Abdullah S. Understanding the needs and challenges of using conversational agents for deaf older adults. New York: Association for Computing Machinery; 2019 Presented at: Computer Supported Cooperative Work and Social Computing; 2019; Austin TX USA p. 161-165 URL: https://doi.org/10.1145/3311957.3359487 [doi: 10.1145/3311957.3359487]
28. Peek ST, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
30. Rosales A, Fernández-Ardèvol M. Smartphones, apps and older people's interests: from a generational perspective. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2016

Presented at: The ACM Conference on Human-Computer Interaction with Mobile Devices and Services; September, 2016; Florence p. 491-403 URL: https://doi.org/10.1145/2935334.2935363 [doi: 10.1145/2935334.2935363]
31. Clark L, Pantidi N, Cooney O, Doyle P, Garaialde D, Edwards J, et al. What makes a good conversation?: challenges in designing truly conversational agents. 2019 May Presented at: The ACM Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-12 URL: https://doi.org/10.1145/3290605.3300705 [doi: 10.1145/3290605.3300705]
32. Luger E, Sellen A. 'Like having a really bad PA': the gulf between user expectation and experience of conversational agents. 2016 Presented at: The ACM conference on human factors in computing systems; May, 2016; San Jose p. 5286-5297 URL: https://doi.org/10.1145/2858036.2858288 [doi: 10.1145/2858036.2858288]
33. Lau J, Zimmerman B, Schaub F. Alexa, are you listening? Proc ACM Hum Comput Interact 2018 Nov;2(CSCW):1-31 [FREE Full text] [doi: 10.1145/3274371]
34. Manikonda L, Deotale A, Kambhampati S. What's up with privacy?: user preferences and privacy concerns in intelligent personal assistants. 2018 Presented at: AAAI/ACM Conference on AI, Ethics, and Society; February, 2018; New Orleans p. 229-225 URL: https://doi.org/10.1145/3278721.3278773 [doi: 10.1145/3278721.3278773]
35. Neven L. 'But obviously not for me': robots, laboratories and the defiant identity of elder test users. Socio of Health & Illness 2010 Feb 01;32(2):335-347 [FREE Full text] [doi: 10.1111/j.1467-9566.2009.01218.x] [Medline: 20149151]
36. Lee C, Coughlin J. Perspective: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759 [FREE Full text] [doi: 10.1111/jpim.12176]

Edited by G Eysenbach, L Buis; submitted 18.05.20; peer-reviewed by K Rector, D Tao, J Offermann-van Heek; comments to author 01.09.20; revised version received 07.09.20; accepted 15.10.20; published 13.01.21.

Please cite as:
Kim S
Exploring How Older Adults Use a Smart Speaker–Based Voice Assistant in Their First Interactions: Qualitative Study
JMIR Mhealth Uhealth 2021;9(1):e20427
URL: http://mhealth.jmir.org/2021/1/e20427/
doi:10.2196/20427
PMID:33439130

Original Paper

Mobile Phone Access and Implications for Digital Health Interventions Among Adolescents and Young Adults in Zimbabwe: Cross-Sectional Survey

Aoife M Doyle[1], BSc, MPhil, PhD; Tsitsi Bandason[2], MSc; Ethel Dauya[2], MPH; Grace McHugh[2], MD; Chris Grundy[1], MSc; Stefanie Dringus[1], PhD; Chido Dziva Chikwari[1,2,3], MSc; Rashida A Ferrand[2,3], PhD

[1]MRC International Statistics and Epidemiology Group, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]Biomedial Resarch and Training Institute, Harare, Zimbabwe

[3]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Aoife M Doyle, BSc, MPhil, PhD
MRC International Statistics and Epidemiology Group
London School of Hygiene & Tropical Medicine
Keppel St
London, WC1E7HT
United Kingdom
Phone: 44 207 927 2041
Email: aoife.doyle@lshtm.ac.uk

Abstract

Background: Mobile phones may help young people (YP) access health information and support health service engagement. However, in low-income settings there is limited knowledge on YP's phone and internet access to inform the feasibility of implementing digital health interventions.

Objective: We investigated access to information and communication technologies among adolescents and young adults in Zimbabwe.

Methods: A cross-sectional population-based survey was conducted from October to December 2018 among YP aged 13-24 years in 5 communities in urban and peri-urban Harare and Mashonaland East, Zimbabwe. Consenting YP completed a self-completed tablet-based questionnaire on mobile phone ownership and use, and use of the internet. The primary outcome was the proportion who reported owning a mobile phone. Secondary outcomes included phone and internet access and use behavior, and ownership and use of other technological devices. Multivariable logistic regression was used to investigate factors associated with mobile phone ownership and with internet access, with adjustment for the one-stage cluster sampling design. A priori exploratory variables were age, sex, marital status, and urban/peri-urban residence.

Results: A total of 634/719 (88.2%) eligible YP, mean age 18.0 years (SD 3.3) and 62.6% (397/634) females, participated. Of the YP interviewed, 62.6% (396/633; 95% CI 58.5-66.5) reported owning a phone and a further 4.3% (27/633) reported having access to a shared phone. Phone ownership increased with age: 27.0% (43/159) of 13-15-year olds, 61.0% (72/118) of 16-17-year olds, 71.5% (103/144) of 18-19-year olds, and 84.7% (171/202) of 20-24-year olds (odds ratio [OR] 1.4, 95% CI 1.3-1.5) per year increase. Ownership was similar among females and males: 61.0% (236/387; 95% CI 55.6-66.1) versus 64.8% (153/236; 95% CI 57.8-71.2), age-adjusted OR 0.7 (95% CI 0.5-1.1); higher in those with secondary level education compared to primary or no education: 67.1% (346/516; 95% CI 62.6-71.2) versus 26% (21/82; 95% CI 16.4-37.7), age-adjusted OR 2.3 (95% CI 1.1-4.8); and similar across other sociodemographic factors. YP reported that 85.3% (361/423) of phones, either owned or shared, were smartphones. Among phone owners, the most commonly used phone app was WhatsApp (71.2%, 282/396), and 16.4% (65/396) reported having ever used their phone to track their health. A total of 407/631 (64.5%; 95% CI 60.3-68.5) currently had access to the internet (used in last 3 months on any device) with access increasing with age (OR 1.2, 95% CI 1.2-1.3 per year increase). In age-adjusted analysis, internet access was higher among males, the unmarried, those with a higher level of education, phone owners, and those who had lived in the community for more than 1 year. The aspect of the internet that YP most disliked was unwanted sexual (29.2%, 136/465) and violent (13.1%, 61/465) content.

Conclusions: Mobile phone–based interventions may be feasible in this population; however, such interventions could increase inequity, especially if they require access to the internet. Internet-based interventions should consider potential risks for participants and incorporate skill-building sessions on safe internet and phone use.

(JMIR Mhealth Uhealth 2021;9(1):e21244) doi:10.2196/21244

KEYWORDS

adolescent; young adult; young person; young people; cross-sectional studies; humans; female; male; mobile phone; smartphone; cell phones; technology; internet; safety; health-related internet use; Zimbabwe

Introduction

There is a growing interest in the use of mobile phones to help young people (YP) access health information, and to support their engagement with health services. Data on YP's use of information and communication technology (ICT), including mobile phones, are limited, particularly in low-income countries [1]. Such information is needed to inform the development of feasible and equitable digital health interventions. Data suggest that gender and socioeconomic gaps in access still exist in many countries [1], and there is a risk that the introduction of digital health interventions may increase inequity in access to health information and services. The importance of enhancing the use of enabling technology to promote sustainable development, gender equality, and the empowerment of all women and girls has been recognized in Sustainable Development Goals 17 and 5 [2].

The 2015 Zimbabwean Demographic and Health Survey found that 87% of households owned a mobile phone [3]. However, there is little quantitative data on YP's access to ICT, for example, mobile phones, their patterns of use of ICT, and whether confidentiality would be a concern when communicating via phone or internet on sensitive topics. In particular, while it is widely believed that mobile phone use among YP is high, the functionality of the phones that are being used, preferences for platforms and apps, and the extent of potential challenges to intervention uptake such as confidentiality, cost, internet coverage or speed remain unknown.

With one-fifth of the Zimbabwean population aged between 15 and 24 years, YP's health is central to the country's development [4]. However, health service uptake by YP lags behind need. In this high HIV prevalence setting approximately half of all HIV-positive 15-24-year olds are unaware of their status [5] and use of preventive services such as contraception and voluntary male circumcision fall below national targets [6].

The aim of this study was to collect data on YP's use of information and communication technology to inform the feasibility of implementing technology-based adolescent-health interventions in Zimbabwe.

Methods

Recruitment

A cross-sectional population-based survey was conducted from October to December 2018 in 3 urban communities (A, B, and C) in Harare province and 2 peri-urban communities (D and E) in Mashonaland East province. These 5 communities participated in formative work for the ongoing CHIEDZA sexual and reproductive health services intervention trial. The communities had been purposively selected to represent the urban and peri-urban communities that would be included in the trial. The survey was conducted in these communities so that the survey team could benefit from the existing research infrastructure and stakeholder relationships. Eligible participants were aged 13-24 years, resident in the study community at the time of the survey, and either provided informed consent (16-24 years) or provided assent with guardian consent (age 13-15 years).

We estimated that the prevalence of mobile phone (Textbox 1) ownership among 13-24-year olds would be 50%. Assuming 10% nonresponse and a design effect of 2 [3], a sample size of 686 YP would provide ±8% precision around this estimate. Using stratified sampling we aimed to recruit 60.1% (412/686) of participants from Harare, and 39.9% (274/686) of participants from Mashonaland East. A simple random sample of 100 GPS coordinates (primary sampling unit) was sampled per cluster from all potential points in the study areas using ArcGIS software version 10.5 (Esri). Points were randomly ordered and then sequentially visited by a team of interviewers. All households with front doors within 20 m of the sampled GPS point were visited. The household head was interviewed to obtain basic demographic information about the household and to obtain consent to interview any eligible YP. If the household head was not available, another household member aged 16+ years or a neighbor was asked to provide information on the composition of the household. Households with YP were visited a further two times in order to interview the household head. All YP in the selected household were eligible for recruitment.

Textbox 1. Definition of terms used in this paper.

Household: A person or a group of related and unrelated persons who live together in the same dwelling unit(s), acknowledge 1 adult male or female as the head of the household, share the same housekeeping arrangements, and are considered a single unit. Household members were defined as individuals who have lived or intended to live in the household for 1 or more months, including school children regularly in residence during the school year [3].

Internet access: A person was considered to have internet access if they reported accessing the internet once or more in the last 3 months, including on a device belonging to a family member or employer (International indicator HH7) [7].

Basic phone: Mobile phone with limited features (no web browser or apps) that is used primarily for phone calls and sending SMS text messages.

Feature phone: Mobile phone with more features than a basic phone and usually has a camera, supports some apps but not all third-party apps, and features a web browser.

Smartphone: Mobile phone built on a mobile computing platform (eg, Apple OS, Android) and supports third-party apps.

Phone ownership: Has sole ownership of a mobile phone.

Phone sharing: Having joint ownership of or access to someone else's mobile phone.

Primary phone: The phone that the respondent reports as the main phone that they use.

Technological devices: Desktop computer, laptop computer, tablet/iPad, mobile phone, iPod or other MP3 player, TV, radio, digital camera, gaming console, handheld gaming device.

Data Collection

Participants responded to a 30-minute audio computer-assisted self-interviewing (ACASI) tablet-based questionnaire (Multimedia Appendix 1). ACASI was facilitated by trained research assistants who oriented the interviewees and were present during the interview to troubleshoot or answer questions. Questionnaire topics included use and ownership of technological devices including mobile phone, and access to and use of the internet. Questions were adapted from pre-existing questionnaires [8-11]. The questionnaire was developed in English and translated into Shona (the local language). Modifications were made to the questionnaire following pretesting with the study team and following the pilot survey which was conducted outside the selected study sites.

Data Management and Analysis

The primary outcome was the prevalence of mobile phone ownership among 13-24-year olds. Secondary outcomes were the characteristics of mobile phones and phone use behavior; internet access and use behaviors; and ownership and use of other technological devices such as tablets, desktops, and laptops. Data were collected and recorded using Open Data Kit survey software with built-in logical checks and skip patterns on Android tablets. Data were analyzed using STATA version 15.1 (StataCorp). Using sampling weights and robust standard errors to account for the clustered sampling design (one-stage cluster sampling). Multivariable logistic regression was used to calculate age-adjusted odds ratios for the association between explanatory factors and mobile phone ownership, and with internet use. Potential explanatory variables were age, sex, marital status, community of residence, highest level of school attended, current occupational status, religion, travel for at least 1 month in past 12 months, length of time living in the community, and orphan status, with age being considered an a priori potential confounder. Wald tests adjusted for the clustered sampling design were used at each step of the analysis.

Ethical Considerations

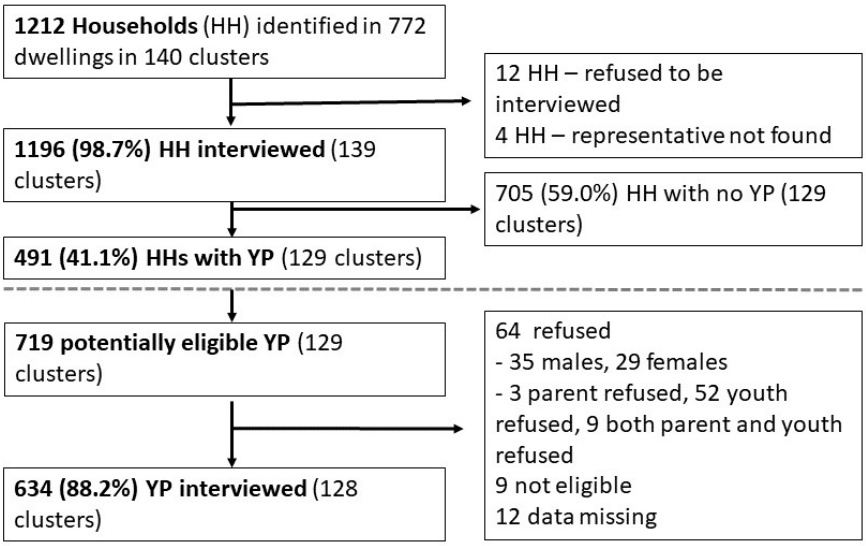
Ethical approval was obtained from the Institutional Review Board of the Biomedical Research and Training Institute (AP149/2018), the Medical Research Council of Zimbabwe (MRCZ/A/2362), and the London School of Hygiene and Tropical Medicine Research Ethics Committee (LSHTM REC, No. 15919). Written informed consent was obtained from parents or guardians of study participants aged below 16 years, along with participant assent. Participants aged 16 years and older consented independently.

Results

Study Population

In total, 1212 households were sampled from 140 GPS point clusters in 5 suburbs (A 25, B 48, C 21, D 42, and E 4; Figure 1). A total of 719 YP in the target age range were identified from 491 households (41.05% [491/1196] of successfully interviewed households); 634/719 (88.2%) YP were included in the study with 633/634 providing information on mobile phone ownership. Fewer GPS point clusters in community E were visited and only 10 participants were interviewed in that community. Community E was included in the descriptive analysis but excluded from regression analysis.

Figure 1. Survey recruitment (HH Household, YP young people).



The mean age of the 634 participants was 18.0 years (SD 3.3) and 62.6% (397/634) were female. The majority (83.9%, 532/634) had never been married, 86.8% (550/634) had attended secondary school or higher, and only 14.7% (51/346) of out-of-school participants reported that they were working. The majority were Christian, had lived in the study community for at least 5 years, and had not traveled for at least 1 month in the past 12 months. Approximately one-third of respondents reported that one or both of their parents were dead or that a parent's location was unknown (Table 1).

Table 1. Demographic characteristics of the study population (N=634).

Demographic characteristic	Sex of respondent		
	Male (N=237, 37.4%)	Female (N=397, 62.6%)	Total (N=634)
Age group (years), n (%)			
13-15	65 (27.4)	96 (24.2)	161 (25.4)
16-17	47 (19.8)	74 (18.6)	121 (19.1)
18-19	55 (23.2)	89 (22.4)	144 (22.7)
20-24	70 (29.5)	138 (34.8)	208 (32.8)
Mean age (years), mean (95% CI)	17.8 (17.4-18.2)	18.2 (17.8-18.5)	18.0 (17.8-18.3)
Marital status, n (%)			
Married	5 (2.1)	64 (16.1)	69 (10.9)
Cohabiting	1 (0.4)	15 (3.8)	16 (2.5)
Never married	225 (94.9)	307 (77.3)	532 (83.9)
Divorced/separated	6 (2.5)	11 (2.8)	17 (2.7)
Highest level of school attended, n (%)			
Primary	32 (13.5)	50 (12.6)	82 (12.9)
Secondary	195 (82.3)	330 (83.1)	525 (82.8)
Higher (Tertiary)	9 (3.8)	16 (4.0)	25 (3.9)
Never been to school	1 (0.4)	1 (0.3)	2 (0.3)
Current occupational status, n (%)			
In school/university	121 (51.1)	167 (42.1)	288 (45.4)
Out of school (working)	22 (9.3)	29 (7.3)	51 (8.0)
Out of school (not working)	94 (39.7)	201 (50.6)	295 (46.5)
Religion[a], n (%)			
Roman Catholic	27 (11.5)	38 (9.6)	65 (10.3)
Protestant	55 (23.5)	102 (25.7)	157 (24.9)
Pentecostal	96 (41.0)	169 (42.6)	265 (42.0)
Apostolic sect	19 (8.1)	67 (16.9)	86 (13.6)
Other Christian/Muslim/Other	5 (2.1)	3 (0.8)	8 (1.3)
No religion	32 (13.7)	18 (4.5)	50 (7.9)
Traveled for at least 1 month in past 12 months, n (%)			
No	170 (71.7)	282 (71.0)	452 (71.3)
Yes	67 (28.3)	115 (29.0)	182 (28.7)
How long lived in community?[b], n (%)			
<1 year	29 (12.3)	83 (20.9)	112 (17.7)
1-4 years	44 (18.7)	98 (24.7)	142 (22.5)
5+ years	162 (68.9)	216 (54.4)	378 (59.8)
Orphan status, n (%)			
Double orphan	23 (9.7)	40 (10.1)	63 (9.9)
Mother dead, father alive	21 (8.9)	31 (7.8)	52 (8.2)
Mother alive, father dead, or unknown	40 (16.9)	70 (17.6)	110 (17.4)
Both parents alive	153 (64.6)	256 (64.5)	409 (64.5)

[a]n=3 no response (men only).

[b]n=2 do not know (men only).

Mobile Phone Ownership and Access

The prevalence of mobile phone ownership was 62.6% (396/633; 95% CI 58.5-66.5). Among the 237 who did not own a phone, 27 (11.4%) reported that they shared a phone (Table 2). In total, 423 out of the 633 YP interviewed (66.8%; 95% CI 62.3, 71.1) reported either owning or sharing a phone, with 18.0% (76/423) currently using (owning or sharing) 2 or more phones (Multimedia Appendix 2). The use of multiple phone numbers was common, with 26.5% (112/423) currently using and 42.8% (181/423) having used more than 1 phone number in the past year. The majority (85.3%, 361/423) of primary phones, either owned or shared, were reported to be smartphones.

Female phone sharers reported sharing phones with their mother (18/50, 36%), partner/boyfriend (14/50, 28%), or siblings (12/50, 24%), whereas male phone sharers reported sharing phones primarily with their siblings (11/20, 55%). Almost all (67/70, 96%) respondents who reported sharing phones did so at least once a week (Multimedia Appendix 2).

Table 2. Prevalence of phone ownership and phone sharing.

	Shares a phone					
	No		Yes		Total[a]	
Owns a phone	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
No	210 (33.2)	29.0-37.7	27 (4.3)	2.6-6.9	237 (37.4)	33.5-41.5
Yes	353 (55.8)	51.2-60.2	43 (6.8)	4.6-9.9	396 (62.6)	58.5-66.5
Total	563 (88.9)	85.0-91.9	70 (11.1)	8.1-15.0	633	Not applicable

[a]One participant did not respond to the question “Do you have or use a mobile phone?”

The main reasons for not owning or sharing a phone were have/had a phone but it is not working (27.6%, 58/210), cost (21.9%, 46/210), and not being allowed (17.1%, 36/210). However, 63.8% (134/210) of those who did not have access to a phone reported planning to buy one in the near future (Multimedia Appendix 3).

The median age at first mobile phone use was 13 years (IQR 12-15) and 15 years (IQR 13-16) for male and female respondents, respectively (Multimedia Appendix 4). First phones were primarily purchased by parents (237/367, 64.6%) or other relatives (83/367, 22.6%).

Prevalence of phone ownership increased with age of the respondent with 27.0% (43/159) of 13-15-year olds, 61.0% (72/118) of 16-17-year olds, 71.5% (103/144) of 18-19-year olds, and 84.7% (171/202) of 20-24-year olds owning a phone (OR 1.4, 95% CI 1.3-1.5) for each year increase (P<.001). In age-adjusted analysis there was weak evidence that mobile phone ownership was higher in those with at least secondary level education compared to those with primary or no education (secondary OR 2.3, 95% 1.1-4.8; tertiary OR 2.6, 95% 0.6-11.9; P=.09). The prevalence of mobile phone ownership was similar among male (153/236, 64.8%; 95% CI 57.8-71.2) and among female respondents (236/387, 61.0%; 95% CI 55.6-66.1; age-adjusted OR 0.7, 95% CI 0.5-1.1; P=.11; Table 3).

Table 3. Factors associated with phone ownership (N=623).

Factors	n	Prevalence, %	Unadjusted odds ratio (95% CI)	Age-adjusted odds ratio (95% CI)
Age group (years)			*P*<.001	
13-15	159	27.0		
16-17	118	61.0		
18-19	144	71.5		
20-24	202	84.7		
Per-year increase			1.40 (1.30-1.52)	
Gender			*P*=.39	*P*=.11
Male	236	64.8	1[b]	1
Female	387	61.0	0.85 (0.58-1.24)	0.73 (0.49-1.07)
Marital status			*P*<.001	*P*=0.13
Married/cohabiting	79	79.8	1	1
Never married	527	59.4	0.37 (0.21-0.64)	1.67 (0.87-3.18)
Divorced/separated	17	76.5	0.83 (0.22-3.12)	0.64 (0.15-2.77)
Religion			*P*=0.13	*P*=0.10
Roman Catholic	64	75.0	1	1
Protestant	153	64.7	0.61 (0.31-1.22)	0.68 (0.33-1.40)
Pentecostal	264	60.2	0.50 (0.28-0.92)	0.53 (0.28-1.00)
Apostolic sect	85	56.5	0.43 (0.20-0.92)	0.46 (0.21-0.99)
Other Christian/Muslim/Other	7	71.4	0.83 (0.17-4.05)	1.06 (0.35-3.20)
No religion	47	57.5	0.45 (0.21-0.95)	0.42 (0.17-1.03)
Community			*P*=.37	*P*=.19
A	178	59.0	1	1
B	140	67.1	1.42 (0.89-2.28)	1.51 (0.91-2.52)
C	147	59.2	1.01 (0.67-1.52)	0.97 (0.60-1.57)
D	158	65.2	1.30 (0.79-2.16)	1.49 (0.82-2.69)
Highest level of school attended			*P*<.001	*P*=.09
None/Primary	82	25.6	1	1
Secondary	516	67.1	5.91 (3.22-10.86)	2.27 (1.08-4.77)
Higher (tertiary)	25	88.0	21.30 (5.44-83.41)	2.64 (0.58-11.94)
Current occupational status			*P*<.001	*P*=.30
In school/university	286	47.2	1	1
Out of school (working)	51	90.2	10.29 (3.97-26.66)	2.24 (0.78-6.45)
Out of school (not working)	286	72.7	2.98 (2.06-4.32)	1.09 (0.70-1.69)
Traveled for at least 1 month in the past 12 months			*P*=.41	*P*=.63
No	443	61.4	1	1
Yes	180	65.0	1.17 (0.81-1.69)	1.11 (0.72-1.71)
How long lived in community?			*P*=.08	*P*=.32
<1 year	112	63.4	1	1
1-4 years	137	70.8	1.40 (0.79-2.48)	1.51 (0.82-2.76)
5+ years	372	59.1	0.84 (0.50-1.39)	1.05 (0.64-1.71)
Orphan status			*P*=.05	*P*=.12

Factors	n	Prevalence, %	Unadjusted odds ratio (95% CI)	Age-adjusted odds ratio (95% CI)
Double orphan	61	62.3	1	1
Mother dead, father alive	52	75.0	1.82 (0.85-3.88)	1.96 (0.84-4.56)
Mother alive, father dead or unknown	107	70.1	1.42 (0.83-2.44)	1.92 (1.03-3.60)
Both parents alive	403	58.8	0.86 (0.50-1.50)	1.89 (1.07-3.34)

[a]n=623 as excludes 10 participants who were interviewed in community E.

[b]Reference.

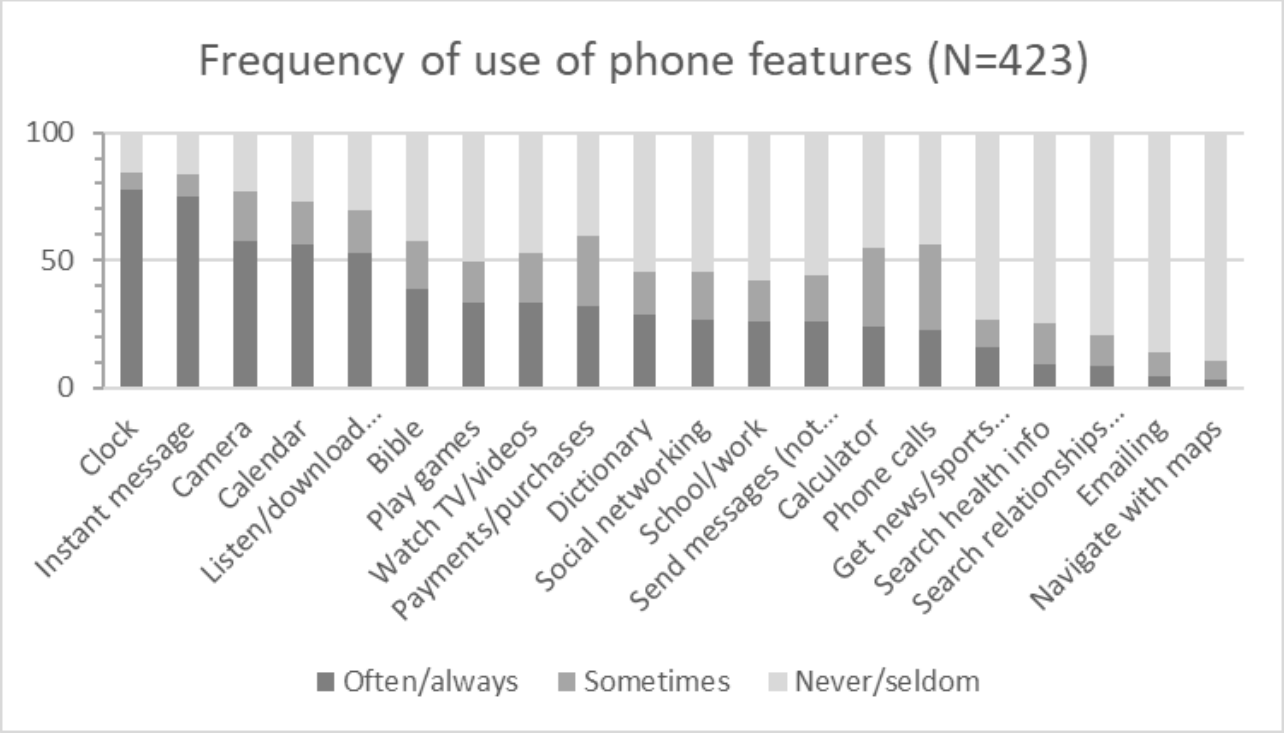
Phone Use Behavior

In total 4 in 10 phone owners reported that they never turned their phone off and 25.9% (103/397) reported that they could not do without their phone for a day. Among school-going phone-using respondents, just over half reported regularly bringing their phone to school. YP reported that the best thing about having a mobile phone was that it was convenient and made life easier (Multimedia Appendix 5).

The majority (280/423, 66.2%) of phone users spent US $1-3 per week on phone credit with 9.9% (42/423) spending nothing. Most phone users reported spending less on airtime in the past week when compared to other personal expenditure but 23.9% (101/423) reported having spent more on airtime. Phone credit was paid for by a combination of the respondent, their family members, or their friends. Half of females and a quarter of males reported that their boyfriend or girlfriend paid for phone credit (Multimedia Appendix 6).

The most commonly used phone features were the clock, instant messaging/chat, camera, and the calendar (Figure 2). The most commonly used app was WhatsApp (70.9%, 300/423). Other commonly used apps were Facebook, Facebook Messenger, internet browser, Instagram, Twitter, YouTube, dictionary, bible, and calculator (Figure 2, Multimedia Appendix 7). As many as 67.1% (108/161) of male and 56.9% (149/262) of female phone users reported playing games on their mobile phone. Candy Crush and Temple Run were the most popular games among females and FIFA, Temple Run, and Dream League were the most popular games among males.

Figure 2. Frequency of use of different phone features among phone users.



Among phone users, a quarter report at least sometimes searching for health information and 20.3% (86/423) for information on relationships (Multimedia Appendix 7). A minority (16.3%, 69/423) report having ever used their phone to track their health. Males were more likely than females to report having used a phone to track their health (21.1% [34/161] versus 13.4% [35/262], *P*=.03). When asked to list the apps that they used to track their personal health, the majority mentioned web browsers, YouTube, and social media platforms. Specific apps mentioned related to fitness, blood pressure, sugar levels, body temperature, HIV testing, skincare, period trackers, home remedies app, and health tips (including daily health tips).

In total, 4 in 10 phone owners (44.2%, 175/396) reported that the information stored on their phone was not private. A similar proportion thought that the information they sent on their phone

(40.4%, 160/396) and received on their phone (40.9%, 162/396) was not private. Only two-thirds (62.1%, 246/396) had a password to lock/unlock their phone and 17.7% (70/396) had passwords for any apps on their phone (Table 4).

Table 4. Security and privacy associated with phone use among phone owners (N=396).

Questions on security and privacy	n (%)	95% CI
How private do you consider the information that you send when using a phone?		
Very private	148 (37.4)	29.4-46.1
Somewhat private	88 (22.2)	17.6-27.7
Not private	160 (40.4)	33.5-47.8
How private do you consider the information that you receive when using a phone?		
Very private	145 (36.6)	29.3-44.7
Somewhat private	89 (22.5)	18.1-27.6
Not private	162 (40.9)	33.8-48.4
How private do you consider the information stored on your phone?		
Very private	146 (36.9)	29.8-44.6
Somewhat private	75 (18.9)	14.7-24.1
Not private	175 (44.2)	37.0-51.7
Do you have passwords to lock/unlock your phone?		
No	150 (37.9)	33.5-42.4
Yes	246 (62.1)	57.6-66.5
Do you have passwords for any apps on your phone?		
No	326 (82.3)	78.4-85.6
Yes	70 (17.7)	14.4-21.6

Internet Use

A total of 407/631 (64.5%) respondents (95% CI 60.3-68.5) had access to the internet (used in the last 3 months), with 73.8% (468/634) reporting ever using the internet. Ever internet users reported accessing the internet frequently (at least once per week) on a mobile phone (78.2%, 366/468) or on a computer at work or school (18.6%, 87/468). Frequent access to the internet on other computers was rare: commercial internet outlet (5.6%, 26/468), at home (6.8%, 32/468), at someone else's house (4.3%, 20/468), or in a library/community facility (4.1%, 19/468; Multimedia Appendix 8).

Internet access in the last 3 months increased with age (OR 1.2, 95% CI 1.2-1.3, per year increase; *P*<.001). In age-adjusted analysis, internet access was lower among females (adjusted OR 0.5, 95% CI 0.4-0.8; *P*=.001). Internet access was higher among the never married compared to the married and cohabitating (adjusted OR 2.8, 95% CI 1.5-5.5; *P*=.001), among those who had secondary education compared to primary or no education (adjusted OR 2.3, 95% CI 1.2-4.2; *P*=.03), and among those who had lived in the community for more than 1 year (*P*=.003; Table 5). Mobile phone owners had 9 times the odds of having access to the internet compared to nonphone owners (adjusted OR 8.7, 95% CI 5.6-13.5; *P*<.001). There was no evidence of a difference in internet access according to religion, community, travel in the past 12 months, or orphan status.

Table 5. Factors associated with internet access (N=621)[a].

Factors	n	Prevalence, %	Unadjusted odds ratio (95% CI)	Age-adjusted odds ratio (95% CI)
Age group (years)			*P*<.001	
13-15	156	37.8		
16-17	119	69.8		
18-19	144	73.6		
20-24	202	77.2		
Per year increase			1.22 (1.15-1.30)	
Gender			*P*=.007	*P*=.001
Male	234	72.7	1[b]	1
Female	387	60.5	0.58 (0.39-0.86)	0.52 (0.35-0.76)
Marital status			*P*=.34	*P*=.01
Married/cohabiting	79	65.8	1	1
Never married	525	64.4	0.94 (0.57-1.55)	2.83 (1.45-5.53)
Divorced/separated	17	82.4	2.42 (0.57-10.37)	2.19 (0.49-9.82)
Religion			*P*=.11	*P*=.13
Roman Catholic	63	77.8	1	1
Protestant	153	68.0	0.61 (0.30-1.24)	0.67 (0.31-1.45)
Pentecostal	264	64.4	0.52 (0.25-1.06)	0.56 (0.26-1.19)
Apostolic sect	84	53.6	0.33 (0.15-0.71)	0.35 (0.15-0.79)
Other Christian/Muslim/Other	8	62.5	0.48 (0.10-2.17)	0.56 (0.15-2.10)
No religion	46	60.9	0.44 (0.18-1.08)	0.44 (0.18-1.09)
Community			*P*=.06	*P*=.05
A	177	63.3	1	1
B	139	77.0	1.94 (1.07-3.51)	2.03 (1.11-3.71)
C	147	61.9	0.94 (0.61-1.46)	0.92 (0.58-1.47)
D	158	59.5	0.85 (0.53-1.38)	0.87 (0.51-1.50)
Highest level of school attended			*P*<.001	*P*=.03
None/Primary	82	35.4	1	1
Secondary	514	68.9	4.04 (2.40-6.82)	2.28 (1.23-4.21)
Higher (Tertiary)	25	84.0	9.59 (3.05-30.18)	2.99 (0.89-10.08)
Current occupational status			*P*=.009	*P*=.37
In school/university	285	58.3	1	1
Out of school (working)	51	80.4	2.94 (1.36-6.33)	0.90 (0.38-2.16)
Out of school (not working)	285	69.1	1.60 (1.07-2.41)	0.72 (0.45-1.17)
Traveled for at least 1 month in the past 12 months			*P*=.25	*P*=.32
No	442	63.6	1	1
Yes	179	68.7	1.26 (0.85-1.87)	1.23 (0.82-1.85)
How long lived in community?			*P*=.01	*P*=.003
<1 year	110	51.8	1	1
1-4 years	137	69.3	2.10 (1.25-3.55)	2.30 (1.29-4.12)
5+ years	372	67.5	1.93 (1.19-3.13)	2.50 (1.47-4.24)
Orphan status			*P*=.74	*P*=.28

Factors	n	Prevalence, %	Unadjusted odds ratio (95% CI)	Age-adjusted odds ratio (95% CI)
Double orphan	61	62.3	1	1
Mother dead, father alive	52	67.3	1.25 (0.59-2.64)	1.24 (0.55-2.79)
Mother alive, father dead, or unknown	108	68.5	1.32 (0.69-2.52)	1.55 (0.75-3.19)
Both parents alive	400	64.3	1.09 (0.60-1.98)	1.79 (0.94-3.41)
Owns a phone			*P*<.001	*P*<.001
No	232	33.6	1	1
Yes	388	84.0	10.38 (6.84-15.75)	8.67 (5.58-13.47)

[a]10 respondents from community E were excluded and 3 respondents did not provide information on the timing of most recent internet access.
[b]Reference.

The most common technological devices that respondents had at home were televisions (89.1%, 565/634), mobile phones (87.5%, 555/634), and radios (71.1%, 451/634). Ever use of technological devices was higher than household ownership but showed similar patterns with a high proportion reporting ever use of televisions, mobile phones, and radios. Over half of respondents had ever used a desktop computer (61.7%, 391/634) and laptop computer (65.6%, 416/634; Multimedia Appendix 9).

YP reported that the thing that they most disliked about the internet was seeing unwanted sexual content (29.2%, 136/465) and violent stories, photos, and videos (13.1%, 61/465). The most common suggestions on how to make the internet better were cheaper data plans (36.8%, 171/465), making access to mobile phones and computers easier (14.6%, 68/465), better internet coverage (12.9%, 60/465), and high-speed connectivity (12.7%, 59/465; Multimedia Appendix 10).

Discussion

Principal Findings

YP in Harare and Mashonaland East had high levels of phone ownership and internet access and access increased with age. A minority of YP used their phones to seek health information or to support their health. Challenges that YP face when using mobile phones and the internet include the cost of data, access to phones/computers, speed of connection, exposure to unwanted sexual and violent content, and concerns about security and confidentiality. Older age groups could be targeted for phone-based interventions but ensuring equitable access to data and charging facilities as well as training on safe internet use are necessary.

Limitations

This study only collected data on YP living in urban and peri-urban areas. An estimated 68% of Zimbabwe's population live in rural areas [12] and there is a well-documented digital divide with internet use much lower in rural compared to urban areas [13-15]. Data on contextual factors, such as the availability of electricity to charge devices, were not collected. A relatively high proportion of respondents reported access to a smartphone but we did not collect data on the functionality of the phones. We gathered limited information on respondents' current use of their phone to access health information and services. In-depth qualitative studies are needed to better understand their current use, including barriers and facilitators, and to explore willingness to use their phones to access information and services. Alternative ways to understand YP's digital lives, which could be considered for future studies, are the use of diaries or qualitative interviews which can probe for a detailed understanding of use throughout the day [16,17].

Comparison With Prior Work

This study provides a unique insight into phone ownership and use among YP in Zimbabwe as there is little published data available for this population. Phone ownership among 20-24-year olds in this study (84.7%, 171/202) was higher than national estimates from the most recent 2015 Zimbabwean DHS (71.1% among females and 76.5% among males) [18], but phone access was in line with levels seen in higher-income Sub-Saharan African (SSA) countries such as South Africa. A 2017 survey among adults aged over 18 in 6 African countries (excluding Zimbabwe) found that 72%-93% of 18-29-year olds reported owning a mobile phone, with the proportion owning smartphones ranging from 17% in Tanzania to 63% in South Africa [19]. A 2013/14 household survey among 9-18-year olds in South Africa, Ghana, and Malawi also found a large variation in phone ownership ranging from 6.2% in females in Malawi to 50.9% among males in South Africa [20,21]. A 2012 survey of secondary school students in South Africa found that 81.1% owned or had access to a mobile phone [14].

Although phone ownership was comparable between genders, males reported increased internet access/usage. Phone sharing was relatively common with differences in phone sharing between males and females, that is, who they share with. The lack of an association between gender and phone ownership observed in this study is consistent with findings from South Africa [20] and there is some evidence to suggest that as the prevalence of phone ownership increases, the gender divide decreases [22]. By contrast, we found higher access to the internet among males, those not currently married, those with greater than primary education, and longer-term residents. Observed gender differences in internet access are in keeping with other studies from SSA. A 2018 multicountry study among adolescents in low- and middle-income countries found that boys were more likely than girls to have smartphones (through which they can access the internet) and used a wider variety of phone features compared to girls [23]. A Ugandan study among

18-24-year olds found high phone ownership among both sexes but lower internet use among females compared to males [24]. The 2019 GSMA mobile gender report also found a bigger gap for internet use than phone use, with women also using a smaller range of services and spending less on their phones than men [25]. Global goals have been set to provide internet access for everyone and access to digital technology is considered an important component to help adolescents achieve their rights [26-28]. The findings from this study contribute to our understanding of which YP have and do not have access to phones and the internet.

Implications for Interventions

Some kinds of phone-based interventions are likely to be feasible in this population and phones are increasingly being used in Zimbabwe to facilitate health information and service delivery [29-31]. A quarter of YP in this study reported sometimes using their phones to access health information or services, suggesting that phones may be an acceptable medium for health information and services. A study of informal mHealth (mobile health) in South Africa, Ghana, and Malawi found that almost a fifth of YP surveyed reported using their phones in the previous year to obtain health advice or information [21]. Similarly, across 7 SSA countries an average of 17% of mobile phone owners reported having used their phone in the past 12 months to get information about health and medicine [32].

Current phone use behaviors and preferences can inform the kind of intervention that might be attractive to YP. Many of the interviewed YP use games and social media, and so health interventions that incorporate gaming and social aspects may be attractive. A qualitative study in the United States identified 3 reasons that adolescents 13-18 years reported using ICT for their health: to gather information, to share experiences and view others' experiences in order to gain social support or inspiration, and to track health behaviors and goals [33]. In this study participants reported using features that would correspond with each of these 3 reasons (eg, used web browsers, social media platforms, and health monitoring apps).

Older age groups could be targeted for phone-based interventions but some issues require consideration when planning mobile phone interventions with YP in this setting.

Access and Feasibility

Careful consideration needs to be given to equitable access to interventions, especially in terms of age and gender. Internet-based interventions that require access to a smartphone or computer may be less feasible and increase inequality. Cost, internet coverage, and speed may hinder intervention uptake [23,25]. However, as with access to mobile phones, these factors are likely to change over time. The functionality of phones and the prevalence of fake smartphones should be explored during formative work as not all smartphones may be capable of running additional apps. Recording whether participants have multiple phones or phone numbers may improve follow-up [34]. Phone sharing is relatively common and YP's access to phones may be controlled by someone else which may limit their access and raises the issue of confidentiality [22,35]. In this study females were most likely to share phones with their mothers and boyfriends/partners, and boys with their siblings. Females also reported primarily sharing phones with their mothers in other SSA countries [23]. Phone sharing behaviors coupled with low confidence in the security of information on phones may lead to poor uptake of interventions. This may be particularly important if phones are to be used to deliver information or interventions on sensitive topics such as sexual health. While the use of shared phones for sexual health interventions could be harmful for the participant, it could equally result in healthy discussions between the participant and phone owner.

Safety and Skills

Whether literacy and technical skills are a barrier to phone and internet use should be assessed, and if so then additional training and support provided [25]. Potential risks associated with phone-based interventions need to be considered when designing interventions and monitored closely during implementation. One potential risk of internet use is exposure to unwanted content [27]. Online risk can be categorized as exposure to "content," "contact" (where the YP participates, even unwillingly), or "conduct" (where the YP is the actor) [14]. In this study YP report exposure to content but they did not mention, nor did we specifically probe about other risks. Research suggests that YP are often resilient and have mechanisms to cope with these risks but that those who are vulnerable offline are often vulnerable online [14]. YP in this study reported low levels of phone security and confidence in the privacy of information. Safety and security concerns can be barriers to mobile phone use [25]. Training for YP should include security and confidentiality (eg, use of passwords), and the development of resilience to navigate risk in the online environment [28].

Conclusions

Mobile phone–based health interventions may be feasible for urban and peri-urban Zimbabwean YP. However, interventions could increase inequity especially if they require access to the internet. Provision of free internet access may remove this inequity, but additional factors such as capability to recharge one's mobile phone and the technological capabilities of phones should be taken into account. Involving YP from the target communities in intervention design teams is recommended to develop more appropriate and feasible interventions. Potential risks to intervention participants should be closely monitored and mitigated against by incorporating skill-building sessions on safe internet and phone use into recruitment activities.

Acknowledgments

Financial support for this study was provided through joint funding under the UK Medical Research Council (MRC)/UK Department for International Development (DFID) Concordat agreement which is supported by the European Union under the EDCTP2

programme (reference MR/K012126/1) with AD receiving support (G0700837). The CHIEDZA trial is funded by the Wellcome Trust through a fellowship for RF (206316Z/17/Z). We thank the field team who conducted the interviews, and other staff at Biomedical Research and Training Institute, Harare, who supported the conduct of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Young people's questionnaire.
[DOCX File , 66 KB - mhealth_v9i1e21244_app1.docx]

Multimedia Appendix 2
Characteristics of phones and phone sharing (among n=423 phone users, 396 phone owners, 70 phone sharers).
[DOCX File , 19 KB - mhealth_v9i1e21244_app2.docx]

Multimedia Appendix 3
Non-phone owners.
[DOCX File , 16 KB - mhealth_v9i1e21244_app3.docx]

Multimedia Appendix 4
First phone.
[DOCX File , 17 KB - mhealth_v9i1e21244_app4.docx]

Multimedia Appendix 5
Frequency of phone use and attitudes towards phone.
[DOCX File , 16 KB - mhealth_v9i1e21244_app5.docx]

Multimedia Appendix 6
Expenditure on phone.
[DOCX File , 16 KB - mhealth_v9i1e21244_app6.docx]

Multimedia Appendix 7
Frequency of use of different features (row %; among n=423 phone users).
[DOCX File , 15 KB - mhealth_v9i1e21244_app7.docx]

Multimedia Appendix 8
How often use the internet at different locations? (among n=468 who have ever used the internet).
[DOCX File , 13 KB - mhealth_v9i1e21244_app8.docx]

Multimedia Appendix 9
Household ownership of other technological devices (n=634).
[DOCX File , 15 KB - mhealth_v9i1e21244_app9.docx]

Multimedia Appendix 10
Attitudes to the internet (among n=465 who have ever used the interneta).
[DOCX File , 20 KB - mhealth_v9i1e21244_app10.docx]

References

1. Livingstone S, Nandi A, Banaji S, Stoilova M. Young Adolescents and Digital Media: Uses, Risks and Opportunities in Low- and Middle-Income Countries: A Rapid Evidence Review. London, UK: Gage; 2017. URL: http://eprints.lse.ac.uk/83753/ [accessed 2020-12-18]
2. United Nations General Assembly. Transforming Our World: The 2030 Agenda for Sustainable Development (A/RES/70/1). New York, NY: United Nations; 2015. URL: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E [accessed 2020-12-18]

3. Zimbabwe National Statistics Agency (ZIMSTAT). 2015 Zimbabwe Demographic and Health Survey 2015 (2015 ZDHS). Harare, Zimbabwe: Zimbabwe National Statistics Agency (ZIMSTAT); 2016 Nov. URL: https://dhsprogram.com/pubs/pdf/FR322/FR322.pdf [accessed 2020-12-18]
4. Zimbabwe National Statistics Agency (ZIMSTAT). Zimbabwe Population Census 2012. Harare, Zimbabwe: Population Census Office; 2012. URL: http://www.zimstat.co.zw/wp-content/uploads/publications/Population/population/census-2012-national-report.pdf [accessed 2020-12-18]
5. Ministry of Health and Child Care (MOHCC). Zimbabwe Population-Based HIV Impact Assessment (ZIMPHIA) 2015-2016: Final Report. Harare, Zimbabwe: Ministry of Health and Child Care (MOHCC); 2019. URL: https://phia.icap.columbia.edu/wp-content/uploads/2019/08/ZIMPHIA-Final-Report_integrated_Web-1.pdf [accessed 2020-12-18]
6. Ministry of Health and Child Care (MOHCC). Accelerated Strategic and Costed Operational Plan 2014-18: Voluntary Medical Male Circumcision. Harare, Zimbabwe: Ministry of Health and Child Care (MOHCC); 2015. URL: https://www.malecircumcision.org/resource/national-launch-accelerated-voluntary-medical-male-circumcision-operational-plan-2014-2018 [accessed 2020-12-18]
7. International Telecommunication Union. Manual for Measuring ICT Access and Use by Households and Individuals. 2014. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/manual2009/ITUManualHouseholds2020_E.pdf [accessed 2020-12-18]
8. GSM Association. A Toolkit for Researching Women's Internet Access and Use. London, UK: GSM Association; 2018. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/05/GSMA-Women-and-Internet-Research-Toolkit_WEB.pdf [accessed 2020-12-18]
9. MACRO. Study of Mobile Phone Usage Among the Teenagers and Youth In Mumbai. 2004. URL: https://www.itu.int/osg/spu/ni/futuremobile/SocialAspects/IndiaMacroMobileYouthStudy04.pdf [accessed 2020-12-18]
10. Kreutzer T. Online and Digital Media Usage on Cell Phones Among Low-Income Urban Youth in Cape Town in Centre for Film and Media Studies. Cape Town: University of Cape Town; 2009. URL: https://open.uct.ac.za/bitstream/handle/11427/8968/thesis_hum_2009_kreutzer_t.pdf?sequence=1&isAllowed=y [accessed 2020-12-18]
11. UNICEF. mHealth and Young People in South Africa. Geneva, Switzerland: UNICEF; 2017. URL: https://www.unicef.org/southafrica/media/236/file/mhealth-and-young-people-in-South-Africa.pdf [accessed 2020-12-18]
12. World Bank. Zimbabwe: Rural Population. Geneva, Switzerland: World Bank; 2018. URL: https://tradingeconomics.com/zimbabwe/rural-population-percent-of-total-population-wb-data.html [accessed 2020-12-18]
13. GMSA. Mobile Internet Connectivity 2019. Sub-Saharan Africa Factsheet. 2019. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/Mobile-Internet-Connectivity-SSA-Factsheet.pdf [accessed 2020-12-18]
14. UNICEF South Africa. Connected Dot Com. Young People's Navigation of Online Risks. 2013. URL: https://www.unicef.org/southafrica/media/1021/file/ZAF-connected-dot-com-2013.pdf [accessed 2020-12-18]
15. UNCTAD. SDG Pulse. 2019. URL: https://sdgpulse.unctad.org/ict-and-digital-gap/#Ref_Itu2018a [accessed 2020-12-18]
16. Caribou Digital. Digital Lives in Ghana, Kenya, and Uganda. Farnham, Surrey, UK: Caribou Digital Publishing; 2015. URL: https://www.cariboudigital.net/wp-content/uploads/2019/01/1474-Caribou-Digital-Digital-Lives-in-Ghana-Kenya-and-Uganda.pdf [accessed 2020-12-18]
17. Ofcom. Digital Day 2016. Results From the Children's Diary Study. 2016. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0017/94013/Childrens-Digital-Day-report-2016.pdf [accessed 2020-12-18]
18. The DHS Program STATcompiler (Zimbabwe DHS 2016). 2015. URL: http://www.statcompiler.com [accessed 2020-12-18]
19. Pew Research Center. Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa. Washington, D.C: Pew Research Center; 2018. URL: https://www.pewresearch.org/global/2018/10/09/internet-connectivity-seen-as-having-positive-impact-on-life-in-sub-saharan-africa/ [accessed 2020-12-18]
20. Porter G, Hampshire K, Milner J, Munthali A, Robson E, de Lannoy A, et al. Mobile Phones and Education in Sub-Saharan Africa: From Youth Practice to Public Policy. J. Int. Dev 2015 Jun 11;28(1):22-39. [doi: 10.1002/jid.3116]
21. Hampshire K, Porter G, Owusu SA, Mariwah S, Abane A, Robson E, et al. Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc Sci Med 2015 Oct;142:90-99 [FREE Full text] [doi: 10.1016/j.socscimed.2015.07.033] [Medline: 26298645]
22. Porter G, Hampshire K, Abane A, Munthali A, Robson E, Mashiri M, et al. Youth, mobility and mobile phones in Africa: findings from a three-country study. Information Technology for Development 2012 Feb 03;18(2):145-162. [doi: 10.1080/02681102.2011.643210]
23. Girl Effect: Real Girls, Real Lives, Connected: A Global Study of Access and Usage of Mobile, Told Through 3000 Voices. 2018. URL: https://prd-girleffect-corp.s3.amazonaws.com/documents/GE_VO_Full_Report-compressed.pdf?AWSAccessKeyId=AKIAIWVYO5R6RMTXA2NA&Signature=B4A%2BFxiZIZUSILB4F8NbvoQforo%3D&Expires=1608287592 [accessed 2020-12-18]
24. Wanyama JN, Nabaggala SM, Kiragga A, Owarwo NC, Seera M, Nakiyingi W, et al. High mobile phone ownership but low internet access and use among young adults attending an urban HIV clinic in Uganda. Vulnerable Children and Youth Studies 2018 Jan 03;13(3):207-220. [doi: 10.1080/17450128.2017.1418037]
25. GSM Association. Connected Women. The Mobile Gender Gap Report 2019. 2019. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf [accessed 2020-12-18]

26. UN Secretary-General's High-level Panel on Digital Cooperation. The Age of Digital Interdependence. 2019. URL: https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf [accessed 2020-12-18]
27. UNICEF. State of the World's Children 2017. Children in a Digital World. 2017. URL: https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf [accessed 2020-12-18]
28. UNICEF. Global Kids Online: Comparative Report. Florence, Italy: UNICEF Office of Research; 2019. URL: https://www.unicef-irc.org/publications/pdf/GKO%20LAYOUT%20MAIN%20REPORT.pdf [accessed 2020-12-18]
29. Feldacker C, Murenje V, Holeman I, Xaba S, Makunike-Chikwinya B, Korir M, et al. Reducing Provider Workload While Preserving Patient Safety. JAIDS Journal of Acquired Immune Deficiency Syndromes 2020;83(1):16-23. [doi: 10.1097/qai.0000000000002198]
30. Venables E, Ndlovu Z, Munyaradzi D, Martínez-Pérez G, Mbofana E, Nyika P, et al. Patient and health-care worker experiences of an HIV viral load intervention using SMS: A qualitative study. PLoS One 2019 Apr 11;14(4):e0215236 [FREE Full text] [doi: 10.1371/journal.pone.0215236] [Medline: 30973925]
31. Praekelt Foundation. Tune Me Zimbabwe. 2017. URL: https://www.praekelt.org/worktunemelearnmore [accessed 2020-05-21]
32. Pew Research Center. Cell Phones in Africa: Communication Lifeline. 2015. URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2015/04/Pew-Research-Center-Africa-Cell-Phone-Report-FINAL-April-15-2015.pdf [accessed 2020-12-18]
33. Radovic A, McCarty CA, Katzman K, Richardson LP. Adolescents' Perspectives on Using Technology for Health: Qualitative Study. JMIR Pediatr Parent 2018 Mar 14;1(1):e2 [FREE Full text] [doi: 10.2196/pediatrics.8677] [Medline: 30740590]
34. 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]
35. Ames H, Glenton C, Lewin S, Tamrat T, Akama E, Leon N. Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis. Cochrane Database Syst Rev 2019 Oct 14;10:CD013447 [FREE Full text] [doi: 10.1002/14651858.CD013447] [Medline: 31608981]

Abbreviations

ACASI: audio computer-assisted self-interviewing
ICT: information and communication technology
SSA: sub-Saharan Africa
YP: young people

Edited by G Eysenbach; submitted 09.06.20; peer-reviewed by G Khalil, S Owusu; comments to author 15.08.20; revised version received 08.10.20; accepted 30.10.20; published 13.01.21.

Please cite as:
Doyle AM, Bandason T, Dauya E, McHugh G, Grundy C, Dringus S, Dziva Chikwari C, Ferrand RA
Mobile Phone Access and Implications for Digital Health Interventions Among Adolescents and Young Adults in Zimbabwe: Cross-Sectional Survey
JMIR Mhealth Uhealth 2021;9(1):e21244
URL: https://mhealth.jmir.org/2021/1/e21244
doi:10.2196/21244
PMID:33439136

Original Paper

Usability and Preliminary Effectiveness of a Preoperative mHealth App for People Undergoing Major Surgery: Pilot Randomized Controlled Trial

Miriam van der Velde[1,2], PT, MSc; Karin Valkenet[1,2], PT, PhD; Edwin Geleijn[3], PT; Marjoke Kruisselbrink[4], PT, MSc; Marije Marsman[5], MSc, MD; Liedewij MJ Janssen[6], MSc, MD; Jelle P Ruurda[7], MD, PhD; Donald L van der Peet[8], MD, PhD; Jesse J Aarden[9], PT, MSc; Cindy Veenhof[1,2], PT, PhD; Marike van der Leeden[3,10], PT, PhD

[1]Innovation of Human Movement Care Research Group, HU University of Applied Sciences, Utrecht, Netherlands
[2]Department of Rehabilitation, Physical Therapy Science and Sports, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
[3]Department of Rehabilitation Medicine, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[4]Clinical Health Sciences, Program Physiotherapy Sciences, Utrecht University, Utrecht, Netherlands
[5]Department of Anesthesiology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
[6]Department of Anesthesiology, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[7]Department of Surgery, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
[8]Department of Surgery, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[9]European School of Physiotherapy, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[10]Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Miriam van der Velde, PT, MSc
Innovation of Human Movement Care Research Group
HU University of Applied Sciences
Heidelberglaan 7
Utrecht,
Netherlands
Phone: 31 064 221 4251
Email: miriam.vandervelde@hu.nl

Abstract

Background: Major surgery is associated with negative postoperative outcomes such as complications and delayed or poor recovery. Multimodal prehabilitation can help to reduce the negative effects of major surgery. Offering prehabilitation by means of mobile health (mHealth) could be an effective new approach.

Objective: The objectives of this pilot study were to (1) evaluate the usability of the Be Prepared mHealth app prototype for people undergoing major surgery, (2) explore whether the app was capable of bringing about a change in risk behaviors, and (3) estimate a preliminary effect of the app on functional recovery after major surgery.

Methods: A mixed-methods pilot randomized controlled trial was conducted in two Dutch academic hospitals. In total, 86 people undergoing major surgery participated. Participants in the intervention group received access to the Be Prepared app, a smartphone app using behavior change techniques to address risk behavior prior to surgery. Both groups received care as usual. Usability (System Usability Scale), change in risk behaviors 3 days prior to surgery, and functional recovery 30 days after discharge from hospital (Patient-Reported Outcomes Measurement Information System physical functioning 8-item short form) were assessed using online questionnaires. Quantitative data were analyzed using descriptive statistics, chi-square tests, and multivariable linear regression. Semistructured interviews about the usability of the app were conducted with 12 participants in the intervention group. Thematic analysis was used to analyze qualitative data.

Results: Seventy-nine people—40 in the intervention group and 39 in the control group—were available for further analysis. Participants had a median age of 61 (interquartile range 51.0-68.0) years. The System Usability Scale showed that patients considered the Be Prepared app to have acceptable usability (mean 68.2 [SD 18.4]). Interviews supported the usability of the app. The major point of improvement identified was further personalization of the app. Compared with the control group, the intervention group showed an increase in self-reported physical activity and muscle strengthening activities prior to surgery. Also, 2 of 2

frequent alcohol users in the intervention group versus 1 of 9 in the control group drank less alcohol in the run-up to surgery. No difference was found in change of smoking cessation. Between-group analysis showed no meaningful differences in functional recovery after correction for baseline values (β=–2.4 [95% CI –5.9 to 1.1]).

Conclusions: The Be Prepared app prototype shows potential in terms of usability and changing risk behavior prior to major surgery. No preliminary effect of the app on functional recovery was found. Points of improvement have been identified with which the app and future research can be optimized.

Trial Registration: Netherlands Trial Registry NL8623; https://www.trialregister.nl/trial/8623

(JMIR Mhealth Uhealth 2021;9(1):e23402) doi:10.2196/23402

KEYWORDS

preoperative care; smartphone; mhealth; risk behavior; prehabilitation; usability

Introduction

Every year, approximately 635,000 adults in the Netherlands undergo major surgery [1]. Major surgery is associated with negative postoperative outcomes such as complications and delayed or poor recovery [2]. Risk behaviors, such as smoking, excessive alcohol consumption, and physical inactivity, and risk factors like poor nutritional status may elevate the chance of poor postoperative outcomes in people undergoing major surgery [2-6].

Prehabilitation programs can be used to address these risks prior to major surgery. Prehabilitation is the process of improving an individual's functional capacity by modifying risk behaviors to enable them to withstand the forthcoming stressor of major surgery [7]. There is increasing evidence for the effectiveness of prehabilitation in reducing the negative effects of major surgery [3,4].

As risk behaviors rarely occur alone, prehabilitation programs will often need to address multiple factors. These programs often consist of a combination of physical training (unsupervised or supervised by a physical therapist), nutritional support by a dietitian, smoking and alcohol cessation support, and/or psychological support [4,8,9]. Changing risk behaviors before major surgery offers opportunities and challenges. The preoperative period is considered a teachable moment, a useful time to facilitate change [10]. Some research indicates that hospitalization or upcoming surgery increases motivation to change risk behaviors [11]. In addition, patients scheduled for elective surgery might be more willing to change their risk behavior preoperatively given the restricted period of behavior change.

Even though the preoperative period is considered a teachable moment, it is also a stressful period for many patients, which poses a challenge in terms of changing risk behavior [12,13]. Furthermore, patients experience various barriers to participate in prehabilitation (eg, problems with transportation, finding time, and bearing costs). Offering prehabilitation in one's own environment by means of mobile health (mHealth) could be an effective new approach, making prehabilitation easily accessible to many patients and helping to overcome experienced barriers to participation [14]. However, evidence for the use of mHealth apps for multimodal prehabilitation is lacking.

In this study, we evaluate the first version of the Be Prepared (Beter Voorbereid, in Dutch) mHealth app. The content of the app was developed by a team of health care professionals and health care researchers collaboratively with patients to optimize the process of prehabilitation and overcome barriers. The app uses behavior change techniques to address risk behaviors and enhance patients' health prior to surgery in order to achieve a better postoperative functional recovery [15].

The Centre for eHealth Research (CeHRes) roadmap, a 5-step development, evaluation, and implementation approach, was used as a guideline during the development and evaluation of the Be Prepared app [16]. In this pilot study, we describe the first steps in the evaluation.

This evaluation focuses on the usability of the app and the first effects on functional outcomes. Usability has been identified as an essential criterion for the evaluation of digital apps in health care since poor usability can influence the use of the app and thereby affect its effectiveness [17].

Therefore, the primary aim of this pilot study is to evaluate the usability of the Be Prepared app prototype. In addition, we explore whether the app is capable of bringing about a change in risk behaviors in people undergoing major surgery, and we estimate a preliminary effect of the Be Prepared app on functional recovery after major surgery.

Methods

Design

This multicenter pilot randomized controlled trial (RCT) is part of the development approach following the CeHRes roadmap [16]. The first stages of the CeHRes roadmap (contextual inquiry, value specification, and design) and this pilot study provide the basis for the subsequent steps (operationalization and summative evaluation). These steps will be conducted in the next phase of this project. This pilot study was conducted to evaluate the usability and preliminary effectiveness of the Be Prepared app and feasibility of study procedures to identify points of improvement before conducting a large multicenter RCT. In this article, we focus on the evaluation of usability and preliminary effectiveness of the app.

Patients scheduled for major elective surgery were recruited from the preoperative assessment outpatient clinic of two academic hospitals in the Netherlands between November 2018 and March 2019. Patients were informed about the study by the

anesthetist, anesthesiology nurse, or anesthesiology assistant during their preoperative assessment. Patients who were interested in participation received written information about the study and were called the next day by the investigator (at least 24 hours after being informed about the study). During the phone call, the patient was asked whether they wanted to participate, and inclusion criteria were checked. Eligible patients signed the informed consent form provided with the written information and sent it back to the investigators. After completing the baseline questionnaires, participants were randomly assigned to either the intervention or control group using a web-based randomization system. The intervention group received a link to the Be Prepared app via email. Both groups received care as usual (eg, verbal information and information leaflets). Quantitative data were collected at baseline, 3 days before surgery, and 30 days after hospital discharge. Patient characteristics and perioperative factors were collected via the electronic health record. All other data were collected through questionnaires completed by the participant on a secured web-based system. Qualitative data were collected pre- or postoperatively from a selection of app users. Participants could indicate on the informed consent form whether they gave permission to be approached for a telephone interview.

The medical ethical committee of the Amsterdam University Medical Center approved this study (NL61503.029.18). No changes were made to the design after the study was approved. We followed the Consolidated Standards of Reporting Trials (CONSORT) guidelines and the CONSORT EHEALTH (Electronic and Mobile Health Applications and Online Telehealth) checklist [18,19]. The main study is registered at the Netherlands Trial Registry [NL8623], but this pilot study is not.

Participants

Patients scheduled for major elective surgery were eligible to be included if they were aged 18 years or older, had an indication for postoperative hospital stay of at least 2 nights, and had one or more risk behaviors. For the purpose of inclusion, we categorized risk behaviors into binary variables representing risk status based on the recommendations of the Dutch Health Council and evidence (ie, currently smoking, alcohol consumption 1 or more drinks every day, moderate intensity physical activity less than 30 minutes every day, muscle strengthening activities on fewer than 2 days per week, and/or unintentional weight loss of more than 3kg in the last month) [2,20-22]. Participants were excluded if they had no access to a mobile device or had an insufficient command of the Dutch language. Patient characteristics (age, gender, BMI, physical functioning, type of surgery, American Society of Anesthesiologists physical status classification, presence of risk behaviors) were collected in order to describe the sample.

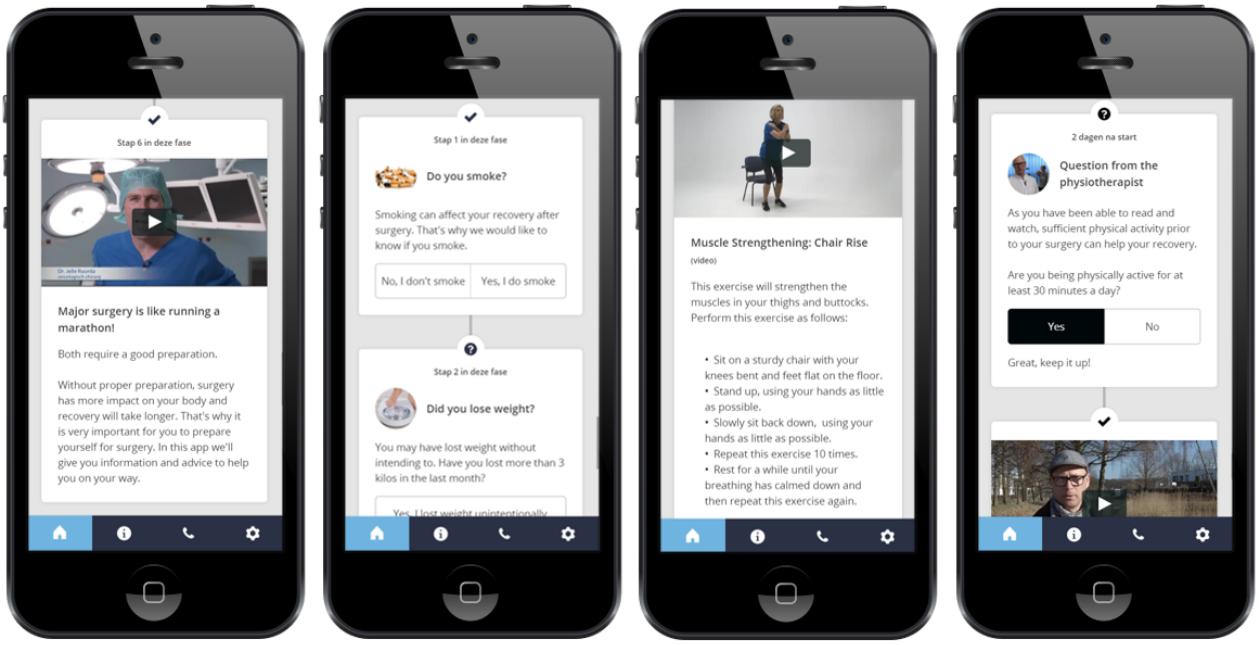
Intervention

The Be Prepared app prototype (Patient Journey platform by Interactive Studios BV) is a smartphone app which uses behavior change techniques to support patients in optimizing their health and risk behaviors prior to surgery. A behavior change technique is a strategy that helps an individual change their behavior to promote better health. Techniques like setting goals, advice on stop smoking medication, social support, feedback on behavior, and providing information on the health consequences of alcohol consumption and alcohol cessation were used in this app [15,23]. For example, participants were encouraged to exercise with a buddy and call someone when they felt the urge to smoke and were informed by a pulmonologist about the health risks of smoking before a major operation. In the app, participants answered questions about their risk behavior and received tailored information and advice based on the given answers. Current smokers were supported with smoking cessation prior to surgery. Frequent alcohol users were supported to decrease their alcohol intake. Inactive participants were supported to increase their amount of physical activity to at least 30 minutes of moderate intensity physical activity per day. Participants who did muscle strengthening activities less than twice a week were supported to increase these activities to at least twice a week in combination with increasing protein-rich food in their diet. Participants who unintentionally lost more than 3 kg during the past month were advised on protein and energy enriched food. Additionally, participants received information about preoperative fasting and the use of blood coagulation medication prior to surgery.

Participants in the intervention group received access to the Be Prepared app for use on their own mobile device. The introduction screen showed only basic information about the goal and use of the app. They could unlock the additional information and advice by entering a personal code they received via email.

The information and advice in the app were displayed on a dynamic timeline based on the patient's operation date. Through this dynamic timeline, day-to-day information was offered in various ways to meet different needs. The timeline provided written information and videos, tips on healthy behavior and changing risk behaviors, quizzes, and exercise videos (Figure 1). Furthermore, participants were asked whether they succeeded in following the advice and received feedback based on their response. Push notifications informed the patient about available new information and advice. This prototype version of the app provided information and advice for a maximum of 14 days prior to surgery. An overview of the app content is provided in Multimedia Appendix 1.

Figure 1. Screenshots of the Be Prepared app, translated from Dutch. From left to right, information about the purpose of the app, screening questions for app personalization, muscle strengthening exercise video, and check for progress with feedback video.



Data Collection

Usability

Usability of the mHealth intervention was assessed both quantitatively and qualitatively. Usability is the extent to which a product can be used by specified users to achieve specified goals with regard to effectiveness, efficiency, and satisfaction [24].

App Use

Whether a participant in the intervention group activated the app with their personal code was logged anonymously in the database of Interactive Studios (ISO 27001 and NEN7510 certified). In the online questionnaire, participants in the intervention group were asked whether they had used the app and, if applicable, were asked about reasons for nonuse.

Quantitative Data

The Dutch translation of the System Usability Scale (SUS) was used to assess usability among all app users. Participants completed the SUS 3 days prior to surgery. The SUS is a reliable and valid 10-statement usability scale suitable to assess a wide range of eHealth technologies. The total SUS score ranges from 0 to 100, and higher scores reflect higher usability [25,26]. An SUS score of at least 62.7 was considered acceptable [25,27], and 68 or above was regarded as above average in terms of usability quality.

Qualitative Data

For the qualitative data collection, semistructured telephone interviews were conducted with a selection of participants to gain more detailed insight into the usability of the app. Interviews took 20 to 30 minutes. During the interviews, a topic list was used based on the usability components described by Nielsen [28]: efficiency, satisfaction, learnability, memorability, and tolerance for errors. The topic list is provided in Multimedia Appendix 2. Preoperative interviews were conducted 3 or more days before surgery and postoperative interviews between 1 and 2 weeks after hospital discharge. Participants were included until maximum variation in patient characteristics (age, gender, days of app use, interviewing pre- or postoperatively, and length of hospital stay) and data saturation in 3 consecutive interviews was reached.

Risk Behaviors

Risk behaviors were assessed through self-report at baseline and 3 days before surgery. Participants were asked to indicate on how many days per week they were physically active for at least 30 minutes, on how many days per week they performed muscle strengthening activities, and on how many days per week they consumed one or more alcoholic beverages. Participants were asked whether they were currently smoking. Current smokers were asked to indicate how many cigarettes they had smoked last week.

Participants were also asked to self-report their change in risk behaviors using a set of closed-ended questions at 3 days before surgery. All participants were asked whether they had made positive changes to their risk behaviors regarding smoking, alcohol intake, unintentional weight loss, physical activities, and/or muscle strengthening activities prior to surgery.

Functional Recovery

Functional recovery after surgery was assessed by the Patient-Reported Outcomes Measurement Information System physical functioning 8-item short form (PROMIS-PF) at baseline and 30 days after hospital discharge. The PROMIS-PF is derived from the Dutch PROMIS physical function item bank consisting of 121 items concerning daily activities. The short form consists of 8 questions which can be scored on a 5-point Likert scale

from 1 (unable to do) to 5 (performed without any difficulty). The PROMIS-PF has a high reliability and validity and is applicable to patients with different conditions [29,30].

Data Analysis

Descriptive statistics were used to describe patient characteristics and app use. No a priori level of statistical significance was set as this pilot study was not powered to assess effect. Point estimators and confidence intervals are given to estimate the effects, and *P* values are provided to give an impression of the evidence against the null hypothesis [31]. Complete case analyses were performed according to the intention-to-treat principle. Quantitative data were analyzed using SPSS Statistics version 25.0 (IBM Corporation).

Usability

Quantitative Data

The total SUS score was calculated using the method by Brooke [25], and descriptive statistics were used to describe scores. In order to have a better insight into the different aspects of usability, the SUS statements were subdivided into the categories learnability, efficiency, and satisfaction [32]. Mean scores per SUS statement and per category of usability were calculated using the method by Brooke [25] with the addition of a factor 10 to get a range of 0 to 100 per statement.

Qualitative Data

Qualitative analysis of interview data was done following the steps of thematic analysis: compiling, disassembling, reassembling, interpreting, and concluding [33-35]. Interviews were transcribed verbatim, and data were coded through open coding. Axial coding was discussed by two researchers to define definitive codes. Subsequently, codes were put into context with each other to create themes. Differences were discussed until consensus was reached. Next, analytical conclusions were made from the data presented as codes and themes. The main themes and findings regarding usability are described in the Results section as an addition to the quantitative analysis of usability.

Risk Behaviors

Descriptive statistics were used to describe changes in risk behaviors. Bootstrapping methods (1000 samples) were used to calculate confidence intervals for medians. Chi-square tests for linear trend were performed to examine the relation between group allocation and change scores in days of performing physical activities and muscle strengthening activities. Chi-square tests or Fisher exact tests were used to test the difference in distribution of self-reported change of all risk behaviors between allocation groups.

Functional Recovery

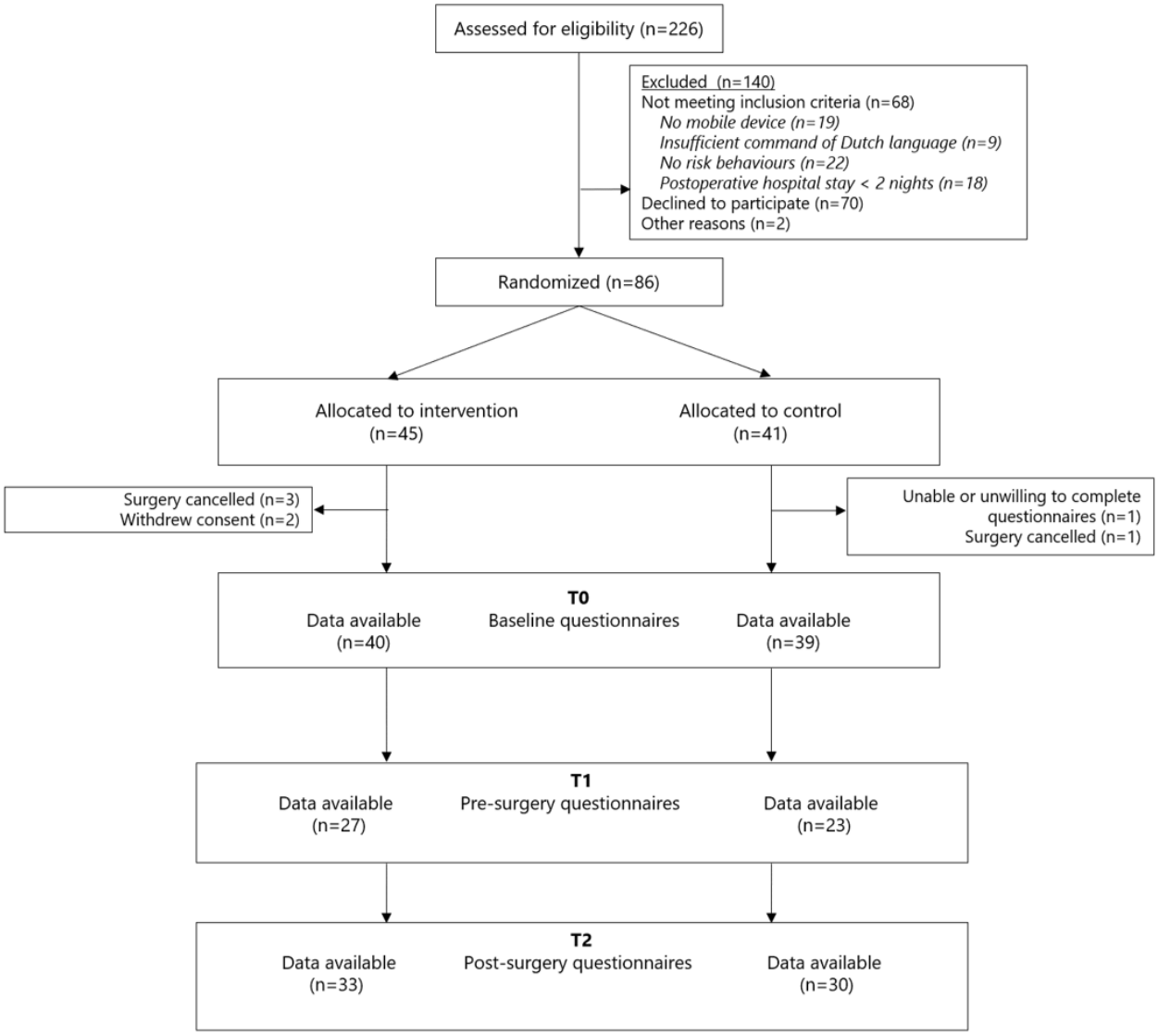
Total raw scores on the PROMIS-PF were translated into a t score for each participant using the PROMIS score conversion table [36]. The t score rescales total raw score into a standardized score with a mean of 50 and a standard deviation of 10. Descriptive statistics were used to describe physical functioning at baseline and 30 days after discharge from hospital. Bootstrapping methods were used to estimate confidence intervals for medians. A PROMIS-PF score at 30 days after discharge from hospital greater than or equal to the score at baseline was considered functionally recovered. Between-group differences in functional recovery measured by the postsurgery PROMIS-PF corrected for baseline PROMIS-PF were analyzed using multivariable linear regression.

Results

Flow of Participants Through the Study

Between November 2018 and February 2019, 226 patients were screened for eligibility; 86 people were eligible and signed informed consent, with 45 participants randomized to the experimental group and 41 to the control group. In the control group, one participant did not complete baseline questionnaires and was therefore excluded from further analysis, and 4 participants were excluded from further analysis because their surgeries were cancelled. In the intervention group, 2 participants withdrew informed consent due to nursing home admission and start of palliative care. Thus, there were 40 evaluable participants in the intervention group and 39 in the control group (Figure 2).

Figure 2. Design of study and flow of participants through the trial.



The median age of participants was 61.0 (interquartile range [IQR] 51.0-68.0) years, 49% (39/79) were female; 34% (27/79) of participants had 1 risk behavior at baseline, 48% (38/79) had 2 risk behaviors, and 18% (14/79) had 3 or more. Of the participants, 81% (64/79) were insufficiently physically active, and the median waiting time for surgery was 28 (IQR 16-52) days. The groups were similar at baseline in terms of demographic and clinical characteristics (Table 1).

Table 1. Baseline characteristics of participants.

Characteristic	Total (n=79)	Intervention (n=40)	Control (n=39)
Age in years, median (IQR)	61.0 (51.0-68.0)	59.0 (43.8-64.0)	63.0 (53.0-70.0)
Female, n (%)	39 (49)	22 (55)	17 (44)
BMI (kg/m^2), median (IQR)	25.8 (23.9-28.3)	25.5 (23.1-27.7)	26.5 (24.5-28.8)
ASA PS[a] classification, n (%)			
I	9 (12)	7 (17)	2 (5)
II	42 (53)	21 (53)	21 (54)
III	24 (30)	11 (28)	13 (33)
IV	3 (4)	1 (2)	2 (5)
Unknown	1 (1)	0 (0)	1 (3)
Surgical specialty, n (%)			
Neurosurgical	16 (20)	9 (23)	7 (18)
Cardiothoracic	16 (20)	7 (18)	9 (23)
Gastrointestinal	15 (19)	8 (18)	7 (18)
Oral and maxillofacial	12 (15)	4 (10)	8 (20)
Urologic and gynecologic	13 (17)	8 (20)	5 (13)
Orthopedic	3 (4)	0 (0)	3 (8)
Vascular	3 (4)	3 (8)	0 (0)
Other	1 (1)	1 (3)	0 (0)
Waiting time for surgery in days, median (IQR)	28 (16-52)	27 (16-46)	29 (16-65)
Risk behaviors[b], n (%)			
Smoking	7 (9)	4 (10)	3 (8)
Alcohol consumption	12 (15)	2 (5)	10 (26)
Physical activities	64 (81)	35 (88)	29 (74)
Muscle strengthening activities	57 (72)	29 (73)	28 (72)
Unintentional weight loss	7 (9)	3 (8)	4 (10)
Number of risk behaviors, n (%)			
1	27 (34)	12 (30)	15 (39)
2	38 (48)	23 (58)	15 (39)
3	12 (15)	5 (12)	7 (17)
4	2 (2)	0 (0)	2 (5)
PROMIS-PF[c] (t score), median (IQR)	47.8 (40.8-60.1)	47.8 (42.3-60.1)	46.7 (40.1-60.1)

[a]ASA PS: American Society of Anesthesiologists physical status.

[b]Multiple response options.

[c]PROMIS-PF: Patient-Reported Outcomes Measurement Information System physical functioning 8-item short form.

Usability

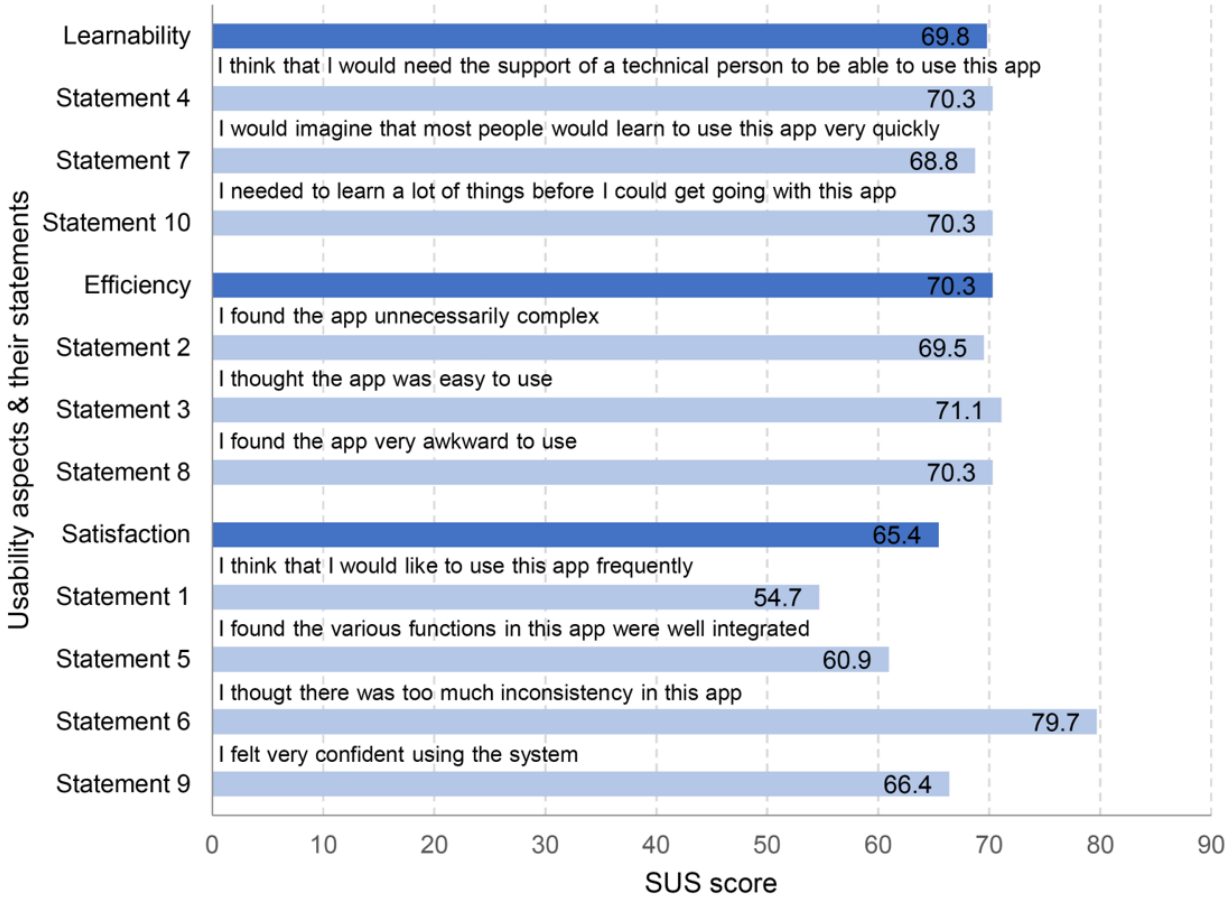
App Use

Of the participants in the intervention group, 73% (29/40) activated the app with their personal code. Reasons for nonuse were unable to access the app (2/29) and not seeing added value of app use (1/29). For the other participants, reasons for nonuse are unknown. There were no apparent differences in characteristics between app users and nonusers.

Quantitative Results

The SUS was completed by 80% (32/40) of participants from the intervention group. The usability of the Be Prepared app scored 68.2 (SD 18.4). The mean SUS scores per usability aspect were: learnability 69.8 (SD 22.5), efficiency 70.3 (SD 23.5), and satisfaction 65.4 (SD 21.9). Figure 3 shows the SUS scores per statement and usability aspect.

Figure 3. System Usability Scale scores per statement and usability aspect (higher scores reflect higher usability).



Qualitative Results

Interviewees

Of the participants in the intervention group, 33% (13/40) were approached for a telephone interview. One participant declined to participate because the interview would take too much time. After 12 interviews, data saturation was reached. Table 2 presents the characteristics of the interviewees. After axial coding, 5 themes were formed. The qualitative results will be described per theme.

Table 2. Characteristics of interviewees.

Code	Gender	Age (years)	App use (number of days)	Interview pre- or postoperative	Days of hospitalization
1	F	35	15	Pre	4
2	M	65	3	Pre	2
3	M	76	4	Pre	3
4	M	68	3	Pre	5
5	F	63	17	Post	4
6	M	49	42	Pre	Waitlisted
7	F	43	34	Post	3
8	F	59	16	Post	7
9	F	63	24	Post	7
10	M	36	22	Post	1
11	F	52	38	Pre	6
12	M	77	84	Post	7

Ease of Use of the App

The log-in procedure was difficult for many interviewees, and some needed help from family or the research team to log in.

> *I had some trouble opening the app, but then I went back to the instructions and it clearly stated what I had to do.* [Male, age 65]

The interviewees did not experience problems when using the app.

> *It is self-explanatory.* [Male, age 76]

Contents of the App Should Meet Personal Preferences

The interviewees were satisfied with the various ways in which information was presented in the app and appreciated the practical advice.

> *Suggestions are being made in the app that I found useful. For me it was combining walking with doing groceries. I thought that was a clever one, for me as well; if you go somewhere anyway, go walking. I even did that yesterday.* [Male, age 65]

The app did not match everyone's level of functioning, and interviewees saw the need for more personalization in the app.

> *There are, of course, very fit people and people who have been living unhealthy for a long time. So, it will be difficult to differentiate. There could be an option menu of some sort in the app.* [Male, age 65]

App as Motivational Tool for Behavior Change

For the majority of the interviewees, getting a reminder to exercise more or eat more protein-rich food was enough to change their behavior while some others preferred advice from a health care professional.

> *An app is useful because you read it. Only knowing something, does not mean you will do it, but when you read it in an app you are reminded that you have to do it.* [Female, age 52]

The app emphasized the importance of changing risk behavior before major surgery. This motivated many interviewees. The push notifications supported most interviewees in their behavior change.

> *Some days I forgot or I was busy, so on those days the notifications came in handy.* [Female, age 35]

General Motivation for Behavior Change Before Major Surgery

The urgency to prepare for the upcoming surgery seemed to be the greatest motivation for the interviewees. Interviewees with a waiting time of a few weeks until their surgery felt the need to change their risk behavior.

> *I feel the need, I'm going into surgery next week. It has to stop raining, because this afternoon I have to work in my garden for at least 30 minutes as a physical activity.* [Male, age 65]

Interviewees who had to wait more than 4 weeks for surgery or who didn't know their surgery date did not feel the urge to change until a few weeks before surgery.

> *The longer you have to use something like that, the greater the chance that you will not finish it.* [Male, age 65]

Views on What Constitutes a Good Preparation for Major Surgery

According to the majority of interviewees, a good preparation should benefit postoperative recovery.

> *So you can manage better when you get back home.* [Female, age 59]

Being well-informed was also mentioned as a key factor for good preparation.

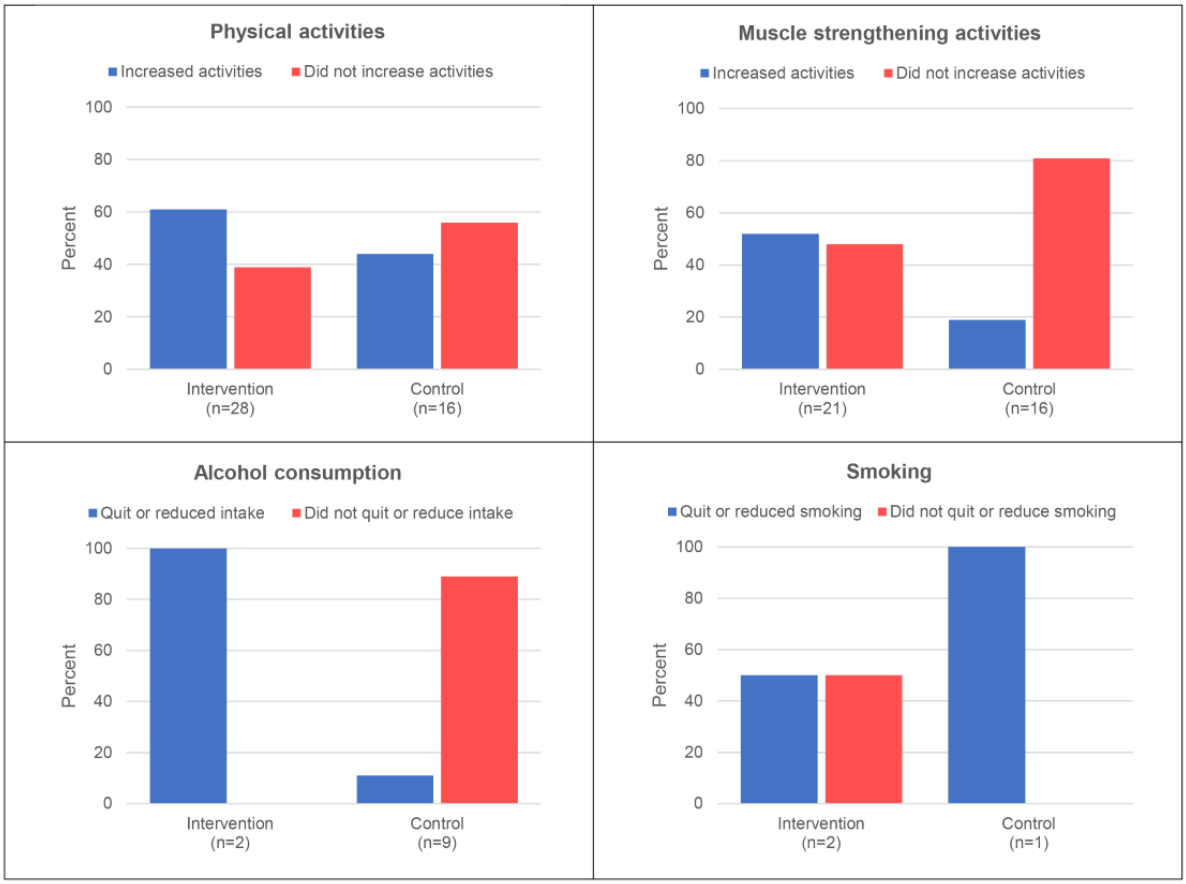
> *To know what you can expect, what you can and cannot do. That is important, because of course you don't know, you don't know what's wise to do.* [Male, age 68]

Risk Behaviors

Physical Activities

Of the participants, 81% (64/79) were physically inactive at baseline, of whom 69% (44/64) completed the presurgery follow-up questionnaire. At baseline, the median number of days on which participants were physically active for at least 30 minutes was 3 (95% CI 2.5 to 4.0) in the intervention group (n=28) and 4 (95% CI 2.0 to 5.0) in the control group (n=16). The intervention group became active on more days of the week after the intervention period (+1.0 day [95% CI 0.0 to +2.0]) compared with the control group (0.0 days [95% CI –0.5 to +1.0], *P*=.12). Figure 4 shows that a bigger proportion of participants in the intervention group reported having increased their physical activities prior to surgery (17/28, 61%) compared with the control group (7/16, 44%; *P*=.28).

Figure 4. Self-reported change in risk behavior.



Muscle Strengthening Activities

Of the participants, 72% (57/79) had risk behavior regarding muscle strengthening activities, of whom 65% (37/57) completed the presurgery follow-up questionnaire. At baseline, the median number of days on which participants performed muscle strengthening activities was 0 (95% CI 0.0 to 0.0) in both the intervention (n=21) and control group (n=16). The median number of days of muscle strengthening activities increased by +2.0 days (95% CI 0.0 to +3.0) in the intervention group and did not increase in the control group (0.0 [95% CI 0.0 to +1.0], *P*=.06). In the intervention group, 52% (11/21) of participants reported having increased their muscle strengthening activities before surgery compared with 19% (3/16) of participants in the control group (*P*=.04; Figure 4).

Alcohol Consumption

Of the participants, 15% (12/79) had risk behavior regarding alcohol consumption, and 92% (11/12) of those completed the presurgery questionnaire (intervention group [2/11], control group [9/11]). Both participants in the intervention group (100%) reported a reduction in alcohol consumption before surgery compared with 11% (1/9) in the control group (*P*=.06; Figure 4).

Smoking

Of the participants 9% (7/79) were current smokers at baseline, and 43% (3/7) completed the follow-up questionnaire (intervention group [2/3] and control group [1/3]). In each group, 1 participant indicated having stopped smoking or smoked less in the run-up to surgery (Figure 4). Due to the small groups, no test was performed to compare groups.

Unintentional Weight Loss

Of the participants, 9% (7/79) had unintentionally lost 3 kg of weight in the last month. Due to an error in the online questionnaire, this outcome could not be evaluated.

Functional Recovery

Of the participants, 81% (64/79) completed the questionnaire on physical functioning at baseline and 30 days after hospital discharge. At baseline, the median PROMIS-PF score was 47.8 (95% CI 43.9 to 60.1) in the intervention group (n=33) and 50.8 (95% CI 41.6 to 60.1) in the control group (n=31). Compared with baseline, both the intervention group (–6.2 [95% CI –10.6 to –2.1]) and the control group (–3.6 [95% CI –7.2 to 0.0]) had a lower level of physical functioning 30 days after discharge from hospital. Of the participants, 27% (9/40) in the intervention group and 35% (11/39) in the control group could be considered functionally recovered at 30 days after discharge from hospital. Using the complete cases, between-group analysis showed no meaningful difference in functional recovery after correction for baseline values (β=–2.4 [95% CI −5.9 to 1.1]).

Discussion

Principal Findings

The primary aim of this pilot study was to investigate the usability of the Be Prepared app prototype in patients undergoing major surgery. Both quantitative and qualitative data support the usability of the app and provide insight into adjustments that can be made to improve the app. Data from the SUS showed that patients considered the Be Prepared app to have acceptable usability (mean 68.2 [SD 18.4]) [25,27]. The app scored the best on the aspect of efficiency, the speed and ease with which the user gets something done [28,37]. Appreciation of the efficiency of the app is supported by our qualitative findings, as the majority of the interviewees found the app easy to use. The SUS score for learnability, the ease with which users accomplish basic tasks in the app for the first time [28], was above average. This was also supported by positive responses from the interviewees who found the app self-explanatory. Only the log-in procedure was described as difficult. The usability aspect of satisfaction had the highest and lowest SUS scores but could still be considered acceptable. The lowest scoring statement was "I think that I would like to use this app frequently," which, in the context of the use of this app prior to surgery, is potentially confusing for participants. The major point of improvement suggested by the interviewees was further personalization of the app.

Our results suggest the Be Prepared app to be capable of bringing about a change in risk behaviors prior to major surgery. Participants who used the app became physically more active and increased their number of days of performing muscle strengthening activities. Furthermore, a larger proportion of app users indicated that they drank less alcohol in the run-up to their operation. No difference was found in smoking behavior between groups, but the sample size of both groups was too small to draw any conclusions. In our study, risk behavior has been assessed by means of online questionnaires. Although this applies to both groups, it may lead to socially desirable answers and perhaps an underestimation of risk behavior at baseline or an overestimation of the positive change in risk behavior. In spite of that, our findings are consistent with other research suggesting that individuals awaiting surgery welcome support to increase physical activities and reduce alcohol consumption but are less positive about smoking cessation [2].

Despite the positive trend in our intermediate outcome, change of risk behaviors, no preliminary effect of the Be Prepared app prototype on functional recovery 30 days after major surgery was found. The fact that we have not been able to show an effect on functional recovery after surgery can have several causes. Based on the concept that improving an individual's functional capacity before major surgery helps them withstand the forthcoming stressor of major surgery, it makes sense to target those at high risk of postoperative complications and functional decline after surgery [4]. A recent systematic review describes that in many prehabilitation trials, inadequate patient selection may have led to an underestimation of the benefits of prehabilitation in patients undergoing major intra-abdominal surgery [38]. This is supported by the finding that only those studies targeting high-risk patients found significant improvements in physiological parameters and postoperative outcomes. In this pilot study, we included patients undergoing major surgery with an indication for postoperative hospital stay of a minimum of 2 nights and at least 1 risk behavior, with the aim to exclude low-risk patients. Nevertheless, participants in our pilot were on average younger and relatively fit in comparison with patients in other successful prehabilitation trials [38,39].

Furthermore, the Be Prepared app was a home-based unsupervised intervention. Evidence suggests that supervised prehabilitation may have a greater effect and higher adherence than unsupervised prehabilitation programs [4,38]. Offering prehabilitation by means of mHealth might not be sufficient for the entire group of surgical patients. Literature shows that especially patients with limited (digital) health literacy, advanced age, and chronic health conditions could benefit from extra support (blended care) in addition to mHealth [40-42]. Additional support by a health care provider during the preoperative period should therefore be considered for those patients in need of extra supervision.

Limitations

This study is one of the first studies to explore the usability and preliminary effectiveness of an app for multimodal prehabilitation in patients undergoing major surgery. Our objective was to determine whether the Be Prepared app was usable and beneficial for a wide range of surgical patients. Therefore, our study sample was deliberately more heterogeneous than samples in other prehabilitation trials [38,39,43,44]. In hindsight, the 30-day follow-up may not have been the ideal time to identify an improvement in functional recovery for this diverse group of surgical patients, as the course of recovery is highly dependent on, among other things, the type of surgery [45]. Our results show that the majority of patients have not yet functionally recovered within 30 days of discharge from hospital. Future evaluation of the Be Prepared app will therefore include a longer follow-up with multiple time points to be able to compare (the course of) functional recovery between groups.

This study suffered from a large amount of missing data, especially at the presurgery follow-up. This can be explained, in part, by the timing of this measurement. Participants were invited to complete the presurgery follow-up questionnaire 3 days preoperatively, which might not have been an ideal time as patients have their minds on other things right before their surgery [2]. Changing the timing of this measurement to 1 week after discharge from hospital may help to reduce the amount of missing data when conducting the definitive trial.

In this study, 73% of the possible app users activated the app. The difficult log-in procedure and differences in patient (digital) health literacy may have contributed to the substantial number of nonusers in the intervention group [46]. The proportion of patients accessing the app at least once was comparable to that of other apps [47]. Altering the log-in procedure and providing support during installation and first use of the app could increase initial app use, but it is well known that user engagement decreases over time [48]. This is consistent with our qualitative

data showing that patients who had access to the app for more than 4 weeks before surgery did not use the app during the entire preoperative period. Besides patients not feeling the urge to change until shortly before surgery, the limited days of preoperative content in the Be Prepared prototype (14 days) could have contributed to these patients stopping use of the app. Expanding the content for those with a longer preoperative period could help increase user engagement prior to surgery.

Conclusions

Overall, the results of this pilot RCT demonstrate that the Be Prepared app prototype for patients undergoing major surgery has potential in terms of usability and changing risk behavior prior to major surgery. The app seems to fit the needs of patients preparing for major surgery. Several points of improvement for the app and study procedures have been identified, which supports the further development of the Be Prepared app and adjustment of study procedures before evaluating its effectiveness in a large multicenter RCT. These include adaptation of the timing of follow-up measurements, additional support by the physiotherapist during the preoperative period, and expanding the preoperative content of the Be Prepared app.

Acknowledgments

This work was supported by the Foundation Innovative Alliance—Regional Attention and Action for Knowledge circulation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of app content.
[PDF File (Adobe PDF File), 102 KB - mhealth_v9i1e23402_app1.pdf]

Multimedia Appendix 2
Topic list for semistructured interviews.
[PDF File (Adobe PDF File), 10 KB - mhealth_v9i1e23402_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 564 KB - mhealth_v9i1e23402_app3.pdf]

References

1. StatLine—Operaties in het ziekenhuis, 1995-2010. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80386NED/table?fromstatweb [accessed 2020-06-03]
2. McDonald S, Yates D, Durrand JW, Kothmann E, Sniehotta FF, Habgood A, et al. Exploring patient attitudes to behaviour change before surgery to reduce peri-operative risk: preferences for short- vs. long-term behaviour change. Anaesthesia 2019 Dec;74(12):1580-1588 [FREE Full text] [doi: 10.1111/anae.14826] [Medline: 31637700]
3. Tew GA, Ayyash R, Durrand J, Danjoux GR. Clinical guideline and recommendations on pre-operative exercise training in patients awaiting major non-cardiac surgery. Anaesthesia 2018 Jun;73(6):750-768 [FREE Full text] [doi: 10.1111/anae.14177] [Medline: 29330843]
4. Scheede-Bergdahl C, Minnella EM, Carli F. Multi-modal prehabilitation: addressing the why, when, what, how, who and where next? Anaesthesia 2019 Jan;74 Suppl 1:20-26 [FREE Full text] [doi: 10.1111/anae.14505] [Medline: 30604416]
5. Thomsen T, Villebro N, Møller AM. Interventions for preoperative smoking cessation. Cochrane Database Syst Rev 2014 Mar 27(3):CD002294. [doi: 10.1002/14651858.CD002294.pub4] [Medline: 24671929]
6. Dronkers JJ, Chorus AMJ, van Meeteren NLU, Hopman-Rock M. The association of pre-operative physical fitness and physical activity with outcome after scheduled major abdominal surgery. Anaesthesia 2013 Jan;68(1):67-73 [FREE Full text] [doi: 10.1111/anae.12066] [Medline: 23121372]
7. Banugo P, Amoako D. Prehabilitation. BJA Education 2017 Dec;17(12):401-405. [doi: 10.1093/bjaed/mkx032]
8. Cabilan CJ, Hines S, Munday J. The effectiveness of prehabilitation or preoperative exercise for surgical patients: a systematic review. JBI Database System Rev Implement Rep 2015 Jan;13(1):146-187. [doi: 10.11124/jbisrir-2015-1885] [Medline: 26447015]
9. Li C, Carli F, Lee L, Charlebois P, Stein B, Liberman AS, et al. Impact of a trimodal prehabilitation program on functional recovery after colorectal cancer surgery: a pilot study. Surg Endosc 2013 Apr;27(4):1072-1082. [doi: 10.1007/s00464-012-2560-5] [Medline: 23052535]
10. Durrand J, Singh SJ, Danjoux G. Prehabilitation. Clin Med (Lond) 2019 Nov;19(6):458-464 [FREE Full text] [doi: 10.7861/clinmed.2019-0257] [Medline: 31732585]

11. Lawson PJ, Flocke SA. Teachable moments for health behavior change: a concept analysis. Patient Educ Couns 2009 Jul;76(1):25-30 [FREE Full text] [doi: 10.1016/j.pec.2008.11.002] [Medline: 19110395]
12. Ferreira V, Agnihotram RV, Bergdahl A, van Rooijen SJ, Awasthi R, Carli F, et al. Maximizing patient adherence to prehabilitation: what do the patients say? Support Care Cancer 2018 Aug;26(8):2717-2723. [doi: 10.1007/s00520-018-4109-1] [Medline: 29478189]
13. Carli F, Scheede-Bergdahl C. Prehabilitation to enhance perioperative care. Anesthesiol Clin 2015 Mar;33(1):17-33. [doi: 10.1016/j.anclin.2014.11.002] [Medline: 25701926]
14. van der Meij E. Improving Postoperative Recovery by eHealth [Thesis]. Amsterdam: Vrije Universiteit Amsterdam; 2019.
15. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512. [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
16. van Gemert-Pijnen EWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
17. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Dec;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
18. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
19. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [Medline: 20332509]
20. Neelemaat F, Kruizenga HM, de Vet HCW, Seidell JC, Butterman M, van Bokhorst-de van der Schueren MAE. Screening malnutrition in hospital outpatients. Can the SNAQ malnutrition screening tool also be applied to this population? Clin Nutr 2008 Jun;27(3):439-446. [doi: 10.1016/j.clnu.2008.02.002] [Medline: 18395946]
21. Dutch dietary guidelines 2015. The Hague: Health Council of the Netherlands URL: https://www.healthcouncil.nl/binaries/healthcouncil/documents/advisory-reports/2015/11/04/dutch-dietary-guidelines-2015/advisory-report-dutch-dietary-guidelines-2015.pdf [accessed 2020-12-16]
22. Health COTN. Physical activity guidelines 2017. The Hague: Health Council of the Netherlands URL: https://www.healthcouncil.nl/binaries/healthcouncil/documents/advisory-reports/2017/08/22/physical-activity-guidelines-2017/physical_activity_guidelines_2017_0.pdf [accessed 2020-12-16]
23. Bartlett YK, Sheeran P, Hawley MS. Effective behaviour change techniques in smoking cessation interventions for people with chronic obstructive pulmonary disease: a meta-analysis. Br J Health Psychol 2014 Feb;19(1):181-203 [FREE Full text] [doi: 10.1111/bjhp.12071] [Medline: 24397814]
24. Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models: ISO/IEC 25010:2011. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en [accessed 2020-12-16]
25. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. Boca Raton: CRC Press; 1996.
26. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
27. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. Burlington: Morgan Kaufmann; 2012.
28. Nielsen J. Usability Engineering, 1st Edition. Burlington: Morgan Kaufmann; 1994.
29. Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, et al. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol 2016 May;73:89-102 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.038] [Medline: 26952842]
30. Schalet BD, Hays RD, Jensen SE, Beaumont JL, Fries JF, Cella D. Validity of PROMIS physical function measured in diverse clinical samples. J Clin Epidemiol 2016 May;73:112-118 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.039] [Medline: 26970039]
31. Lee EC, Whitehead AL, Jacques RM, Julious SA. The statistical interpretation of pilot trials: should significance thresholds be reconsidered? BMC Med Res Methodol 2014 Mar 20;14:41 [FREE Full text] [doi: 10.1186/1471-2288-14-41] [Medline: 24650044]
32. Sauro J. 10 Things to Know About the System Usability Scale (SUS). MeasuringU. 2013 Jun 18. URL: https://measuringu.com/10-things-sus/ [accessed 2020-06-03]
33. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate—service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
34. Castleberry A, Nolen A. Thematic analysis of qualitative research data: Is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
35. Creswell J, Poth C. Qualitative Inquiry and Research Design. 4th Edition. Thousand Oaks: Sage Publishers; 2018.

36. PHYSICAL FUNCTION—A brief guide to the PROMIS Physical Function instruments. 2020. URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Physical_Function_Scoring_Manual.pdf [accessed 2020-06-23]
37. Cockton G. The Encyclopedia of Human-Computer Interaction. Chapter 15. Usability evaluation. 2013. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/usability-evaluation [accessed 2020-12-16]
38. Thomas G, Tahir MR, Bongers BC, Kallen VL, Slooter GD, van Meeteren NL. Prehabilitation before major intra-abdominal cancer surgery: a systematic review of randomised controlled trials. Eur J Anaesthesiol 2019 Dec;36(12):933-945 [FREE Full text] [doi: 10.1097/EJA.0000000000001030] [Medline: 31188152]
39. Sanchez-Lorente D, Navarro-Ripoll R, Guzman R, Moises J, Gimeno E, Boada M, et al. Prehabilitation in thoracic surgery. J Thorac Dis 2018 Aug;10(S22):S2593-S2600. [doi: 10.21037/jtd.2018.08.18]
40. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066]
41. de Vries HJ, Kloek CJJ, de Bakker DH, Dekker J, Bossen D, Veenhof C. Determinants of adherence to the online component of a blended intervention for patients with hip and/or knee osteoarthritis: a mixed methods study embedded in the e-exercise trial. Telemed J E Health 2017 Dec;23(12):1002-1010. [doi: 10.1089/tmj.2016.0264] [Medline: 28525310]
42. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
43. Bolshinsky V, Li MH, Ismail H, Burbury K, Riedel B, Heriot A. Multimodal prehabilitation programs as a bundle of care in gastrointestinal cancer surgery: a systematic review. Dis Colon Rectum 2018 Jan;61(1):124-138. [doi: 10.1097/DCR.0000000000000987] [Medline: 29219922]
44. Furze G, Dumville JC, Miles JNV, Irvine K, Thompson DR, Lewin RJP. "Prehabilitation" prior to CABG surgery improves physical functioning and depression. Int J Cardiol 2009 Feb 06;132(1):51-58 [FREE Full text] [doi: 10.1016/j.ijcard.2008.06.001] [Medline: 18703241]
45. Ljungqvist O, Francis N, Urman R. Enhanced Recovery After Surgery: A Complete Guide to Optimizing Outcomes. New York: Springer International Publishing; 2020.
46. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550] [Medline: 26281194]
47. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12). [doi: 10.3390/ijerph15122838] [Medline: 30551555]
48. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]

Abbreviations

CeHRes: Centre for eHealth Research
CONSORT: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
IQR: interquartile range
mHealth: mobile health
PROMIS-PF: Patient-Reported Outcomes Measurement Information System physical functioning 8-item short form
RCT: randomized controlled trial
SUS: System Usability Scale

Edited by L Buis; submitted 13.08.20; peer-reviewed by D Bossen, I Schiering; comments to author 28.09.20; revised version received 20.10.20; accepted 03.11.20; published 07.01.21.

Please cite as:
van der Velde M, Valkenet K, Geleijn E, Kruisselbrink M, Marsman M, Janssen LMJ, Ruurda JP, van der Peet DL, Aarden JJ, Veenhof C, van der Leeden M
Usability and Preliminary Effectiveness of a Preoperative mHealth App for People Undergoing Major Surgery: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(1):e23402
URL: http://mhealth.jmir.org/2021/1/e23402/
doi:10.2196/23402
PMID:33410758

Original Paper

Web-Based Patient Self-Reported Outcome After Radiotherapy in Adolescents and Young Adults With Cancer: Survey on Acceptance of Digital Tools

Marco M E Vogel[1,2], MD; Kerstin A Eitz[1,2,3], PhD; Stephanie E Combs[1,2,3], MD

[1]Department of Radiation Oncology, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

[2]Institute of Radiation Medicine, Helmholtz Zentrum München, Neuherberg, Germany

[3]Deutsches Konsortium für Translationale Krebsforschung, DKTK Partner Site Munich, Munich, Germany

Corresponding Author:
Kerstin A Eitz, PhD
Department of Radiation Oncology
Klinikum rechts der Isar
Technical University of Munich
Ismaninger Str. 22
Munich, 81675
Germany
Phone: 49 89 4140 ext 9705
Email: kerstin.eitz@tum.de

Abstract

Background: eHealth and mobile health (mHealth) are an evolving trend in the medical field. The acceptance of digital tools is high, and the need is growing.

Objective: Young adults (18-40 years) confronted with a cancer diagnosis present unique needs and require special care. They often have a strong affinity and are familiar with modern technology. On that account, we implemented a web-based symptom and quality of life (QoL) assessment to address patients' attitudes and willingness to use mHealth tools. The study also aims to evaluate sociodemographic parameters that could influence patients' opinions.

Methods: A total of 380 young patients aged 18-40 treated with radiotherapy between 2002 and 2017 were included in the trial. We assessed QoL via the European Organization for Research and Treatment of Cancer-Core 30 (EORTC C30) questionnaire and added general questions about mHealth technology. The added questions inquired patients' opinions regarding general aspects, including technical advances in medicine, mobile and app assistance during cancer treatment, data transfer, and app-specific features. The survey was conducted for 12 months. Participation was voluntary and pseudonymized; prior written consent was obtained.

Results: We achieved a participation rate of 57.6% (219/380) and a completion rate of 50.2% (110/219). The median age was 33 years (range 18-40). Of all participants, 89.1% (98/110) considered new technologies in medicine as positive; 10.9% (12/110) answered with neutral. Nearly all patients (96.4%, 106/110) stated that they would send further data via a web-based platform. Of all, 96.4% (106/110) considered the provided pseudonymization of their data as safe. We further asked the patients if they would use a mobile app for symptom and QoL assessment similar to the present web-based system: 74.5% (82/110) answered with yes and 25.5% (28/110) said they would not use a mobile app in the future. We tested the willingness to use an app on several sociodemographic parameters, such as age, gender, education, health insurance status, and cancer-related parameters: tumor stage, time since radiation treatment, and treatment intention. None of these parameters correlated with app use in this group of young adults. Patients who were generally positive regarding using an app rated several possible functions of a future app. The 3 most requested features were appointment reminders (89.0%, 73/82), contact overview of all involved clinics and physicians (87%, 71/82), and making an appointment via app (78%, 64/82).

Conclusions: eHealth and mHealth tools should be available as an integrated part of a comprehensive cancer care approach. It provides automated, thorough documentation of health parameters during therapy and follow-up for doctors, medical staff, and tumor patients to optimize treatment. With this study, we could show that young adults are the ideal patient population to use eHealth/mHealth tools. Such tools offer further digital support and improve the patients' need for constant QoL during cancer care.

(JMIR Mhealth Uhealth 2021;9(1):e19727) doi:10.2196/19727

KEYWORDS

mHealth; eHealth; young adults

Introduction

eHealth and mobile health (mHealth) are an evolving trend in the medical field. The World Health Organization (WHO) defines mHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [1]. Apps for various health areas exist, supporting our everyday life in cases such as diabetes, weight loss, and depression, or tracking our healthy lifestyle with wearables and devices, such as smartwatches, fitness trackers, and blood pressure monitors [2-4]. Therefore, the application of such tools in the oncologic setting should be discussed. Especially, with the recent COVID-19 pandemic, the desire for health tracking of patients with active treatment for cancer is high. The University of Oklahoma initiated a trial evaluating an app which tracks the symptoms (including COVID-19 symptoms) for patients undergoing chemotherapy (NCT04397614). In previous surveys, we showed the positive acceptance of using such tools: 48.5% of the surveyed patients with cancer and 84.3% of the health care professionals (HCPs) support an oncological app complementing treatment [5,6].

Patient-reported outcome (PRO) is an essential tool. PRO is "any report of the status of a patient's health condition that comes directly from the patient without interpretation of the patient's response by a clinician or anyone else" [7]. Convincing studies were performed by Basch et al [8] and Denis et al [9], which suggest that regular contact between patient and HCP via eHealth tools also improves overall survival. Furthermore, Broderick et al [10] reported that the performance status assessment can be improved before initiation of oncologic treatment by PRO. In the literature review by Anatchkova et al [11] regarding PRO use, it became apparent that PRO is still not commonly used in clinical practice. It is emphasized that PRO can support many aspects of cancer care, such as treatment management, monitoring treatment outcomes, quality of life (QoL), and patient communication. The U.S. Food and Drug Administration (FDA) defines QoL as "a general concept that implies an evaluation of the impact of all aspects of life on general well-being" [7].

Young adults (age 18-40 years) confronted with a cancer diagnosis present unique needs and require special care [12]. They differ from pediatric or elderly patients in survival outcomes or epidemiology incidence. Younger patients seem to suffer more in their QoL than older patients. Among others, Champion et al [13] evaluated QoL of breast cancer survivors and showed that younger patients (aged ≤45), compared with older patients (aged 55-70), showed a worse index of well-being (P<.001) as well as worse scoring in most of the scales (eg, fatigue, sleep, and overall sexual functioning) [13]. It seems comprehensible as young adults are the group of patients that are in the middle of life. Therefore, a cancer diagnosis may disrupt their employment, relationships, social life, fertility, or independence [14,15]. The constant measure of QoL is essential in this particular group. It can improve their needs in terms of cancer treatment and aftercare. It might influence decisions by HCPs, which often underestimate patients' preferences and support needs.

In previous studies, we investigated the opinion of 375 patients [5] and 108 HCPs [6] in terms of using mHealth tools in cancer care. We showed that younger patients were more open to modern technologies to support their health (P=.032, r=–0.12) [5]. Of all, 68.7% believed that an app would be an ideal complement to the standard follow-up [5]. In total, 98% of HCPs found regular QoL assessment essential, and 93.5% supported the idea of using such an app for scientific research [6]. Basch et al [16] investigated the self-monitoring of chemotherapy toxicity. They reported an 85% compliance of patients with cancer for using an online platform. For young patients with cancer, tracking symptoms and information seeking with eHealth/mHealth tools is found to be relevant by several authors [17,18]. Ramsey et al [19] performed a literature review on eHealth/mHealth for pediatric patients with cancer (mean age 21 years or younger at the time of diagnosis; mean age 39 years or younger at the time of intervention) and concluded that such interventions may play a crucial role in improving health outcomes of young patients undergoing cancer treatment [19]. These previous studies suggest the overall demand for eHealth/mHealth tools in oncology, especially for younger patients. Combined with the fact that young adults are digital natives and familiar with modern tools, it makes sense to incorporate such apps into their regular cancer care. On that account, we implemented exemplarily a web-based symptom and QoL assessment to address patients' attitudes and acceptance of eHealth/mHealth tools. This study also aims to evaluate sociodemographic parameters that could influence patients' (cohort of young adults in our case) opinions.

Methods

This publication is part of the FABIUS trial, which was designed as a prospective study within the Department of Radiation Oncology, Klinikum rechts der Isar, Technical University of Munich (TUM). We included a total of 380 young patients aged 18-40 years treated with radiotherapy between 2002 and 2017.

We assessed symptoms and QoL after radiotherapy via the European Organization for Research and Treatment of Cancer-Core 30 (EORTC C30) questionnaire [20] and added 5 general questions about mHealth technology. The questions added to the EORTC C30 questionnaire are appended as an English translation (see Multimedia Appendices 1 and 2 for the Questionnaire [German] and English translation of the additional questions, respectively).

Patients were contacted via postal mail and asked to participate in the study via a web-based survey system (Survio sro). The

platform ensured data protection and security (2048-bit SSL security, ISO/IEC 270001 standards, daily backups).

The added questions inquired patients' opinions about general aspects, including technical advances in medicine, mobile, and app assistance during cancer treatment, data transfer, and app-specific features. The questions were developed explicitly for the purpose of this study; however, some questions were similar to those used in previous studies by Kessel et al [5,21]. In these studies, we investigated the general attitude of patients with cancer toward mHealth in clinical routine [5] and performed a usability test of an in-house app for QoL evaluation [21]. We descriptively compared the results of this survey with our previously published data [5,21].

One question per page was displayed. The questions for symptom and QoL assessment were designed in multiple-choice format with a single answer and forced entry according to the EORTC C30 questionnaire [20]. The added questions regarding mHealth allowed either single answers (Q: 31, 32, 34, 36), multiple answers (Q: 38), or optional free text (Q: 33, 35, 37). Q31 was designed as a 3-scale question (yes–neutral–no). Q32, Q34, and Q36 were designed as polar questions (yes or no questions) with branching logic. These questions followed a free-text question, and it was only displayed if the previous question regarding mHealth was answered with "no" and personal concerns and problems were inquired. If necessary, we explained technical terms in a footnote. Because all questions were designed with forced entries or with optional free text, only completed questionnaires could be submitted by the user and were analyzed. The participant was able to revise answers using a back button.

The survey was conducted for 12 months between January and December 2017, according to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) guidelines [22]. Participation was voluntary and pseudonymized; prior written consent was obtained. Each patient received a pseudonym via letter and entered it and the answers in the web-based platform. This way, we were able to reidentify the patient and prevent duplicate entries. The Ethics Committee of the Medical Faculty of Technical University of Munich (TUM) approved the nature and content of the study (Ethics vote: 438/16 S).

In this analysis, we report on the results focusing on the questions about mHealth technology; hence, we will not focus on the QoL measures as there are no comparative values.

We calculated the participation rate as the ratio of unique visitors to the survey site and the total number of contacted patients via letter. The completion rate was calculated using the ratio of completed surveys and the number of unique visitors to the survey.

Statistical calculations were performed using SPSS Statistics version 23 (IBM) in a primarily descriptive way. We used the chi-squared test to test the influence of age, gender, education, health insurance status, tumor stage, time since radiation treatment, and treatment intention. A P value <.05 was considered significant.

Results

Of all patients contacted by letter (n=380), we registered 219 unique visitors. Of those, 110 patients submitted the online survey completely. Fifteen patients left the survey incomplete, and 94 never started the survey. Hence, this results in a participation rate of 57.6% (219/380) and a completion rate of 50.2% (110/219). Median age was 33 years (range 18-40 years); gender distribution was 3:2 (female:male). Table 1 presents the complete participants' characteristics.

Table 1. Participants' characteristics (N=110).

Characteristic	Values
Gender	
Male, n (%)	45 (40.9)
Female, n (%)	65 (59.1)
Age (years), median (range)	33 (18-40)
18-30, n (%)	37 (33.6)
30-40, n (%)	73 (66.4)
Health insurance status	
Privately insured, n (%)	17 (15.5)
State insured, n (%)	93 (84.5)
Education	
High school, n (%)	28 (25.5)
Above high school, n (%)	50 (45.5)
University degree, n (%)	24 (21.8)
Unknown	8 (7.3)
Tumor entity	
Breast cancer and gynecological tumor, n (%)	23 (20.9)
Prostate cancer and urological tumor, n (%)	4 (3.6)
Neurooncological tumor, n (%)	31 (28.2)
Upper and lower gastrointestinal cancer, n (%)	1 (0.9)
Hematological cancer, n (%)	28 (25.5)
Skin cancer, n (%)	2 (1.8)
Head and neck cancer, n (%)	6 (5.5)
Bone cancer, n (%)	4 (3.6)
Soft tissue tumor, n (%)	9 (8.2)
Benign tumor, n (%)	2 (1.8)
Tumor stage	
Advanced/Metastatic, n (%)	48 (44.0)
Low grade, n (%)	62 (56.4)
Treatment intention	
Curative, n (%)	103 (93.6)
Palliative, n (%)	7 (6.4)
Time since radiotherapy (months), median (range)	27 (0.2-178)

Of all participants, 89.1% (98/110) considered new technologies in medicine as positive; 10.9% (12/110) answered with neutral. Nearly all patients (96.4%, 106/110) stated that they would send further data via a web-based platform. Of all, 96.4% (106/110) considered the provided pseudonymization of their data as safe. We further asked the patients if they would use a mobile app for symptom and QoL assessment similar to the present web-based system: 74.5% (82/110) answered with yes and 25.5% (28/110) said they would not use a mobile app in the future.

We tested the willingness to use such an app for symptom and QoL assessment on several sociodemographic parameters, such as age, gender, education, health insurance status, and cancer-related parameters: tumor stage, time since radiation treatment, and treatment intention. None of these parameters correlated with the willingness to use an app in this group of young adults (Table 2).

Table 2. Evaluation of sociodemographic parameters on the willingness to use an app (according to Pearson chi-square tests).

Parameter	*P* value
Age (18-30 vs 31-40)	.846
Gender (male vs female)	.257
Education (low vs medium vs high)	.413
Health insurance status (private vs state)	.843
Tumor stage (advanced vs low-grade)	.220
Treatment intention (curative vs palliative)	.110
Time since radiotherapy (<24 months vs ≥24 months)	.327

The most mentioned reasons against using an app were as follows: smartphones are less safe (7/28), the patient does not want to be reminded of the illness on a smartphone (3/28), no need for one as there is no current treatment (2/28), and not owning a smartphone (2/28). Patients who were generally positive regarding using an app rated several possible functions of a future app (Figure 1). The 3 most requested features were appointment reminders (89%, 73/82), contact overview of all involved clinics and physicians (87%, 71/82), and making an appointment via app (78%, 64/82; Figure 1).

Figure 1. Rating of possible app features by patients willing to use an app (N=82).

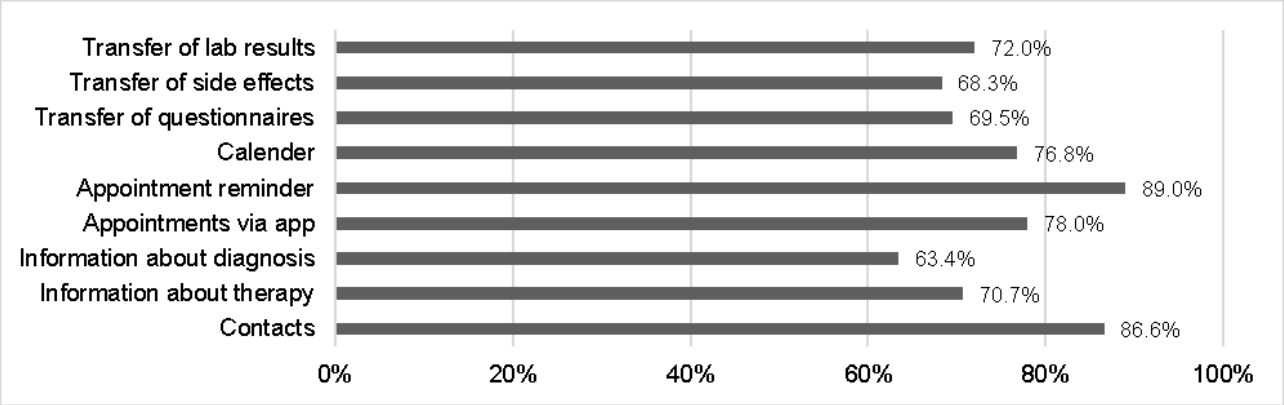
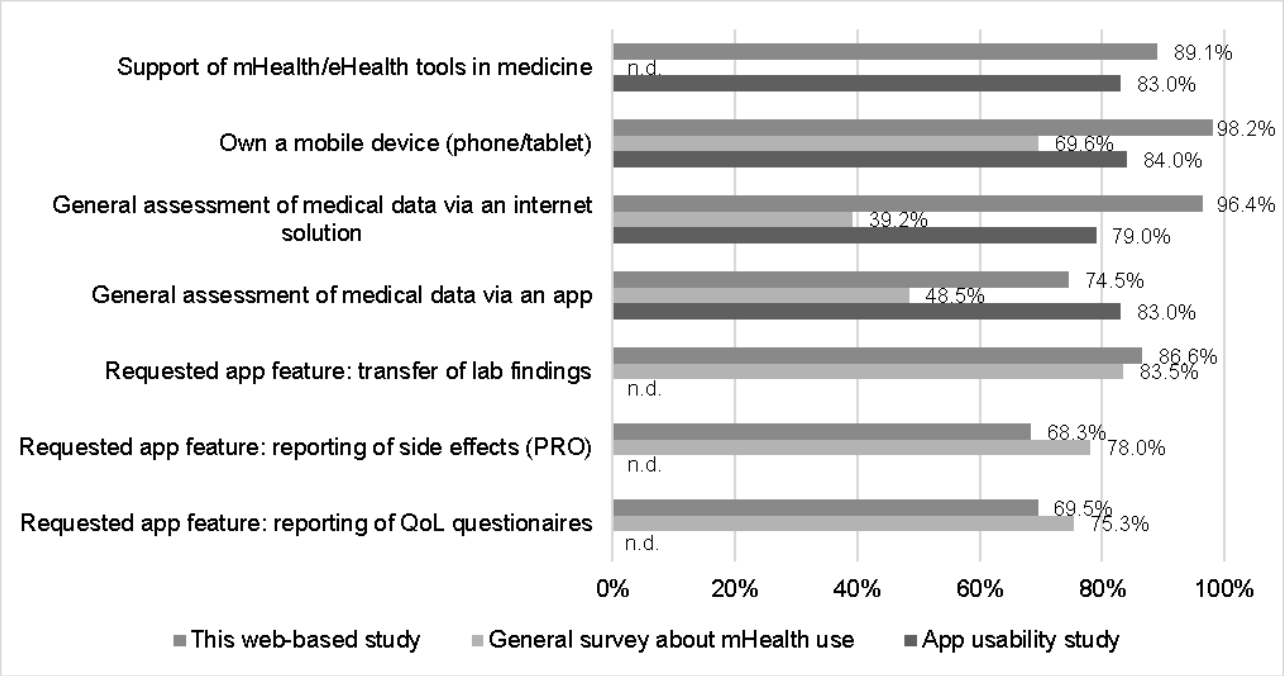


Figure 2 shows the descriptive comparison of questions of this study (median age 33 years; range 18-40 years) with previous surveys by Kessel et al: a general survey of 375 patients with cancer about mHealth use in oncology (median age 59 years; range 18-92 years) [5] and a usability study with 81 patients with cancer on a prototype for an oncologic app (median age 55 years; range 21-80 years) [21]. The surveys were conducted with similar questions as this study.

Figure 2. Comparison of results for this web-based study with the previously published results by Kessel et al regarding a general survey about mHealth use [5] and an app usability study [21]. mHealth: mobile health; PRO: patient-reported outcome; QoL: quality of life.



Discussion

Principal Findings

Patients' compliance with a web-based symptom and QoL assessment depends on their general technical affinity. Our cohort of young adults confirmed that these are the ideal patients to be supported by digital health care as they show high acceptance (96.4%, 106/110). No sociodemographic or cancer-related factors could be found influencing the attitude and willingness to use eHealth/mHealth tools.

The advancing digitalization offers countless possibilities in the health sector: from simple pedometers to complex behavioral therapy for patients with depression. Fitness bracelets, digital blood pressure meter, and blood glucose meters are now connected to the smartphone as a matter of course, and the data are evaluated [2,23,24]. Especially in oncology, due to the many and complex prognostic factors, a broad database including diagnostic, therapy, and regular reported PRO data combined with artificial intelligence–based analyses could have a lasting positive effect on the success of the therapy [25-27].

Previously we showed that young adults are most likely to use modern solutions such as apps and web-based tools to support their cancer treatment [5]. Nearly all patients (96.4%, 106/110) stated that they would send further medical data via a web-based platform. Compared with an app-based assessment, still, 74.5% (82/110) would be willing to use such a tool to support their treatment course. This is comparable to the app usability test we performed on a group of patients with cancer (median age 55; range 21-80 years) [21]. Here 83% would use a web-based and 79% an app-based tool to assess medical data. In a further study where we asked patients with cancer in several departments about their attitude to modern technologies, 39.2% indicated that they would use a web-based and 48.5% an app-based solution [5]. These numbers are relatively smaller as in the other cohorts; however, the survey was conducted without preselecting for age or favorable attitude toward modern technologies. Hence, all the critics and opponents of the idea of implementing an app account for the smaller percentages of acceptance.

In our group of young adults, no sociodemographic or cancer-related factors could be found influencing the attitude and willingness to use mHealth tools. This corresponds well with the high acceptance (89.1%, 98/110) of new technologies in medicine. Certainly, irrespective of whether an app or web-based data transfer is applied, a secure and safe approach must be ensured. Generally, patients understand the concept of pseudonymization and accept it as a safe way of data transmission (96.4%, 106/110). Informing the patient extensively about data management is an important part that should not be underestimated to gain the patients' trust. Besides, to ensure institutional review board approval and fulfill all legal requirements, the data transfer is the most critical part when implementing an eHealth/mHealth solution into a clinical environment [25]. Still, significant obstacles are the lack of technical standards and often difficulties integrating a system that needs external access to the clinic network [25,28,29].

The 3 most desired features of an app were the possibility of making an appointment via app (78%, 64/82), an appointment reminder (89%, 73/82), and the general possibility to store all contacts of the involved physicians in one place (87%, 71/82). The latter makes much sense, as during complex and interdisciplinary cancer treatment, many HCPs from the treating clinic as well as external care providers (eg, radiologists, oncologists, family physicians) are involved. In a review by Iribarren et al [30], mHealth apps and their activities were investigated. All these features were also named as essential activities. Compared to the results of our previous survey about

mHealth [5], the desired app features in this study are equally important to patients of all age groups (Figure 2).

Especially during the time of the COVID-19 virus pandemic, clinicians, especially oncologists, wish for digital/mobile options to contact patients to minimize patient presence while guaranteeing access to treatment and safety of the patients and their families. Patients with cancer are individuals confronted with the most challenging impact as they have acute or chronic medical conditions and often a weakened immune system. With a mobile or web-based connection, it is possible to get regular feedback, such as current health status reported by patients themselves, and decide if a visit, for example, for a chemotherapy session, is possible [31]. During the COVID-19 times, it is evident that the digitalization and implementation of eHealth/mHealth tools are missing and need to be permanently installed in a clinic [32,33].

Limitations

Our study has some limitations. We sent the study invitation to 380 patients, of which 110 participated and completed the survey. Unfortunately, we have no information about the critics' and opponents' attitude to the evaluated web-based QoL assessment and can only present the results of the supporters. We did not subselect patients by tumor type and invited all treated patients between 2002 and 2015. Patients with benign tumors that are no longer in treatment or follow-up might consider themselves as healthy and are most likely not willing to participate in a cancer-related survey.

Future Directions

In the future, to provide a comprehensive solution of eHealth-/mHealth-supported cancer care, all involved parties must agree to an individual, age-specific approach. This includes a web-based and app-based assessment of medical data and the thorough integration into the clinical environment to connect the patient-reported parameters with all health-relevant data. It must be assured that the used software solutions are professional and validated to guarantee patients' safety [25,29,34-37]. Because our data are promising, our goal is to implement an app into our day-to-day clinical routine. However, with the first attempts of developing an own app [21], we quickly realized that such projects must be seen in a broader context. We need to develop across-the-board apps with a variety of interfaces for various medical disciplines. Such projects can only be accomplished with strong partners in politics and industry.

Conclusion

eHealth and mHealth tools should be available as an integrated part of a comprehensive cancer care approach. Such tools provide automated, comprehensive documentation of health parameters during therapy and follow-up care for doctors, medical staff, and tumor patients to optimize treatment. IT departments need to strengthen the implementation and create a comprehensive eHealth solution integrated into the existing IT infrastructure. With the FABIUS trial, we could show that young adults are the ideal patient population to use eHealth/mHealth tools. Such tools offer further digital support and improve the patients' need for constant QoL during cancer care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based original questionnaire.
[PDF File (Adobe PDF File), 121 KB - mhealth_v9i1e19727_app1.pdf]

Multimedia Appendix 2
English translation of the additional questions.
[PDF File (Adobe PDF File), 63 KB - mhealth_v9i1e19727_app2.pdf]

References

1. World Health Organization (WHO). mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization (WHO); 2011.
2. 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]
3. Goyal S, Lewis G, Yu C, Rotondi M, Seto E, Cafazzo JA. Evaluation of a Behavioral Mobile Phone App Intervention for the Self-Management of Type 2 Diabetes: Randomized Controlled Trial Protocol. JMIR Res Protoc 2016 Aug 19;5(3):e174 [FREE Full text] [doi: 10.2196/resprot.5959] [Medline: 27542325]
4. Hernández-Reyes A, Cámara-Martos F, Molina Recio G, Molina-Luque R, Romero-Saldaña M, Moreno Rojas R. Push Notifications From a Mobile App to Improve the Body Composition of Overweight or Obese Women: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Feb 12;8(2):e13747 [FREE Full text] [doi: 10.2196/13747] [Medline: 32049065]

5. 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]
6. Kessel KA, Vogel MM, Schmidt-Graf F, Combs SE. Mobile Apps in Oncology: A Survey on Health Care Professionals' Attitude Toward Telemedicine, mHealth, and Oncological Apps. J Med Internet Res 2016 Nov 24;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]
7. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006 Oct 11;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
8. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. JCO 2016 Feb 20;34(6):557-565. [doi: 10.1200/jco.2015.63.0830]
9. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Overall survival in patients with lung cancer using a web-application-guided follow-up compared to standard modalities: Results of phase III randomized trial. JCO 2016 May 20;34(15_suppl):LBA9006. [doi: 10.1200/jco.2016.34.15_suppl.lba9006]
10. Broderick JE, May M, Schwartz JE, Li M, Mejia A, Nocera L, et al. Patient reported outcomes can improve performance status assessment: a pilot study. J Patient Rep Outcomes 2019 Jul 16;3(1):41 [FREE Full text] [doi: 10.1186/s41687-019-0136-z] [Medline: 31313047]
11. Anatchkova M, Donelson SM, Skalicky AM, McHorney CA, Jagun D, Whiteley J. Exploring the implementation of patient-reported outcome measures in cancer care: need for more real-world evidence results in the peer reviewed literature. J Patient Rep Outcomes 2018 Dec 27;2(1):64 [FREE Full text] [doi: 10.1186/s41687-018-0091-0] [Medline: 30588562]
12. Sodergren SC, Husson O, Robinson J, Rohde GE, Tomaszewska IM, Vivat B, EORTC Quality of Life Group. Systematic review of the health-related quality of life issues facing adolescents and young adults with cancer. Qual Life Res 2017 Jul 1;26(7):1659-1672 [FREE Full text] [doi: 10.1007/s11136-017-1520-x] [Medline: 28251543]
13. Champion VL, Wagner LI, Monahan PO, Daggy J, Smith L, Cohee A, et al. Comparison of younger and older breast cancer survivors and age-matched controls on specific and overall quality of life domains. Cancer 2014 Aug 01;120(15):2237-2246 [FREE Full text] [doi: 10.1002/cncr.28737] [Medline: 24891116]
14. Sodergren SC, Husson O, Rohde GE, Tomaszewska IM, Vivat B, Yarom N, et al. A Life Put on Pause: An Exploration of the Health-Related Quality of Life Issues Relevant to Adolescents and Young Adults with Cancer. Journal of Adolescent and Young Adult Oncology 2018 Aug;7(4):453-464. [doi: 10.1089/jayao.2017.0110]
15. van Leeuwen M, Husson O, Alberti P, Arraras JI, Chinot OL, Costantini A, EORTC QLG. Understanding the quality of life (QOL) issues in survivors of cancer: towards the development of an EORTC QOL cancer survivorship questionnaire. Health Qual Life Outcomes 2018 Jun 04;16(1):114 [FREE Full text] [doi: 10.1186/s12955-018-0920-0] [Medline: 29866185]
16. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an Online Platform for Cancer Patient Self-reporting of Chemotherapy Toxicities. Journal of the American Medical Informatics Association 2007 May 01;14(3):264-268. [doi: 10.1197/jamia.m2177]
17. Hanghøj S, Boisen K, Hjerming M, Elsbernd A, Pappot H. Usability of a Mobile Phone App Aimed at Adolescents and Young Adults During and After Cancer Treatment: Qualitative Study. JMIR Cancer 2020 Jan 02;6(1):e15008 [FREE Full text] [doi: 10.2196/15008] [Medline: 31895046]
18. Stinson J, Gupta A, Dupuis F, Dick B, Laverdière C, LeMay S, et al. Usability testing of an online self-management program for adolescents with cancer. J Pediatr Oncol Nurs 2015 Jul 18;32(2):70-82. [doi: 10.1177/1043454214543021] [Medline: 25037173]
19. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan 19;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
20. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
21. Kessel KA, Vogel MM, Alles A, Dobiasch S, Fischer H, Combs SE. Mobile App Delivery of the EORTC QLQ-C30 Questionnaire to Assess Health-Related Quality of Life in Oncological Patients: Usability Study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e45 [FREE Full text] [doi: 10.2196/mhealth.9486] [Medline: 29463489]
22. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Tay I, Garland S, Gorelik A, Wark JD. Development and Testing of a Mobile Phone App for Self-Monitoring of Calcium Intake in Young Women. JMIR Mhealth Uhealth 2017 Mar 07;5(3):e27 [FREE Full text] [doi: 10.2196/mhealth.5717] [Medline: 28270379]

24. Stephens JD, Yager AM, Allen J. Smartphone Technology and Text Messaging for Weight Loss in Young Adults. The Journal of Cardiovascular Nursing 2017;32(1):39-46. [doi: 10.1097/jcn.0000000000000307]
25. Vogel MME, Combs SE, Kessel KA. mHealth and Application Technology Supporting Clinical Trials: Today's Limitations and Future Perspective of smartRCTs. Front Oncol 2017 Mar 13;7:37 [FREE Full text] [doi: 10.3389/fonc.2017.00037] [Medline: 28348978]
26. Jim HSL, Hoogland AI, Brownstein NC, Barata A, Dicker AP, Knoop H, et al. Innovations in research and clinical care using patient-generated health data. CA Cancer J Clin 2020 May 20;70(3):182-199 [FREE Full text] [doi: 10.3322/caac.21608] [Medline: 32311776]
27. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Optimizing cancer care through mobile health. Support Care Cancer 2015 Jul 4;23(7):2183-2188. [doi: 10.1007/s00520-015-2627-7] [Medline: 25649121]
28. Lavallee DC, Lee JR, Austin E, Bloch R, Lawrence SO, McCall D, et al. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. Mhealth 2020 Jan;6:8 [FREE Full text] [doi: 10.21037/mhealth.2019.09.17] [Medline: 32190619]
29. Giordano F, Welzel G, Siefert V, Jahnke L, Ganslandt T, Wenz F, et al. Digital Follow-Up and the Perspective of Patient-Centered Care in Oncology: What's the PROblem? Oncology 2020 Dec 5;98(6):379-385 [FREE Full text] [doi: 10.1159/000495294] [Medline: 30517946]
30. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb 2;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
31. COVID-19: Supporting Oncology Professionals. URL: https://www.esmo.org/covid-19-and-cancer/supporting-oncology-professionals [accessed 2021-01-03]
32. Wang X, Bhatt DL. COVID-19: An Unintended Force for Medical Revolution? J Invasive Cardiol 2020 Apr;32(4):E81-E82 [FREE Full text] [Medline: 32229647]
33. Elkbuli A, Ehrlich H, McKenney M. The effective use of telemedicine to save lives and maintain structure in a healthcare system: Current response to COVID-19. Am J Emerg Med 2020 Apr 07 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.003] [Medline: 32303410]
34. 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 27;48(7):509-515. [doi: 10.1080/07853890.2016.1195010] [Medline: 27348761]
35. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
36. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
37. 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]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
EORTC C30: European Organization for Research and Treatment of Cancer-Core 30
FDA: U.S. Food and Drug Administration
HCP: health care professional
PRO: patient-reported outcome
QoL: quality of life
TUM: Technical University of Munich
WHO: World Health Organization

Edited by G Strudwick; submitted 29.04.20; peer-reviewed by A Carcone, L McCann; comments to author 19.07.20; revised version received 09.12.20; accepted 12.12.20; published 11.01.21.

Please cite as:
Vogel MME, Eitz KA, Combs SE
Web-Based Patient Self-Reported Outcome After Radiotherapy in Adolescents and Young Adults With Cancer: Survey on Acceptance of Digital Tools
JMIR Mhealth Uhealth 2021;9(1):e19727
URL: http://mhealth.jmir.org/2021/1/e19727/
doi:10.2196/19727
PMID:33427669

Original Paper

Adverse Childhood Experiences and Building Resilience With the JoyPop App: Evaluation Study

Angela MacIsaac[1], MA; Aislin R Mushquash[1], PhD; Shakira Mohammed[1], MA; Elizabeth Grassia[1], HBA; Savanah Smith[2], HBA; Christine Wekerle[2], PhD

[1]Lakehead University, Thunder Bay, ON, Canada

[2]McMaster University, Hamilton, ON, Canada

Corresponding Author:
Aislin R Mushquash, PhD
Lakehead University
955 Oliver Road
Thunder Bay, ON,
Canada
Phone: 1 8073438771
Email: aislin.mushquash@lakeheadu.ca

Abstract

Background: The effects of adverse childhood experiences (ACEs) on mental health, self-regulatory capacities, and overall resilience are well-known. Given such effects, ACEs may play a role in how individuals adjust to challenges later in life. Of interest in this study is the transition to university, a time of heightened stress when adapting to circumstances is required and when those with ACEs may need additional in-the-moment support to exercise resilience. A smartphone app may provide a worthwhile and readily accessible medium for a resilience intervention, provided behavioral outcomes are adequately evaluated.

Objective: This study evaluates the impact of an innovative, smartphone app–based resilience intervention. The JoyPop app was designed to promote resilience through the use of self-regulatory skills such as emotion regulation and executive functioning. Among a sample of first-year undergraduate students, we explored whether use of the app would be associated with positive changes in resilience and related outcomes, and whether these benefits were influenced by level of childhood adversity.

Methods: Participants (N=156) were requested to use the JoyPop app for 4 weeks, at least twice daily. Changes in resilience, emotion regulation, executive functioning, and depression were assessed after 2 and 4 weeks of app usage using multilevel modeling.

Results: The sample of 156 participants included 123 females and 33 males, with a mean age of 19.02 years (SD 2.90). On average participants used the app on 20.43 of the possible 28 days (SD 7.14). App usage was associated with improvements in emotion regulation (χ^2_1=44.46; *P*<.001), such that it improved by 0.25 points on the 18-point scale for each additional day of app usage, and symptoms of depression (χ^2_1=25.12; *P*<.001), such that depression symptoms were reduced by .08 points on the 9-point scale with each additional day of app usage. An interaction between ACEs and days of app usage existed for emotion regulation, such that participants with more adversity evidenced a faster rate of change in emotion regulation (*P*=.02).

Conclusions: Results highlight that daily incorporation of an app-based resilience intervention can help youth who have experienced adversity to improve emotion regulation skills and experience reductions in depression. The JoyPop app represents an important step forward in the integration of resilience intervention research with a technology-based medium that provides in-the-moment support.

(JMIR Mhealth Uhealth 2021;9(1):e25087) doi:10.2196/25087

KEYWORDS

adverse childhood experiences; resilience; emotion regulation; smartphone; app; childhood; emotion; mental health; transition; innovation; intervention

Introduction

Adverse Childhood Experiences

Adverse childhood experiences (ACEs) have profound and long-lasting effects on a broad spectrum of physical and psychological health outcomes. While the assessment of ACEs has changed somewhat since the seminal study by Felitti et al [1], it is well-recognized that the 10 established categories of childhood adversity (divided into household dysfunction and child maltreatment exposures) are associated with mental health difficulties that can last well into adulthood [2-6]. A dose-response relationship exists, such that risk accumulates with each one-unit increase in ACEs [2]; further, categorically, those with high ACEs (eg, 4 or more in adults) are at a significantly greater risk for mental health issues compared to those with low or no ACEs [3,6]. These impacts have remained consistent across cohorts dating back to the 1900s [7]; thus, the impacts of ACEs are seemingly immune to societal, cultural, or health-related changes but instead represent robust effects on mental health and well-being.

The Mechanism of Action of ACEs

While a moderate amount of stress can support the development of coping skills, ACEs may constitute stress that is too overwhelming for an individual's current regulatory processes [8]. Further, some ACEs involve removal of opportunities to learn these regulatory processes, such as in the case of neglect when appropriate self-regulation behaviors are not being modeled by the caregiver. Together, these experiences may lead to disrupted development of corresponding stress-sensitive brain areas [9-12], the results of which are observed behaviorally. For instance, those with ACEs may experience difficulties with emotion regulation, which are associated with a range of negative outcomes such as difficulties with alcohol and interpersonal relationships [13,14]. In addition to emotion regulation, impacts on self-regulation in the form of executive functioning are observed. Those with ACEs evidence deficits in a wide range of executive functions [15] that also impact well-being, resulting in difficulties such as mental health problems [16].

The Transition to University

Life transitions often represent a period of increased challenges and associated stress, since one's environment is in a state of flux and demands that individuals successfully adapt. The transition to university is worth examining as it is common among youth. In Canada, 916,944 youth aged 18-24 enrolled in university for the 2017-2018 school year [17]. In the United States, over 2,000,000 individuals made the transition to university or college in 2016 [18]. While the transition to university itself does not constitute adversity and is instead a time of new and exciting opportunities, it is also marked by stress and an increased prevalence of mental health difficulties. A systematic review spanning two decades of research found the average prevalence of depression in students was 30.6%, much higher than in the general population [19]. Well-being and anxiety may also worsen during university [20]. Surely, the stress of university is not specific to only those with ACEs, rendering it a period where the successful coping of any student is of the utmost importance. Where those with a history of childhood adversity are at a disadvantage, however, is in lacking the resources to cope; those with ACEs may need extra support in exercising resilience during the transition to university.

Resilience

Those with ACEs may be doubly disadvantaged compared to peers when coping with the increased stress associated with the transition to university. Specifically, ACEs serve to challenge one's capacity for resilience. Resilience refers to the capability, resources, and processes available to a person or system to adapt successfully in the face of adversity [21,22]. The operationalization of resilience has changed over time, progressing from focusing on individuals who seemingly possess a unique quality of invulnerability, to variables explaining resilient individuals, to a more recent developmental systems view, whereby resilience is seen as the result of dynamic interactions of various systems (eg, biological and sociocultural systems) [21,22]. An implication of this shift is that resilience is seen as a result of ordinary survival processes common to humans as adaptive creatures, including self-regulatory skills like emotion regulation and executive functioning [21-23]. Moderate amounts of stress allow one to learn such self-regulatory skills and exercise resilience in the future, whereas stressors that are too challenging, such as some ACEs, overwhelm the individual and increase the risk of negative outcomes in the future [24].

Such self-regulatory skills like emotion regulation and executive functioning are vital underlying processes of resilience. In a systematic review of adolescents and young adults, Fritz et al [25] found that components of emotion regulation and executive functions were included among 13 of 25 resilience factors supported by research. Many studies have implicated emotion regulation and executive functioning in the relationship between adversity and overall resilience, studying these abilities as moderators or mediators of resilience outcomes [16,26-28], or examining the separate influence of both self-reported resilience (eg, the Connor-Davidson Resilience Scale [29]) and self-regulation on later mental health outcomes [14]. In recognition of the myriad of ways resilience is operationalized methodologically, we see examining it across various domains, including underlying processes and self-reported resilience, as important. If resilience results when normal adaptive capacities like self-regulation are properly promoted [21], resilience may be restored by the same mechanism [22]. Experts have highlighted a need for resilience supports external to the individual [30,31]. Moreover, it is futile to screen for ACEs in clinical practice without being able to respond with appropriate interventions [32]. The transition to university may be an opportune time to promote resilience because although it is characterized by increased stress, it may be appropriately challenging such that one can learn to exercise resilience if assisted [24]. Moreover, the movement from adolescence into young adulthood is a time when many of those who struggled early on can move onto more positive paths [21] and when self-regulatory brain regions are still developing [33]. Practically, students would benefit from having strategies available for managing emotion that then become part of their

self-regulatory skill repertoire in exercising resilience against the stress of the transition.

Advances in technology have opened doors for implementing this kind of resilience intervention [21]. This includes the use of smartphone apps. The potential of a smartphone app to act as a resilience-promoting tool lies in the readily accessible nature of smartphones, meaning support can be accessed in the moment when it is needed [34]. There is a growing body of empirical literature supporting digital health interventions, including mobile phone apps, as facilitators of improved health and well-being. Specifically, recent evidence supports the use of various digital health interventions targeting coping, stress reduction, self-management skills, and symptom reporting in improving medication and treatment adherence, health knowledge, and anxiety in those with chronic health conditions [35-37]. Other app-based interventions have demonstrated success in improving mental health–related outcomes such as depression, stress, and substance use [38]. Mental health–focused apps may be a cost-effective way of providing psychological support to a wider population who may not otherwise have access to formal interventions [39]. Among university students, smartphone use is ubiquitous, and students have demonstrated a willingness to engage with smartphone health-related apps [40]. At the same time, researchers have described the gap between the incredible number of apps and the demonstration of their safety and efficacy [34]. For example, Donker et al [38] found only 8 studies that assessed outcomes of apps using a pre-post design or control group of the 5464 abstracts they searched. Similarly, in their review, Lui et al [41] concluded there is not enough evidence for the effectiveness of any one individual mental health app and that it was unknown whether there were any adverse effects associated with existing apps. Instead of behavioral outcomes, studies more often focus on aspects of accessibility and usability [42]. With respect to university students specifically, most app studies have examined smoking and alcohol cessation, thus presenting a need for a broader focus across other issues experienced by this population [40].

Aims of This Study

Although the detrimental effects of ACEs on the self-regulatory functions that underlie resilience are well-known, a gap exists in translating these findings into evidence-based interventions that promote resilience in youth with ACEs. The aim of this study was to test whether a smartphone app (JoyPop) [43] promotes resilience over time among youth with varying degrees of ACEs who are navigating the transition to university. Practically, we theorized that the JoyPop app would benefit those transitioning to university by helping them identify, reflect on, and regulate their emotions and improve executive function skills more broadly, contributing to overall resilience against the stress of the university transition. Further, the ease of access of a smartphone app allows the practical benefit of receiving support as needed, which was thought to be conducive to regular usage and the skills becoming routine over time. Resilience was examined across multiple domains, including self-reported resilience and improvement of self-regulatory functions. We also examined a negative outcome of relevance to university students (depression) to ensure the app did not have any unexpected adverse effects, consistent with the gaps identified by Lui et al [41]. More specifically, we examined whether there was a significant rate of change over app usage in these resilience-related outcomes, as well as the relationship between the rate of change and ACEs score. We hypothesized that using the JoyPop app would be associated with a positive rate of change in resilience-related outcomes, with a direct relationship between the number of days the app was used and the amount of change observed, and that this would be most evident among those with higher ACEs scores as they have more to gain from a resilience-promoting intervention focused on bolstering self-regulatory skills.

Methods

Participants and Procedure

This study was approved by the Research Ethics Board at Lakehead University. Students were eligible if they were first-year undergraduate students, owned an iPhone, and were fluent in English. Data were collected in waves over the duration of the school year (both fall and winter semesters). Participants attended 3 laboratory sessions which were run in a group format (pre-app, mid-app [after 2 weeks of app usage], post-app [after 4 weeks of app usage]). Pre-app sessions were run by authors AM, SM, and EG, while mid-app and post-app sessions were run by these authors or supervised research assistants. During the pre-app session, participants received information about the study, provided informed consent, and then were guided through downloading the JoyPop app and provided with a demonstration of all the features. Participants were asked to use the JoyPop app at least twice per day over 4 weeks; no additional requirements were made with respect to feature usage or time spent using the app. Participants returned to the laboratory for mid-app and post-app group sessions, during which time they were requested to complete the self-report measures. Each morning and evening (ie, twice per day), participants were sent reminder emails to use the app. They also received reminder emails to attend their laboratory sessions. If participants encountered any technical difficulties while using the app, they were encouraged to contact the research team, who would liaise with the app development company to resolve any issues. Participants were provided with contact information for mental health supports in the event that they felt distressed at any point throughout their participation in the study. For completing the study, participants received CAD $90 (US $70) in cash or CAD $60 (US $47) and two bonus points toward an eligible psychology course.

JoyPop App

The JoyPop app (see Figure 1 and Table 1) was developed from a cumulative research and parallel consultation approach. Findings from epidemiological and clinical research projects conducted by a federally funded Canadian team highlighted the importance of addressing the role of self-regulation in the link between adversity and mental health outcomes [44] and the resilience value of increasing self-reflection and self-regulation by fostering well-being [45]. This research also suggested there was value in developing internal assets (eg, positive identity [46], well-being [45], self-compassion [47,48]) to support

self-reflection and self-regulation. Consultation with youth, service providers, and clinician-scientists informed design and discovery exercises with app development company Clearbridge Mobile. App features were consequently developed to target daily self-regulation via evidence-based techniques. For example, the Rate My Mood feature was designed to encourage attention towards positive (versus negative) mood states and help users understand and manage their emotions [49]. The Breathing Exercises support self-regulation and decreased physiological arousal [50]. The Journal feature was included in light of the long-term positive benefits evidenced by expressive writing [51], especially when writing is positively focused to foster self-regulation [52]. SquareMoves (a Tetris-like game) was included as activities of this nature can induce a "flow" state, a form of self-regulation that relinquishes negative self-focus [53]. The Art feature provides an opportunity for unrestricted creativity and is supported by positive benefits in memory and emotional expression through doodling [54]. The JoyPop app also includes connections with one's support network (eg, the Circle of Trust feature) or established helplines (eg, through the telephone icon on the launch screen). Following the initial development of the app and associated features, youth involved with child-welfare and victim services as well as providers working closely with youth reviewed and provided direction on the final app features and functionality. Additional information on the development of the JoyPop app and the various features is available online [55] and in Figure 1 and Table 1.

Figure 1. JoyPop app features.

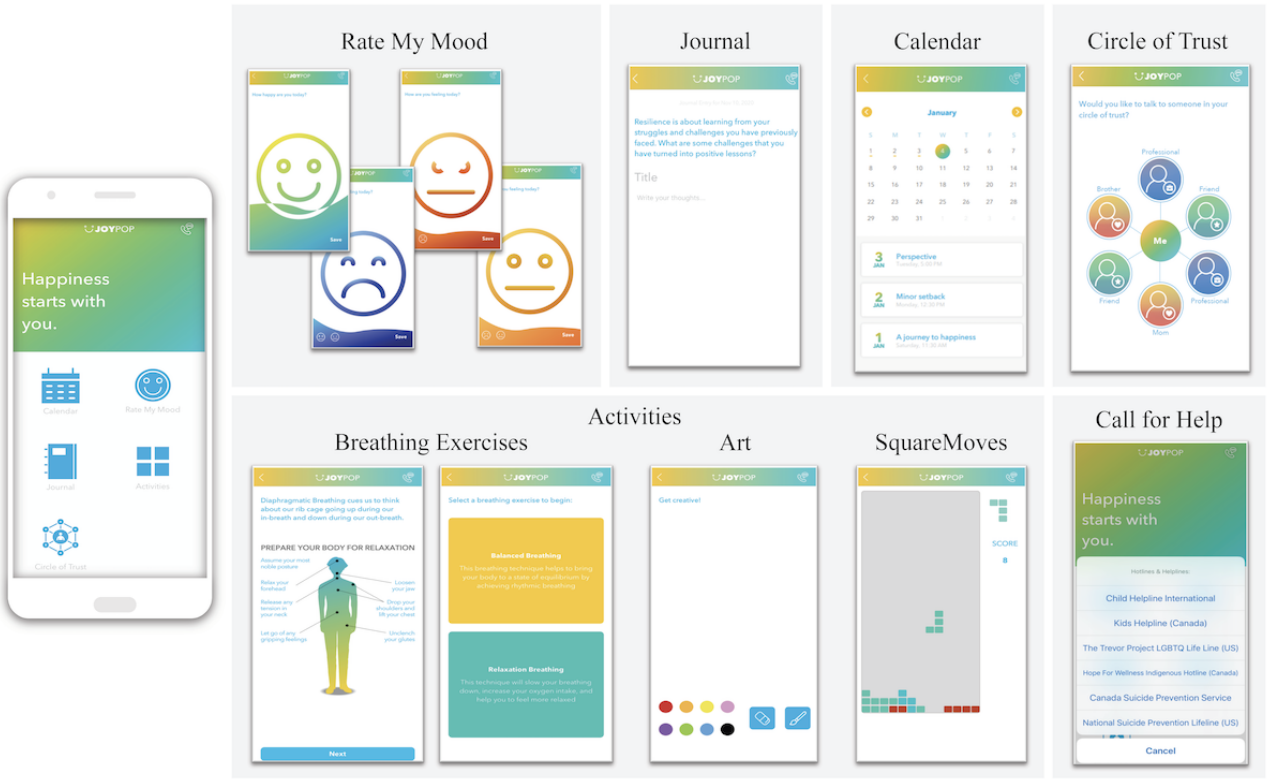


Table 1. JoyPop app features and associated functions.

JoyPop feature	Function
Rate My Mood	Initially prompts users to rate their happiness by sliding a wave of colour up or down to indicate their happiness level. If happiness is rated at 50% or above, the user receives a thumbs up icon and a motivational quote. If happiness rating is lower than 50%, the user is prompted to rate how sad, angry, or "meh" they are feeling using the same technique. Once users have rated their negative emotion, they are given a motivational quote and a prompt to complete an activity from the app.
Journal	Allows the user to complete a journal entry by entering their free-flowing thoughts and emojis or by responding to a resilience-oriented writing prompt at the top of the screen. Users can save their journal entries to the Calendar feature.
Calendar	Allows the user to reflect on previously saved journal entries by date.
Circle of Trust	Allows the user to input up to 6 safe social contacts (ie, by entering their name and phone number) to call if they want to talk or are in need of support. The user can label the contact as a friend, family member, or professional.
Breathing Exercises	Opens to a diagram of the body, with best-practice tips to prepare for relaxation. The user is then prompted to choose between completing a balanced breathing exercise (rhythmic breathing) or a relaxation breathing exercise (slowed breathing). Once the user selects which breathing exercise they would like to complete, they are guided through the breathing exercise with text instructions and an animated diagram.
Art	Allows the user to doodle in colour, swiping their finger across the screen as the paint brush.
SquareMoves	A game in which multi-shaped blocks fall from the top of the screen and the user taps on the shapes to rotate them or swipes them across the screen to move them as they fall to the bottom. Similar to the popular game Tetris, the objective is to form a solid line at the bottom of the screen (with no gaps). Once a line or multiple lines are formed, the blocks in the line break apart and the user is awarded points.
Call for Help	Allows the user to select a 24-hour helpline to call if they are experiencing distress while using the app. The user is provided with culturally specific Canadian and American hotlines (eg, an Indigenous-specific crisis line, LGBTQ helpline) to choose from.

Measures

Adverse Childhood Experiences (ACEs) Questionnaire

A 10-item ACEs Questionnaire [1,56] was used to assess the occurrence of abuse, neglect, and household dysfunction during childhood. This questionnaire was only administered during the pre-app session as responses were not expected to change across time. Items were assessed at the category level as opposed to the level of individual event, and response options were "Never," "At least once," and "Many times" [56]. For this study, an answer of "At least once" or "Many times" was coded as 1, indicating the presence of the category. Previous ACEs questionnaires have assessed domestic violence only against one's mother or stepmother and not that against one's father or stepfather, potentially excluding adverse experiences of a similar nature. We assessed for domestic violence against either of these parental figures in the item focused on domestic violence. The ACEs Questionnaire demonstrates good test-retest reliability [57] and correlates with an inventory of lifetime traumatic exposure, demonstrating construct validity (r=0.69) [56]. In our sample, this measure had a Cronbach α coefficient of .78, indicating acceptable internal consistency.

Connor-Davidson Resilience Scale–10

The Connor-Davidson Resilience Scale–10 (CD-RISC-10) is a 10-item version of the original 25-item CD-RISC [29] designed to measure resilience as grounded in biological, psychological, and social facets [58]. Scores on this scale have demonstrated change in response to intervention [29,59], rendering this measure fitting for our purposes. The revised version correlates highly at r=0.92 with the original measure [29,60]. Items are rated on a 5-point Likert-type scale from "0 – Not true at all" to "4 – True nearly all the time". In undergraduate students, the CD-RISC-10 demonstrated good internal consistency with a Cronbach α of .85, as well as good construct validity [29]. Cronbach α values in our study were .78, .86, and .90 during the pre-app, mid-app, and post-app sessions.

Executive Functioning Index

The Executive Functioning Index is a self-report measure of executive functioning that was created to sample a wide domain of executive functions and to assess the general adult population as opposed to clinical populations [61]. It is a 27-item measure with items rated on a 5-point Likert-type scale from "1 – Not at all much" to "5 – Very." In this study the total score was used. This measure correlates with other measures of executive functioning [61] and predicts impulsive behaviors in delay discounting tasks [62]. Acceptable reliability was demonstrated with a Cronbach α of .82 in the original study [61] and .73, .77, and .76 during the pre-app, mid-app, and post-app sessions of this study.

Difficulties in Emotion Regulation Scale – Short Form

The Difficulties in Emotion Regulation Scale – Short Form (DERS-SF) [63] is a short-form version of the original DERS [64], which measures emotion regulation deficits. This short-form version retains the factor structure of the original measure and correlates highly with it [63]. The DERS-SF consists of 18 items rated on a 5-point Likert-type scale ranging from "Almost Never (0%-10%)" to "Almost Always (91%-100%)." Both the original and short-form scales contain 6 subscales; however, a total score can also be calculated and evidences good reliability and validity [63]. The total score was used in this study, and Cronbach α coefficients demonstrated good reliability at .89 across all timepoints.

Patient Health Questionnaire–9

The Patient Health Questionnaire–9 (PHQ-9) [65] is a self-report measure assessing the diagnostic criteria for depression. It is also used as a general severity measure, as total scores range from 0 to 27 [65]. Construct validity is demonstrated through associations with quality of life, health care utilization, and symptom-related difficulties [65]. In a study of medical patients, Cronbach α values were excellent at .86 and .89 [65], and our sample paralleled these findings with values of .88, .87, and .89 at pre-app, mid-app, and post-app sessions, respectively.

Analytic Plan

Data analyses were conducted using Stata (IBM Corporation). Missing items within questionnaires were imputed with person-mean imputation, while missing data resulting from missing a session were accounted for using maximum likelihood estimation. While person-mean imputation tends to inflate reliability estimates, the risk of this is tolerable if the number of people with missing data and the missing data within each person's measure are less than 15%-20% [66,67]. Maximum likelihood estimation is a preferred method of handling missing data in longitudinal designs and results in relatively unbiased parameters and valid model fit, performing similarly to multiple imputation [68,69]. An attrition analysis was conducted to compare those who attended all 3 sessions to those who missed at least 1 session on their age, sex, ethnicity, ACEs score, and all other pre-app outcome measures. Depending on the variable type, either a *t* test or chi-square test was used.

Multilevel modeling (MLM) was used to assess whether the JoyPop app confers improvements over time in resilience-related outcomes. MLM efficiently assesses longitudinal change while accounting for repeated measurements within the same person by structuring the data in a nested fashion [70]. Specifically, in longitudinal studies, individuals serve as the level 2 variable while time serves as the level 1 variable nested within individuals. Masten and Barnes [21] note that growth modeling (including MLM) represents a statistical advancement to answering questions regarding ACEs and resilience. While early studies separated analyses of person-centred and variable-centred focuses [71], the nested nature of MLM affords the ability to examine resilience-related variables while allowing the individual participant to be the level of analysis. A further departure from previous studies is that the relationship between ACEs score and resilience has been measured in a static fashion, whereas we sought to measure *change* in resilience.

Model building followed steps resembling those outlined by Peugh [72] for each of the four outcome variables separately. Multimedia Appendix 1 contains a description of this process, associated equations, and choice of parameter estimator and covariance structures. In total, two successive models were used for hypothesis testing. The first consisted of time as the only predictor. In this study, time was operationalized as the number of days the individual spent using the app, equal to 0 at pre-app for all participants and to a maximum of 14 and 28 assessed at the mid-app and study completion points, respectively. This approach allows time to vary for each participant, with some participants using the app less than daily or for the full 28 days, whereas using timepoint (pre-app, mid-app, post-app) as the time metric would mask individual differences in app usage [73]. The second model added ACEs score as a predictor to form the ACEs score × days of app usage interaction. A likelihood ratio test assessed whether the second model improved fit compared to the first. This was done for each of the four outcomes.

The interaction effects were interpreted in terms of the slope of days of app usage, that is, testing whether the slope of days of app usage was significantly different from 0 for each ACEs score. Benjamini-Hochberg corrections were used to correct for multiple comparisons; the false discovery rate for these corrections was set to 0.10 [73]. A power analysis was conducted using Monte Carlo simulations. Using parameters similar to the ones obtained in the analyses and a sample of 100, 500 repetitions of the simulation were run to assess power to detect a significant difference between a full model with interactions and a timepoint-only model. Resultant power was 0.92, indicating our final sample size of 156 participants was acceptable after accounting for attrition.

Results

Descriptive Statistics

The sample of 156 participants included 123 females (78.8%) and 33 males (21.2%). The mean age of the sample was 19.02 years (SD 2.90), with a range of 16 to 38. As expected, most of the sample consisted of adolescents (19 years or younger; 137, 87.8%) or youth (24 years or younger; 147, 94.2%) [74]. Of the participants, 109 (69.9%) identified as White, 18 (11.5%) as South Asian, 12 (7.7%) as Black, and the remaining 17 (10.9%) as East or Southeast Asian, Arab, Indigenous, or Latinx. Family income was assessed in categories that ranged from CAD $0-$19,999 (US $15,564) to greater than CAD $200,000 (US $155,649), with a median in the range of CAD $80,000-$99,000 (US $62,260-$77,046). Means and standard deviations for all measures at each timepoint are presented in Table 2. Of the 152 participants who reported on ACEs, 31 (20.4%) reported no ACEs, 25 (16.4%) reported one, 33 (21.7%) reported two, 19 (12.5%) reported three, 44 (28.9%) reported four or more; 4 participants did not provide a response. On average, participants used the app 20.43 of the possible 28 days (SD 7.14). Retention throughout the study was good: of the 156 who enrolled in the study initially, 138 completed the mid-app sessions (88.5%) and 126 (80.8%) completed the post-app sessions. These rates include 3 individuals who missed the mid-app session but then returned for the post-app session. Overall, 123 participants (78.8%) completed all 3 sessions. Figure 2 depicts retention of participants throughout the study. There were no statistically significant differences in age, sex, ethnicity, ACEs score, or other pre-app outcome measures between those who completed all 3 sessions and those who dropped out or missed the mid-app session. Excluding the missing data attributed to missed sessions, the percentages of missing questionnaire items across all participants and items at each timepoint (for which person-mean imputation was used) were as follows: 0.24% for the pre-app sessions, 0.16% for the mid-app sessions, and 0.27% for the post-app sessions. The majority of these missing data resulted from participants who missed one item in a multi-item

questionnaire. As mentioned in our Analytic Plan above, the missing data resulting from a missed session were handled with maximum likelihood estimation.

Table 2. Descriptive statistics for study measures across each timepoint.

Measure	Mean (SD)
ACEs[a]	
Pre-app	2.55 (2.17)
CD-RISC-10[b]	
Pre-app	26.62 (5.95)
Mid-app	27.49 (5.72)
Post-app	27.63 (6.46)
EFI[c]	
Pre-app	95.08 (11.43)
Mid-app	94.50 (12.24)
Post-app	95.63 (12.54)
DERS-SF[d]	
Pre-app	42.91 (13.04)
Mid-app	40.04 (12.59)
Post-app	36.72 (11.40)
PHQ-9[e]	
Pre-app	9.48 (6.09)
Mid-app	8.80 (5.89)
Post-app	7.52 (5.69)

[a]ACEs: Adverse Childhood Experiences Questionnaire.

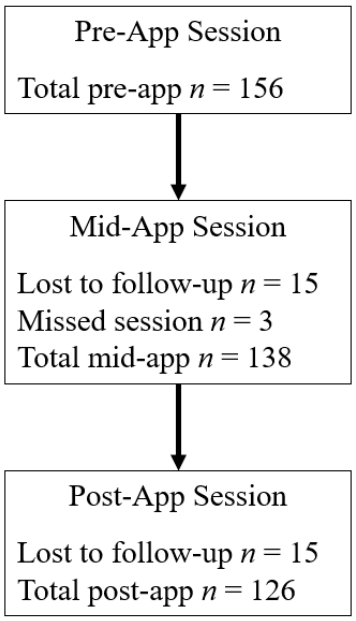
[b]CD-RISC-10: Connor-Davidson Resilience Scale–10.

[c]EFI: Executive Functioning Index.

[d]DERS-SF: Difficulties in Emotion Regulation Scale – Short Form.

[e]PHQ-9: Patient Health Questionnaire–9.

Figure 2. Flow of participant retention throughout the study.



Multilevel Modeling

Table 3 contains the β values, standard errors, and confidence intervals for resilience. The model with days of app usage as the sole predictor was not significant. Adding ACEs score and the 2-way interaction between ACEs score and days of app usage did not further improve the model, and no predictors were significant, meaning that neither app usage nor one's experience of childhood adversity was related to change in resilience.

Table 4 contains the β values, standard errors, and confidence intervals for executive functioning. The model with days of app usage as the sole predictor was not significant. Adding ACEs score and the 2-way interaction resulted in a significant effect of ACEs score at P=.002, such that higher ACEs scores were associated with lower executive functioning (χ^2_3=10.10; P=.02), but the 2-way interaction was not significant, meaning that app usage was not related to a change in executive functioning, nor was there a relationship between app usage and change in executive functioning that depended on one's childhood adversity. This second model demonstrated improved fit to the model with days of app usage as the sole predictor according to a likelihood ratio test (χ^2_2=9.13; P=.01).

Table 3. Coefficients/estimates, standard errors (in brackets), and confidence intervals for the fixed effects of interest and random effects for the Connor-Davidson Resilience Scale–10 models.[a]

Parameter	DAU[b] only, estimate (SE), 95% CI	*P* value	ACEs[c] and DAU, estimate (SE), 95% CI	*P* value
Fixed effects				
Intercept	26.81 (0.45), 25.92 to 27.70	<.001	27.66 (0.69), 26.30 to 29.02	<.001
DAU	0.04 (0.02), 0.00 to 0.08	.08	0.05 (0.03), −0.02 to 0.11	.15
ACEs	N/A[d]	N/A	−0.34 (0.21), −0.74 to 0.07	.105
ACEs × DAU	N/A	N/A	0.00 (0.01), −0.02 to 0.01	.68
Variance components				
Residual	11.95 (1.29), 9.68 to 14.76	N/A	11.94 (1.28), 9.68 to 14.73	N/A
Intercept	21.81 (3.20), 16.36 to 29.09	N/A	21.25 (3.14), 15.92 to 28.38	N/A
Slope	0.01 (0.01), 0.01 to 0.04)	N/A	0.01 (0.01), 0.01 to 0.04	N/A

[a]The number of observations in these models at level 1 (timepoint) is 418; at level 2 (participant) it is 155.

[b]DAU: days of app usage.

[c]ACEs: adverse childhood experiences.

[d]N/A: not applicable.

Table 4. Coefficients/estimates, standard errors (in brackets), and confidence intervals for the fixed effects of interest and random effects for the Executive Functioning Index models.[a]

Parameter	DAU[b] only, estimate (SE), 95% CI	*P* value	ACEs[c] and DAU, estimate (SE), 95% CI	*P* value
Fixed effects				
Intercept	94.77 (0.91), 92.98 to 96.56	<.001	97.92 (1.37), 95.23 to 100.61	<.001
DAU	0.03 (0.04), −0.02 to 0.10	.43	0.00 (0.06), −0.10 to 0.11	.95
ACEs	N/A[d]	N/A	−1.24 (0.41), −2.04 to −0.43	.003
ACEs × DAU	N/A	N/A	0.01 (0.02), −0.02 to 0.04	.55
Variance components				
Residual	30.88 (3.56), 24.63 to 38.72	N/A	30.72 (3.53), 24.52 to 38.48	N/A
Intercept	102.97 (13.55), 79.56 to 133.26	N/A	96.56 (12.83), 74.43 to 125.28	N/A
Slope	0.04 (0.02), 0.02 to 0.12	N/A	0.05 (0.02), 0.02 to 0.12	N/A

[a]The number of observations in these models at level 1 (timepoint) is 416; at level 2 (participant) it is 155.

[b]DAU: days of app usage.

[c]ACEs: adverse childhood experiences.

[d]N/A: not applicable.

Table 5 contains the β values, standard errors, and confidence intervals for difficulties with emotion regulation. Days of app usage was significant in the initial model (P<.001), with a Wald test for the model of χ^2_1=44.46, P<.001, such that difficulties with emotion regulation decreased by an average of 0.25 units on the 18-point scale with each additional day of app usage.

Adding ACEs score and the 2-way interaction to the model, days of app usage, ACE score, and the 2-way interaction were significant at P=.009, P<.001, and P=.02, respectively, with a Wald test for the overall model of χ^2_3=71.22, P<.001. The 2-way interaction was such that difficulties with emotion regulation decreased at a higher rate the higher one's ACEs score: when an individual had no ACEs, their difficulties with emotion regulation score decreased by 0.14 units with each additional day of app usage, but when an individual had an ACEs score of 6 (placing them in the 90th percentile), their score decreased by 0.38 units with each additional day of app usage (see Figure 3; slopes of lines are presented in Table 6). More specifically, for those with no reported childhood adversity, the marginal mean score at pre-app was 37.32, and this decreased to 33.13 when the app was used for 28 days, yet the pre-app marginal mean for those with a high ACEs score of 6 was 50.26 and decreased to 39.57 by 28 days of app usage. Table 6 contains the slopes of days of app usage for each ACEs score. A likelihood ratio test comparing this model to the model that only included days of app usage indicated increased fit (χ^2_2=24.22; P<.001).

Table 5. Coefficients/estimates, standard errors (in brackets), and confidence intervals for the fixed effects of interest and random effects for the Difficulties in Emotion Regulation Scale models.[a]

Parameter	DAU[b] only, estimate (SE), 95% CI	*P* value	ACEs[c] and DAU, estimate (SE), 95% CI	*P* value
Fixed effects				
Intercept	42.81 (0.98), 40.89 to 44.74	<.001	37.32 (1.43), 34.51 to 40.13	<.001
DAU	−0.25 (0.04), −0.32 to −0.18	<.001	−0.15 (0.06), −0.26 to −0.04	.009
ACEs	N/A[d]	N/A	2.16 (0.43), 1.32 to 3.0	<.001
ACEs × DAU	N/A	N/A	−0.04 (0.02), −0.07 to −0.01	.02
Variance components				
Residual	48.72 (4.28), 41.01 to 57.87	N/A	47.86 (4.20), 40.30 to 56.84	N/A
Intercept	108.07 (14.78), 82.66 to 141.29	N/A	93.25 (13.01), 70.95 to 122.57	N/A

[a]The number of observations in these models at level 1 (timepoint) is 417; at level 2 (participant) it is 155.

[b]DAU: days of app usage.

[c]ACEs: adverse childhood experiences.

Figure 3. Changes in difficulties with emotion regulation across time and ACEs score. ACEs: adverse childhood experiences. DERS-SF: Difficulties in Emotion Regulation – Short Form.

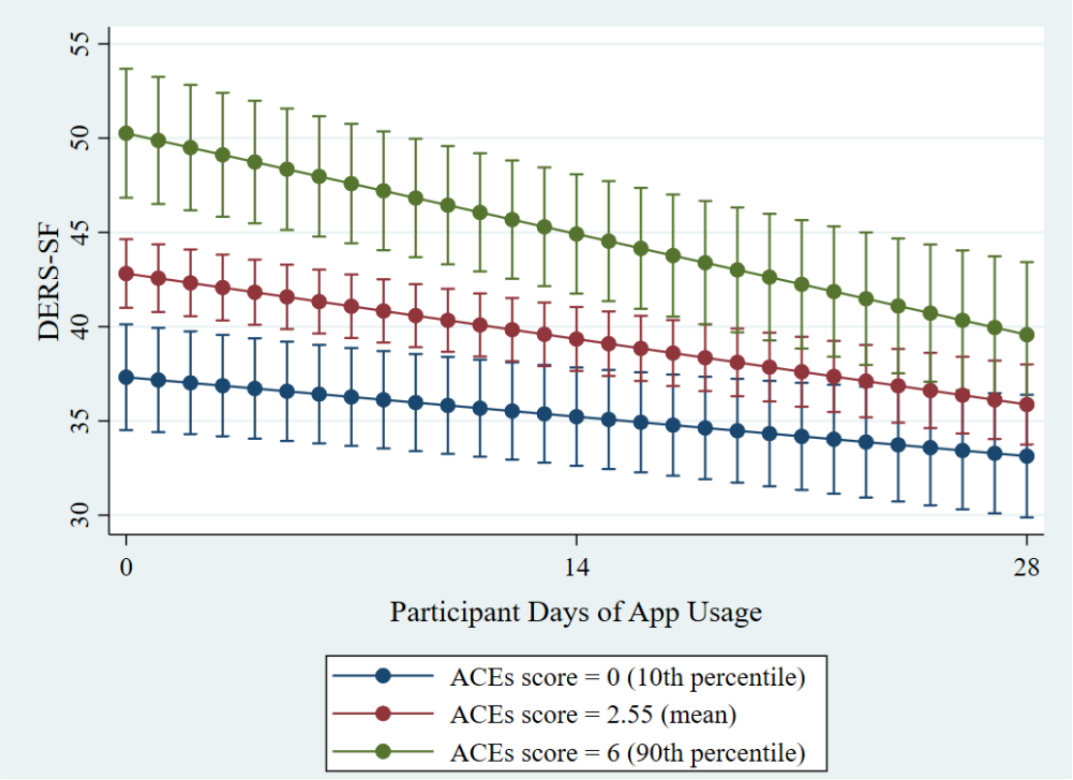


Table 6. DERS-SF – Slopes, standard errors (in brackets), and confidence intervals of participant days of app usage for each ACEs score.

ACEs[a] score	Slope of participant days of app usage, slope (SE), 95% CI	*P* value[b]
0	−0.15 (0.06), −0.26 to −0.04	.009
1	−0.19 (0.05), −0.28 to −0.10	<.001
2	−0.23 (0.04), −0.30 to −0.15	<.001
3	−0.27 (0.04), −0.34 to −0.19	<.001
4	−0.30 (0.04), −0.39 to −0.21	<.001
5	−0.34 (0.05), −0.45 to −0.24	<.001
6	−0.38 (0.07), −0.51 to −0.25	<.001
7	−0.42 (0.08), −0.58 to −0.26	<.001
8	−0.46 (0.09), −0.64 to −0.27	<.001
9	−0.50 (0.11), −0.71 to −0.28	<.001

[a]ACEs: adverse childhood experiences.

[b]After correcting for multiple comparisons using the Benjamini-Hochberg procedure, all significant *P* values remained significant at the .05 level.

Table 7 contains the β values, standard errors, and confidence intervals for depression symptoms. Days of app usage was significant as the sole predictor (*P*<.001), with a Wald test for the model of χ^2_1=25.12, *P*<.001, such that depression symptoms decreased by an average of 0.08 units on the 9-point scale with each additional day of app usage. Adding ACEs score and the 2-way interaction to the model, ACEs score was significant (*P*<.001), and days of app usage remained significant at *P*=.01, meaning that there was a positive conditional effect of ACEs score on depression symptoms and a negative conditional effect of days of app usage. However, the 2-way interaction between ACEs score and days of app usage was not significant, which indicates that the reduction in depression symptoms with increased days of app usage seen in the first model did not depend on one's ACEs score. The model overall was significant at χ^2_3=51.72, *P*<.001, and a likelihood ratio test evidenced improved fit over the days of usage–only model (χ^2_2=24.31; *P*<.001).

Table 7. Coefficients/estimates, standard errors (in brackets), and confidence intervals for the fixed effects of interest and random effects for the Patient Health Questionnaire–9 models.[a]

Parameter	DAU[b] only, coefficient (SE), 95% CI	*P* value	ACEs[c] and DAU, coefficient (SE), 95% CI	*P* value
Fixed effects				
Intercept	9.57 (0.47), 8.65 to 10.49	<.001	6.98 (0.68), 5.65 to 8.31	<.001
DAU	−0.08 (0.02), −0.11 to −0.05	<.001	−0.06 (0.02), −0.11 to −0.01	.01
ACE	N/A[d]	N/A	1.02 (0.20), 0.62 to 1.41	<.001
ACEs × DAU	N/A	N/A	−0.01 (0.01), −0.02 to 0.01	.32
Variance components				
Residual	8.60 (0.75), 7.24 to 10.21	N/A	8.58 (0.75), 7.22 to 10.18	N/A
Intercept	26.79 (0.75), 20.77 to 34.57	N/A	22.52 (2.98), 17.37 to 29.20	N/A

[a]The number of observations in these models at level 1 (timepoint) is 417; at level 2 (participant) it is 155.

[b]DAU: days of app usage.

[c]ACEs: adverse childhood experiences.

[d]N/A: not applicable.

Discussion

Principal Findings

This study sought to assess whether the JoyPop app promotes changes in resilience-related outcomes in first-year undergraduate students over 4 weeks. In particular, this study focused on whether the JoyPop app would be helpful not just for those whose life experiences have paved the way for positive, adaptive development, but for youth whose histories of adversity put them at risk for lacking the foundational capacities that underlie resilience. Our hypotheses were partially supported. We found a dose-response relationship between days of app usage over the study and improvements in difficulties with emotion regulation and depression symptoms. Without asking for whom these changes are evident, it would remain unknown how this relationship is qualified by past adversity. Including ACEs score as a covariate in our models, we saw changes in difficulties in emotion regulation over time depending on one's

ACEs score. Specifically, those with higher ACEs scores had higher mean pre-app difficulties with emotion regulation, but faster rates of change such that the discrepancy with their lower-ACEs counterparts was reduced by the end of the study period. This is consistent with our expectation that those with more adversity have more to gain from an intervention. For depression symptoms, however, reductions with app usage did not differ depending on one's ACEs score, indicating the app was seemingly effective for those with and without histories of adversity.

The main finding with respect to emotion regulation is in line with two claims from previous literature: first, that those with ACEs have poorer self-regulatory capacities [11], and second, that this can be restored through external supports [30,31]. Specifically, this finding is consistent with previous research that asserts a relationship between ACEs and poorer emotion regulation [13,14,75]. Moreover, the observation that both individuals with and without adversity responded to the intervention in terms of emotion regulation, albeit at different rates, supports the notion of resilience as a common process and the merit of bolstering basic adaptive capacities in all individuals. With respect to depressive symptoms, although the effect was not moderated by ACEs score, the finding of a dose-response relationship to app usage is also important considering university students face higher rates of depression [19], heightened distress compared to their pre-university levels [20], and the potential for suicidal ideation with high PHQ-9 scores [76]. Finding improvement in both emotion regulation and depression symptoms could also be viewed as consistent with prior research showing emotion regulation mediates the relationship between ACEs and negative mental health outcomes [13,14]. Specifically, considering the design of the app, reductions in depression may have been related to improved emotion regulation during the university transition. This possibility, however, was not formally tested in this study and could be a direction of future research.

This study contrasts with other intervention studies demonstrating improvements in resilience using the CD-RISC-10 [29] and the notion that resilience should be promoted when basic self-regulatory capacities (including executive functioning) are bolstered [22]. Instead, the JoyPop app did not impact self-reported resilience or executive functioning as hypothesized. At the same time, our findings could be construed as consistent with two alternate views suggested by research. First, if emotion regulation is one of the core components of resilience and plays a mediating role in this regard, then it should change first before more broad changes to general measures of resilience are seen. Consistent with this, Wright et al [22] highlighted that the effects of resilience interventions may take time to occur or may manifest indirectly, and as such, outcomes in multiple domains must be monitored over time. The same logic may be applied to depressive symptoms: emotion regulation is a potential mediating factor between childhood adversity and mental health difficulties like depression symptoms [4]. Second, an alternate view would hold that resilience and emotion regulation, although related, are not inextricably linked. The two constructs have indeed been operationalized as completely independent: for example, Poole et al [75] studied how the effect of emotion regulation on anxiety varies according to resilience scores, examining them separately in a moderating fashion as opposed to a mediating one. Although emotion regulation is indeed a component of resilience, the precise way in which resilience should change when emotion regulation changes has not been addressed by past research. A future direction could be to investigate at which point one's overall trait-like resilience changes as a function of app-related changes in emotion regulation.

Another possibility is that many of the app features are more directly related to emotion regulation than to the domains that did not appear to change. For example, the Rate My Mood feature was meant to increase awareness of emotion, and the Breathing Exercises and Journal feature were meant to regulate emotion. Even SquareMoves, although meant to be more cognitively oriented, can be used as a means of emotion regulation in the form of providing a distraction. In fact, preliminary qualitative exploration of user experiences with the app found just this: many reasons for using certain features had to do with their impact on one's emotions [77]. In addition to this, the Rate My Mood feature opened upon launching the app. Thus, this feature, which is directly tied to emotional capacities, may have been used more often than other features targeting other self-regulatory functions.

There was some consistency with previous literature in the relationship between ACEs and executive functioning. In the initial model with days of app usage, ACEs score, and the 2-way interaction between these variables, the only significant effect was for ACEs score, which suggests higher ACEs scores were associated with lower executive functioning overall. The relationship between ACEs and resilience, however, contrasted with previous research. In our sample, those with a relatively high number of ACEs scored no lower on the CD-RISC-10 resilience measure, on average, than those with fewer. One interpretation is that our sample could be considered resilient from the outset, considering they are doing well enough to attend university, which constitutes higher educational attainment than might be expected in those with difficult life circumstances. It could be that these individuals have succeeded academically because they have exercised resilience despite deficits in depressive symptoms or executive functioning that were associated with higher ACEs scores. For these students, adversity may not have been a barrier to developing resilience but instead functioned as the opposite, equipping them with the hardiness to deal with future stressors. This is consistent with stress-inoculation theory [23] and the challenge model of resilience [8]. Regardless of the reason, finding that the JoyPop app helps with emotion regulation and depression symptoms for a sample of students paves the way for future research to examine the impact of the app with more vulnerable populations.

Strengths, Limitations, and Future Directions

This study had many methodological strengths. The use of MLM represents a sophisticated statistical approach that accounts for the nonindependence of longitudinal data. Inclusion of an individualized, time-varying covariate in the model (that is, days of app usage) allowed a direct dose-response link between app usage and outcomes. This contrasts with studies comparing

post-intervention scores to pre-intervention scores with no consideration for individual differences in engagement with the intervention. Further, our intervention was unique in having few requirements for usage other than encouraging participants to try to use the app twice daily. This flexibility in allowing users to decide when, how often, and in what way to use the app more closely reflects app usage practices outside of a research context. Finally, an asset of this study is the inclusion of multiple facets of resilience, including its theoretical underlying processes and overall self-reported resilience.

In terms of limitations, the study lacked a control group of non–app users, which is a limitation of many app studies [41]. By using individual days of app usage over the study period as the metric of time, however, we are able to conclude that those using the app more evidenced more change than those using it a smaller proportion of days, with individuals who did not often use the app serving as a control group of sorts. Thus, while having a control group would represent the most convincing evidence in asserting that observed changes are due to app use, the relationship between days of app usage and the outcomes of interest helps refute the argument that changes were due to the passage of time. Still, participants who used the app more could have been more motivated to respond to successive questionnaires in a way that would demonstrate improvement.

As Lui et al [41] describe the lack of independent investigations and replications of results with respect to app-based interventions, further studies should attempt to replicate the findings outlined here to continue to build the evidence base for the JoyPop app. Moreover, exploring users' experiences and satisfaction with the app will be important, as positive evaluations of the look, feel, and relevance of a smartphone app can influence ongoing use and engagement [78]. Additional research exploring the economic impact of app-based interventions is also required, as noted in two recent reviews [79,80]. For instance, it will be important to examine the cost-effectiveness and cost utility of integrating an intervention like the JoyPop app into usual care for youth struggling with their mental health. There also remains a need to evaluate long-term outcomes once app usage has ceased to determine whether ongoing usage is needed for maintained benefits or if improvements are sustained when the app is no longer in use. A meta-analysis of mental health–related smartphone apps could not assess this kind of sustainability because so few studies included long-term follow-up [38]. Having more timepoints in general would also allow nonlinear multilevel models to be tested, which could further describe the changes taking place. Finally, while data collection was staggered throughout the school year, another potential limitation is that students in our sample were at different stages of their transition to university. To better understand the impact of the JoyPop app on the transition to university, it would be ideal to recruit students within the first month of attendance.

A final limitation is the dearth of male participants in our sample, which prevented analysis of an interaction between ACEs score and sex in response to the app. Investigating the relationship between sex, adversity, and intervention response is important because there may be several potential sex differences, including rates of childhood adversity [81], neuroendocrine and brain responses to stress and trauma [82,83], susceptibility to certain psychological responses to adversity [84], and patterns of maturation of brain regions involved in stress regulation during adolescence [11]. The lack of male participants may have been a by-product of the relatively large number of psychology students enrolled, who tend to be female, but may also reflect a difference in interest in the app. Indeed, a meta-analysis of mental health interventions aimed at university students found only 24.7% of participants across all studies were male [85], which poses an important question about whether the needs of male students are being met with such interventions. A goal of future investigations will be to recruit a larger sample of male participants whose collective experiences span the range of possible ACEs scores.

In addition to the future research directions described thus far, the relative points of agreement and disagreement between our findings and previous research suggest future areas of study. First, addressing the relationship between emotion regulation and more distal outcomes, including depression and resilience, could be assessed using a longer-term study that may capture changes to the more trait-like CD-RISC-10. A mediation model could be explored to assess the precise way in which changes in emotion regulation might then relate to changes in resilience and depression later on. Further, it will be important to assess the effectiveness of the JoyPop app in a more vulnerable sample, for whom resilience may be a more pertinent construct. Lastly, measures mandated during the COVID-19 pandemic, including abrupt changes to education, mental health services, and social practices, have caused significant disruptions for youth [86]. Consequently, it has become increasingly important to optimize digital approaches to service delivery for youth, including smartphone apps, telemedicine, and virtual therapies [36]. As such, exploring whether the JoyPop app can aid in buffering against the negative effects of stress brought on by the current global pandemic is warranted.

Conclusion

While the JoyPop app appears beneficial for emotion regulation and depressive symptoms, we found the effect on emotion regulation difficulties was qualified by the adversity participants had experienced during their childhood, such that those with ACEs improved the most and more quickly. The fact that those with ACEs were able to benefit in a similar fashion to their non–adversity-exposed counterparts speaks to the capacity for positive change within these individuals and the malleable nature of functions underlying resilience among youth. Importantly, not only did the features of the JoyPop app draw on ACEs research in promoting this positive change, but the smartphone medium of the intervention is an asset in terms of readily accessing help when it is needed. Our findings add to the growing literature on the importance of protecting and promoting self-regulatory capacities through well-timed interventions grounded in theory and research. The JoyPop app represents a positive step forward in catalyzing efforts to support resilience in youth with ACEs, particularly when it comes to regulating emotions during a time of transition.

Conflicts of Interest

The majority of the authors have no conflicts of interest to declare. CW is the creator of the JoyPop app. To mitigate any risk related to conflict of interest, CW was not involved in collecting or analyzing the data. CW's main role was to support the team in evaluating the JoyPop app, and to liaise between the research team and app developers. CW also reviewed the manuscript prior to submission.

Multimedia Appendix 1
Model-building steps and equations.
[DOCX File , 17 KB - mhealth_v9i1e25087_app1.docx]

References

1. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [Medline: 9635069]
2. Campbell JA, Walker RJ, Egede LE. Associations Between Adverse Childhood Experiences, High-Risk Behaviors, and Morbidity in Adulthood. Am J Prev Med 2016 Mar;50(3):344-352 [FREE Full text] [doi: 10.1016/j.amepre.2015.07.022] [Medline: 26474668]
3. Giovanelli A, Reynolds AJ, Mondi CF, Ou S. Adverse Childhood Experiences and Adult Well-Being in a Low-income, Urban Cohort. Pediatrics 2016 Apr;137(4) [FREE Full text] [doi: 10.1542/peds.2015-4016] [Medline: 26966132]
4. Moretti MM, Craig SG. Maternal versus paternal physical and emotional abuse, affect regulation and risk for depression from adolescence to early adulthood. Child Abuse Negl 2013 Jan;37(1):4-13 [FREE Full text] [doi: 10.1016/j.chiabu.2012.09.015] [Medline: 23253857]
5. Schalinski I, Teicher MH, Nischk D, Hinderer E, Müller O, Rockstroh B. Type and timing of adverse childhood experiences differentially affect severity of PTSD, dissociative and depressive symptoms in adult inpatients. BMC Psychiatry 2016 Aug 19;16:295 [FREE Full text] [doi: 10.1186/s12888-016-1004-5] [Medline: 27543114]
6. Shin SH, McDonald SE, Conley D. Patterns of adverse childhood experiences and substance use among young adults: A latent class analysis. Addict Behav 2018 Mar;78:187-192 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.020] [Medline: 29179155]
7. Dube SR, Felitti VJ, Dong M, Giles WH, Anda RF. The impact of adverse childhood experiences on health problems: evidence from four birth cohorts dating back to 1900. Prev Med 2003 Sep;37(3):268-277. [Medline: 12914833]
8. Fergus S, Zimmerman MA. Adolescent resilience: a framework for understanding healthy development in the face of risk. Annu Rev Public Health 2005;26:399-419. [doi: 10.1146/annurev.publhealth.26.021304.144357] [Medline: 15760295]
9. Giedd JN, Rapoport JL. Structural MRI of pediatric brain development: what have we learned and where are we going? Neuron 2010 Sep 9;67(5):728-734 [FREE Full text] [doi: 10.1016/j.neuron.2010.08.040] [Medline: 20826305]
10. Sheridan MA, McLaughlin KA. Neurobiological Models of the Impact of Adversity on Education. Curr Opin Behav Sci 2016 Aug;10:108-113 [FREE Full text] [doi: 10.1016/j.cobeha.2016.05.013] [Medline: 29046891]
11. Tyrka AR, Burgers DE, Philip NS, Price LH, Carpenter LL. The neurobiological correlates of childhood adversity and implications for treatment. Acta Psychiatr Scand 2013 Dec;128(6):434-447 [FREE Full text] [doi: 10.1111/acps.12143] [Medline: 23662634]
12. Leussis M. Is adolescence a sensitive period for depression? Behavioral and neuroanatomical findings from a social stress model. Synapse 2008;62(1):22-30 [FREE Full text] [doi: 10.1002/syn.20462] [Medline: 17957735]
13. Espeleta HC, Brett EI, Ridings LE, Leavens ELS, Mullins LL. Childhood adversity and adult health-risk behaviors: Examining the roles of emotion dysregulation and urgency. Child Abuse Negl 2018 Aug;82:92-101 [FREE Full text] [doi: 10.1016/j.chiabu.2018.05.027] [Medline: 29879586]
14. Poole JC, Dobson KS, Pusch D. Do adverse childhood experiences predict adult interpersonal difficulties? The role of emotion dysregulation. Child Abuse Negl 2018 Jun;80:123-133 [FREE Full text] [doi: 10.1016/j.chiabu.2018.03.006] [Medline: 29604503]
15. DePrince AP, Weinzierl KM, Combs MD. Executive function performance and trauma exposure in a community sample of children. Child Abuse Negl 2009 Jun;33(6):353-361 [FREE Full text] [doi: 10.1016/j.chiabu.2008.08.002] [Medline: 19477515]
16. Horn SR, Roos LE, Beauchamp KG, Flannery JE, Fisher PA. Polyvictimization and externalizing symptoms in foster care children: The moderating role of executive function. J Trauma Dissociation 2018;19(3):307-324 [FREE Full text] [doi: 10.1080/15299732.2018.1441353] [Medline: 29547079]
17. Table 37-10-0015-01 Postsecondary enrolments, by credential type, age group, registration status, program type and gender. Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710001501 [accessed 2020-12-12]

18. National Center for Education Statistics. Enrollment rates [Recent high school completers and their enrollment in 2-year and 4-year colleges, by sex: 1960 through 2012]. 2015. URL: http://nces.ed.gov/programs/digest/d13/tables/dt13_302.10.asp [accessed 2020-12-12]
19. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
20. Bewick B, Koutsopoulou G, Miles J, Slaa E, Barkham M. Changes in undergraduate students' psychological well-being as they progress through university. Studies in Higher Education 2010 Aug 18;35(6):633-645 [FREE Full text] [doi: 10.1080/03075070903216643]
21. Masten AS, Barnes AJ. Resilience in Children: Developmental Perspectives. Children (Basel) 2018 Jul 17;5(7) [FREE Full text] [doi: 10.3390/children5070098] [Medline: 30018217]
22. Wright MO, Masten AS, Narayan AJ. Resilience Processes in Development: Four Waves of Research on Positive Adaptation in the Context of Adversity. In: Goldstein S, Brooks RB, editors. Handbook of Resilience in Children. New York: Springer US; 2013:15-37.
23. Yehuda R, Flory JD, Southwick S, Charney DS. Developing an agenda for translational studies of resilience and vulnerability following trauma exposure. Ann N Y Acad Sci 2006 Jul;1071:379-396 [FREE Full text] [doi: 10.1196/annals.1364.028] [Medline: 16891584]
24. Zimmerman MA. Resiliency theory: a strengths-based approach to research and practice for adolescent health. Health Educ Behav 2013 Aug;40(4):381-383 [FREE Full text] [doi: 10.1177/1090198113493782] [Medline: 23863911]
25. Fritz J, de Graaff AM, Caisley H, van Harmelen A, Wilkinson PO. A Systematic Review of Amenable Resilience Factors That Moderate and/or Mediate the Relationship Between Childhood Adversity and Mental Health in Young People. Front Psychiatry 2018;9:230 [FREE Full text] [doi: 10.3389/fpsyt.2018.00230] [Medline: 29971021]
26. Buckner JC, Mezzacappa E, Beardslee WR. Characteristics of resilient youths living in poverty: the role of self-regulatory processes. Dev Psychopathol 2003;15(1):139-162. [Medline: 12848439]
27. Buckner JC, Mezzacappa E, Beardslee WR. Self-regulation and its relations to adaptive functioning in low income youths. Am J Orthopsychiatry 2009 Jan;79(1):19-30 [FREE Full text] [doi: 10.1037/a0014796] [Medline: 19290722]
28. Obradović J. Effortful Control and Adaptive Functioning of Homeless Children: Variable- and Person-focused Analyses. J Appl Dev Psychol 2010 Mar;31(2):109-117 [FREE Full text] [doi: 10.1016/j.appdev.2009.09.004] [Medline: 20401161]
29. Connor KM, Davidson JRT. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
30. Ungar M. Resilience after maltreatment: the importance of social services as facilitators of positive adaptation. Child Abuse Negl 2013;37(2-3):110-115 [FREE Full text] [doi: 10.1016/j.chiabu.2012.08.004] [Medline: 23260114]
31. Wekerle C, Waechter R, Chung R. Contexts of Vulnerability and Resilience: Childhood Maltreatment, Cognitive Functioning and Close Relationships. In: Ungar M, editor. The Social Ecology of Resilience. New York: Springer; 2012:187-198.
32. Finkelhor D. Screening for adverse childhood experiences (ACEs): Cautions and suggestions. Child Abuse Negl 2018 Nov;85:174-179 [FREE Full text] [doi: 10.1016/j.chiabu.2017.07.016] [Medline: 28784309]
33. Romeo RD. The impact of stress on the structure of the adolescent brain: Implications for adolescent mental health. Brain Res 2017 Jan 01;1654(Pt B):185-191 [FREE Full text] [doi: 10.1016/j.brainres.2016.03.021] [Medline: 27021951]
34. Turvey CL, Roberts LJ. Recent developments in the use of online resources and mobile technologies to support mental health care. Int Rev Psychiatry 2015;27(6):547-557. [doi: 10.3109/09540261.2015.1087975] [Medline: 26523397]
35. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
36. Badawy SM, Radovic A. Digital Approaches to Remote Pediatric Health Care Delivery During the COVID-19 Pandemic: Existing Evidence and a Call for Further Research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
37. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
38. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
39. Prentice JL, Dobson KS. A review of the risks and benefits associated with mobile phone applications for psychological interventions. Canadian Psychology/Psychologie canadienne 2014;55(4):282-290. [doi: 10.1037/a0038113]
40. Johnson K, Kalkbrenner M. The Utilization of Technological Innovations to Support College Student Mental Health: Mobile Health Communication. Journal of Technology in Human Services 2017 Sep 11;35(4):314-339 [FREE Full text] [doi: 10.1080/15228835.2017.1368428]
41. Lui JHL, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Professional Psychology: Research and Practice 2017;48(3):199-210. [doi: 10.1037/pro0000122]
42. Chib A, Lin SH. Theoretical Advancements in mHealth: A Systematic Review of Mobile Apps. J Health Commun 2018;23(10-11):909-955 [FREE Full text] [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]

43. Wekerle C. Finding joy in an app. Futurum Careers. 2019. URL: https://futurumcareers.com/Dr-Christine-Wekerle-Finding-Joy-In-An-App.pdf [accessed 2020-12-12]
44. Park T, Thompson K, Wekerle C, Al-Hamdani M, Smith S, Hudson A, et al. Posttraumatic Stress Symptoms and Coping Motives Mediate the Association Between Childhood Maltreatment and Alcohol Problems. J Trauma Stress 2019 Dec;32(6):918-926. [doi: 10.1002/jts.22467] [Medline: 31742776]
45. Sanders J, Munford R, Thimasarn-Anwar T, Liebenberg L, Ungar M. The role of positive youth development practices in building resilience and enhancing wellbeing for at-risk youth. Child Abuse Negl 2015 Apr;42:40-53 [FREE Full text] [doi: 10.1016/j.chiabu.2015.02.006] [Medline: 25770347]
46. Bell T, Romano E, Flynn RJ. Profiles and predictors of behavioral resilience among children in child welfare. Child Abuse Negl 2015 Oct;48:92-103 [FREE Full text] [doi: 10.1016/j.chiabu.2015.04.018] [Medline: 26002600]
47. Strickland NJ, Wekerle C, Kehayes IL, Thompson K, Dobson K, Stewart SH. Self-compassion as a compensatory resilience factor for the negative emotional outcomes of alcohol-involved sexual assault among undergraduates. International Journal of Child and Adolescent Resilience 2019;6(1):52-69 [FREE Full text] [doi: 10.7202/1069076ar]
48. Tanaka M, Wekerle C, Schmuck ML, Paglia-Boak A, Research Team M. The linkages among childhood maltreatment, adolescent mental health, and self-compassion in child welfare adolescents. Child Abuse Negl 2011 Oct;35(10):887-898 [FREE Full text] [doi: 10.1016/j.chiabu.2011.07.003] [Medline: 22018519]
49. Hilt LM, Hanson JL, Pollak SD. Emotion dysregulation. In: Brown BB, Prinstein MJ, editors. Encyclopedia of Adolescence, 3. San Diego: Academic Press; 2011:160-169.
50. Arch JJ, Craske MG. Mechanisms of mindfulness: emotion regulation following a focused breathing induction. Behav Res Ther 2006 Dec;44(12):1849-1858 [FREE Full text] [doi: 10.1016/j.brat.2005.12.007] [Medline: 16460668]
51. Pascual-Leone A, Yeryomenko N, Morrison O, Arnold R, Kramer U. Does Feeling Bad, Lead to Feeling Good? Arousal Patterns during Expressive Writing. Review of General Psychology 2016 Sep;20(3):336-347 [FREE Full text] [doi: 10.1037/gpr0000083]
52. King L. The health benefits of writing about life goals. Personality and Social Psychology Bulletin 2001;27(7):798-807 [FREE Full text] [doi: 10.1177/0146167201277003]
53. Rankin K, Walsh LC, Sweeny K. A better distraction: Exploring the benefits of flow during uncertain waiting periods. Emotion 2019 Aug;19(5):818-828 [FREE Full text] [doi: 10.1037/emo0000479] [Medline: 30265082]
54. Andrade J. What does doodling do? Appl. Cognit. Psychol 2010 Jan;24(1):100-106 [FREE Full text] [doi: 10.1002/acp.1561]
55. Wekerle C. Resilience in Youth: JoyPop App. 2019. URL: https://youthresilience.net/joypop-app [accessed 2020-12-12]
56. Frewen P, Zhu J, Lanius R. Lifetime traumatic stressors and adverse childhood experiences uniquely predict concurrent PTSD, complex PTSD, and dissociative subtype of PTSD symptoms whereas recent adult non-traumatic stressors do not: results from an online survey study. Eur J Psychotraumatol 2019;10(1):1606625 [FREE Full text] [doi: 10.1080/20008198.2019.1606625] [Medline: 31105905]
57. Pinto R, Correia L, Maia Â. Assessing the reliability of retrospective reports of adverse childhood experiences among adolescents with documented childhood maltreatment. Journal of Family Violence 2014 [FREE Full text] [doi: 10.1007/s10896-014-9602-9]
58. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. J Trauma Stress 2007 Dec;20(6):1019-1028. [doi: 10.1002/jts.20271] [Medline: 18157881]
59. Davidson JRT, Payne VM, Connor KM, Foa EB, Rothbaum BO, Hertzberg MA, et al. Trauma, resilience and saliostasis: effects of treatment in post-traumatic stress disorder. Int Clin Psychopharmacol 2005 Jan;20(1):43-48. [Medline: 15602116]
60. Burns R, Anstey K. The Connor–Davidson Resilience Scale (CD-RISC): Testing the invariance of a uni-dimensional resilience measure that is independent of positive and negative affect. Personality and Individual Differences 2010 Apr;48(5):527-531 [FREE Full text] [doi: 10.1016/j.paid.2009.11.026]
61. Spinella M. Self-rated executive function: development of the executive function index. Int J Neurosci 2005 May;115(5):649-667. [doi: 10.1080/00207450590524304] [Medline: 15823930]
62. Weatherly J, Richard Ferraro F. Executive Functioning and Delay Discounting of Four Different Outcomes in University Students. Personality and Individual Differences 2011 Jul;51(2):183-187 [FREE Full text] [doi: 10.1016/j.paid.2011.03.042]
63. Kaufman EA, Xia M, Fosco G, Yaptangco M, Skidmore CR, Crowell SE. The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and Replication in Adolescent and Adult Samples. J Psychopathol Behav Assess 2015 Nov 23;38(3):443-455. [doi: 10.1007/s10862-015-9529-3]
64. Gratz K, Roemer L. Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. Journal of Psychopathology and Behavioral Assessment 2004 Mar;26(1):41-54 [FREE Full text] [doi: 10.1023/b:joba.0000007455.08539.94]
65. 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] [Medline: 11556941]
66. Bono C, Ried LD, Kimberlin C, Vogel B. Missing data on the Center for Epidemiologic Studies Depression Scale: a comparison of 4 imputation techniques. Res Social Adm Pharm 2007 Mar;3(1):1-27 [FREE Full text] [doi: 10.1016/j.sapharm.2006.04.001] [Medline: 17350555]

67. Downey RG, King C. Missing data in Likert ratings: A comparison of replacement methods. J Gen Psychol 1998 Apr;125(2):175-191. [doi: 10.1080/00221309809595542] [Medline: 9935342]
68. Dong Y, Peng CJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
69. Jelicić H, Phelps E, Lerner RM. Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. Dev Psychol 2009 Jul;45(4):1195-1199 [FREE Full text] [doi: 10.1037/a0015665] [Medline: 19586189]
70. Singer JD, Willett JB. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.
71. Masten AS, Hubbard JJ, Gest SD, Tellegen A, Garmezy N, Ramirez M. Competence in the context of adversity: pathways to resilience and maladaptation from childhood to late adolescence. Dev Psychopathol 1999;11(1):143-169 [FREE Full text] [doi: 10.1017/s0954579499001996] [Medline: 10208360]
72. Peugh JL. A practical guide to multilevel modeling. J Sch Psychol 2010 Feb;48(1):85-112 [FREE Full text] [doi: 10.1016/j.jsp.2009.09.002] [Medline: 20006989]
73. McDonald J. Handbook of Biological Statistics (3rd ed.). Baltimore, Maryland: Sparky House Publishing; 2014.
74. World Health Organization. Adolescent health. URL: https://www.who.int/southeastasia/health-topics/adolescent-health [accessed 2020-12-12]
75. Poole JC, Dobson KS, Pusch D. Anxiety among adults with a history of childhood adversity: Psychological resilience moderates the indirect effect of emotion dysregulation. J Affect Disord 2017 Aug 01;217:144-152 [FREE Full text] [doi: 10.1016/j.jad.2017.03.047] [Medline: 28410477]
76. Garlow SJ, Rosenberg J, Moore JD, Haas AP, Koestner B, Hendin H, et al. Depression, desperation, and suicidal ideation in college students: results from the American Foundation for Suicide Prevention College Screening Project at Emory University. Depress Anxiety 2008;25(6):482-488. [doi: 10.1002/da.20321] [Medline: 17559087]
77. MacIsaac A, Mushquash A, Mohammed S, Wekerle C, Smith S. Promoting Resilience in Undergraduate Students: User Experience with a Resilience App [poster]. 2020 Presented at: St. Joseph's Care Group 2020 Showcase of Health Research; 2020; Thunder Bay, Ontario.
78. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
79. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
80. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
81. Meng X, D'Arcy C. Gender moderates the relationship between childhood abuse and internalizing and substance use disorders later in life: a cross-sectional analysis. BMC Psychiatry 2016 Nov 15;16(1):401 [FREE Full text] [doi: 10.1186/s12888-016-1071-7] [Medline: 27846829]
82. Barr CS, Newman TK, Schwandt M, Shannon C, Dvoskin RL, Lindell SG, et al. Sexual dichotomy of an interaction between early adversity and the serotonin transporter gene promoter variant in rhesus macaques. Proc Natl Acad Sci U S A 2004 Aug 17;101(33):12358-12363 [FREE Full text] [doi: 10.1073/pnas.0403763101] [Medline: 15302939]
83. McClure EB, Monk CS, Nelson EE, Zarahn E, Leibenluft E, Bilder RM, et al. A developmental examination of gender differences in brain engagement during evaluation of threat. Biol Psychiatry 2004 Jun 01;55(11):1047-1055 [FREE Full text] [doi: 10.1016/j.biopsych.2004.02.013] [Medline: 15158422]
84. Olff M, Langeland W, Draijer N, Gersons BPR. Gender differences in posttraumatic stress disorder. Psychol Bull 2007 Mar;133(2):183-204 [FREE Full text] [doi: 10.1037/0033-2909.133.2.183] [Medline: 17338596]
85. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
86. Findlay LC, Arim R, Kohen D. Understanding the Perceived Mental Health of Canadians During the COVID-19 Pandemic. Health Rep 2020 Jun 24;31(4):22-27 [FREE Full text] [doi: 10.25318/82-003-x202000400003-eng] [Medline: 32644764]

Abbreviations

ACE: adverse childhood experience
CD-RISC-10: Connor-Davidson Resilience Scale–10
DERS-SF: Difficulties in Emotion Regulation – Short Form
MLM: multilevel modeling
PHQ-9: Patient Health Questionnaire–9

Edited by G Eysenbach; submitted 17.10.20; peer-reviewed by S Abegglen, S Badawy; comments to author 02.11.20; revised version received 27.11.20; accepted 12.12.20; published 04.01.21.

Please cite as:
MacIsaac A, Mushquash AR, Mohammed S, Grassia E, Smith S, Wekerle C
Adverse Childhood Experiences and Building Resilience With the JoyPop App: Evaluation Study
JMIR Mhealth Uhealth 2021;9(1):e25087
URL: http://mhealth.jmir.org/2021/1/e25087/
doi:10.2196/25087
PMID:33393908

Review

Effectiveness of Mobile Apps to Promote Health and Manage Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials

Sarah J Iribarren[1], PhD; Tokunbo O Akande[2*], MD; Kendra J Kamp[3*], PhD; Dwight Barry[4], PhD; Yazan G Kader[1], BA, MSc; Elizabeth Suelzer[5], MLiS

[1]Department of Biobehavioral Nursing and Health Informatics, University of Washington, Seattle, WA, United States

[2]Department of Pediatrics, Sanford Health, Bemidji, MN, United States

[3]Division of Gastroenterology, School of Medicine, University of Washington, Seattle, WA, United States

[4]Enterprise Analytics, Seattle Children's Hospital, Seattle, WA, United States

[5]Medical College of Wisconsin Libraries, Medical College of Wisconsin, Milwaukee, WI, United States

[*]these authors contributed equally

Corresponding Author:
Sarah J Iribarren, PhD
Department of Biobehavioral Nursing and Health Informatics
University of Washington
Health Science Building, T602C, Box 357266
1959 NE Pacific Street
Seattle, WA, 98195
United States
Phone: 1 2065435211
Email: sjiribar@uw.edu

Abstract

Background: Interventions aimed at modifying behavior for promoting health and disease management are traditionally resource intensive and difficult to scale. Mobile health apps are being used for these purposes; however, their effects on health outcomes have been mixed.

Objective: This study aims to summarize the evidence of rigorously evaluated health-related apps on health outcomes and explore the effects of features present in studies that reported a statistically significant difference in health outcomes.

Methods: A literature search was conducted in 7 databases (MEDLINE, Scopus, PsycINFO, CINAHL, Global Index Medicus, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews). A total of 5 reviewers independently screened and extracted the study characteristics. We used a random-effects model to calculate the pooled effect size estimates for meta-analysis. Sensitivity analysis was conducted based on follow-up time, stand-alone app interventions, level of personalization, and pilot studies. Logistic regression was used to examine the structure of app features.

Results: From the database searches, 8230 records were initially identified. Of these, 172 met the inclusion criteria. Studies were predominantly conducted in high-income countries (164/172, 94.3%). The majority had follow-up periods of 6 months or less (143/172, 83.1%). Over half of the interventions were delivered by a stand-alone app (106/172, 61.6%). Static/one-size-fits-all (97/172, 56.4%) was the most common level of personalization. Intervention frequency was daily or more frequent for the majority of the studies (123/172, 71.5%). A total of 156 studies involving 21,422 participants reported continuous health outcome data. The use of an app to modify behavior (either as a stand-alone or as part of a larger intervention) confers a slight/weak advantage over standard care in health interventions (standardized mean difference=0.38 [95% CI 0.31-0.45]; I2=80%), although heterogeneity was high.

Conclusions: The evidence in the literature demonstrates a steady increase in the rigorous evaluation of apps aimed at modifying behavior to promote health and manage disease. Although the literature is growing, the evidence that apps can improve health outcomes is weak. This finding may reflect the need for improved methodological and evaluative approaches to the development and assessment of health care improvement apps.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42018106868; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=106868

(JMIR Mhealth Uhealth 2021;9(1):e21563) doi:10.2196/21563

KEYWORDS

systematic review; mobile apps; mobile phone

Introduction

Background

The health care field is experiencing exponential growth in the use of mobile apps to deliver interventions aimed at modifying behavior to promote health and manage disease. Behavior change interventions are broadly defined as "coordinated sets of activities designed to change specific behavior patterns" [1]. Traditionally, behavior change interventions are resource intensive and difficult to scale [2]. The enthusiasm for apps stems from their broad reach and capacity to perform multiple functions, including sophisticated features that can enhance person-centered care and improve health outcomes [3]. Research into the effectiveness of health apps to change behaviors is in its early stage, and there is no clear consensus on which specific features of apps can assist in behavior change [4]. Furthermore, most apps contain only a few features that could be considered to have the potential to change behavior [5]. Examples of features within apps that may promote health behavior change include *reminders or notifications* (eg, to prompt patients to take their medication at a specified time), *tracking* activity (eg, to encourage increased physical activity), goal planning, and tailored *information* (eg, provide information on the consequences of continuing a behavior) [6]. Despite the field being in its infancy and the lack of consensus on the efficacy of apps to promote health and manage disease, the use of health apps has become increasingly common [7].

With an estimated 325,000 health care–related apps now in existence [8], the global health app market is expected to reach US $236 billion by 2026 [9]. However, systematic reviews assessing the effectiveness of mobile apps for the management of various conditions such as asthma [10], cardiovascular disease [11], diabetes [12], physical and mental health [13], self-management of medication [14], and smoking cessation [15] have shown mixed results [16].

Similarly, findings from studies exploring the impact of app features on health outcomes have been mixed. For instance, Bonoto et al [12] observed that when the number of features in an app was more than 2, the efficiency of the app appeared to increase. However, other researchers have reported inconclusive results regarding the number of features [17,18]. To inform the future direction of health app development, there is a need to understand if certain features or *active ingredients*, such as self-tracking, feedback, and journaling, within apps are important contributors to improved outcomes [1]. Thus far, only a few systematic reviews of rigorously evaluated apps (eg, through randomized controlled trials [RCTs]) have been performed to assess which features are found in successful apps. In addition, these reviews have typically only looked at disease-specific apps [12,15,18-21]. Furthermore, none of those studies specifically evaluated which features of the apps had an effect on the outcomes.

Objectives

Therefore, a broader review of this technology is warranted, given the widespread availability of health apps and the current gap in knowledge in the understanding of their effectiveness on health outcomes or the features that may have an effect on the outcomes. This study aims to summarize the evidence from RCTs of the effect of health apps on health outcomes. Our secondary aim is to explore the effect of features present in studies that reported a statistically significant difference in health outcomes (eg, feature count and which features were more likely to have a positive effect on the outcome). To this end, all studies rigorously evaluating a health app published between 2008 and early 2019 were reviewed and analyzed.

Methods

This study was designed and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement [22]. A protocol outlining the methods of this systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO, CRD42018106868).

Search Strategy

The search strategy was developed and executed in consultation with an experienced research librarian (ES). We executed our original search on October 2016 and completed an updated search on March 2019. The search strategy was created for Ovid MEDLINE using a combination of MeSH (Medical Subject Heading) terms, keywords, and phrases (Multimedia Appendix 1). The search terms targeted mobile apps, a broad range of disease/illnesses, and health-related outcomes. The strategy was translated for other databases—Scopus, PsycINFO, CINAHL, Global Index Medicus, Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews—using their respective thesaurus terms and advanced search features. A manual search using reference lists of retrieved citations was conducted for other relevant studies.

Eligibility Criteria

Textbox 1 summarizes the PICOS (participants, interventions, comparison, outcomes, and study design) to define the inclusion criteria strategy. A behavior change intervention is broadly defined as a "coordinated set of activities designed to change a specified behavior pattern" [2]. Given this broad definition, we included studies in which an intervention used a mobile health (mHealth) app to change a behavior to promote health or manage disease, whether or not authors explicitly labeled the intervention as *behavior change technique*. For example, apps that included

techniques to increase exercise, improve medication adherence, or self-manage a chronic disease were included. Multimedia Appendix 2 [1] describes and defines the app features. There is currently no consensus on all app features and their definitions. We first developed this list of features based on the literature [20,23] and then, by consensus within the team, iteratively expanded and refined from our review of the included studies. We then assessed how the features corresponded to behavior change techniques [24]. We excluded apps that only collected data passively, with no other intervention or behavior change component (eg, step count collection, blood glucose automatic readings/continuous glucose reading). As app stores from which users could download apps were first launched in 2008, we used this year for the start of search [25]. We excluded articles published in languages other than English or Spanish, not meeting rigorous evaluation criteria, and reporting on behavior change interventions that did not include an app for delivery. In addition, the following studies were excluded: studies on app-based interventions targeting health care providers (eg, training, evaluation of prescription habits, diagnostic assistance, medical information references); interventions designed for nonmobile devices (eg, web-based interventions for computer use), using only wearable devices, or if the intervention was delivered within a health care facility (eg, hospital unit); testing of a smart-health device (eg, blood pressure monitoring, environmental sensors, blood glucose monitoring) or for monitoring a device (eg, pacemaker); reporting on app development; nonrigorously designed studies (nonrandomized studies, not controlled, quasi-experimental); virtual reality studies; systematic review of mobile apps; assessment of web-based networking (eg, blogs, Facebook groups); devices used in health care settings (eg, tablet-based intervention in a hospital setting); use of drones; and app only for notifying users of test results.

Textbox 1. Participants, interventions, comparison, outcomes, and study design criteria.

Inclusion and exclusion criteria

Population

- Adults or children with any disease or health-related issue.

Intervention

- Any intervention using a mobile health app aimed to modify a behavior to promote health and manage disease whether or not authors explicitly labeled the intervention as "behavior change technique." For example, an intervention that allows users to input data, receive feedback, connect with health professionals, learn about a disease, or manage their illness or disease.
- Intervention could have any length of follow-up on outcomes.

Comparator

- Routine practice
- Usual care
- Control
- Attention control

Outcomes

- Any direct health outcome that could be assessed for clinical effectiveness, such as medication adherence, treatment outcome (eg, blood pressure, hemoglobin A_{1c}), health care promotion (eg, cardiovascular disease screening), behavior change (eg, smoking cessation, weight loss), scores measured using any validated standard instrument (eg, pain level, Patient Health Questionnaire for depression, quality of life).

Setting

- Participants in any country in their natural environment (eg, home, community) and not in a controlled health care setting (eg, hospital).

Study design

- Rigorous evaluation conditions: parallel randomized controlled trials (RCTs), cluster RCTs, quasi RCTs, controlled before-after studies, or interrupted time series studies with at least three time points before and after the intervention.

Study Screening and Selection

Studies were screened for eligibility in duplicates under blinded conditions by 2 independent reviewers (SI, KK, YK, Hannah Erdy, and TA) as the best practice for systematic reviews [22]. Covidence software was used to aid in this process. Search results were first screened by title and abstract, and any studies that appeared to meet the eligibility criteria or where eligibility was unclear progressed to full-text screening. Next, 2 independent reviewers screened the full texts to determine eligibility for inclusion in the review. Results from each round of screening were compared and discussed until consensus was attained. When more than one publication referred to the same trial, the publication reporting on the primary outcomes of the study was selected. Reports of ongoing trials were excluded.

Risk of Bias

The quality of the studies was assessed using the Cochrane Risk of Bias tool for RCTs [26]. We assessed each study for random sequence generation, allocation concealment, blinding of participants, blinding of personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias as low, unclear, or high. If intention-to-treat analysis was conducted, the study was considered to have a low risk for completeness of outcome data.

Data Extraction and Variable Definitions

A standardized data extraction form was developed to include standard data (eg, title, year published, health condition) and study-specific data (eg, app features, level of personalization). Data extracted from each full-text paper included the following: title, publication year, authors, country, intervention name, health condition addressed, participant population age, study design, intervention focus, intervention description, mode of intervention delivery (eg, stand-alone app or app as part of a larger intervention such as an in-person component), comparator, intervention frequency, follow-up time, sample size calculation, use of theoretical framework to guide and develop intervention, app name, app features, level of personalization, health outcome, sample size at randomization and at final analysis, and outcome results. Features were extracted as reported or not reported. In cases where the features were not explicitly listed, we reviewed supplementary files, screenshots, or any source cited by the authors (protocol; website; or previous publications by authors detailing intervention development, testing, or initial evaluation) to determine if features were present or not. The level of personalization of the app was categorized as one-size-fits-all, static (one-time tailoring to individual), or dynamic (adaption occurs periodically or in real time during intervention). For health outcomes, we extracted the primary health outcome used to power the study; if there were multiple outcomes or if the power calculation was not reported, we selected the main health outcome. If the health outcome was a secondary outcome, this was noted. Where studies compared more than 2 interventions, we extracted the data from the control and the intervention groups that represented the most stand-alone app. Outcome data were extracted from the longest (last) follow-up. We abstracted summary measures such as means, SDs, event counts, and total n, calculating SD values from other summary metrics (eg, SE, CIs) when the raw value was not included in a study [27,28]. These data were compiled in a spreadsheet independently by 2 of the authors for a subset of 20 articles and were reviewed by the team for consensus of data extraction. Data extraction was then continued by one author and verified by a second extractor. When there was uncertainty for any data point, the article was reviewed as a team, and disagreements were resolved by discussion.

Narrative Synthesis for Feature Definition and Outcome Categorization

Initial narrative synthesis is recommended before undertaking a quantitative synthesis of complex interventions to look at patterns and characteristics of the data identified [27]. Recommendations include organizing studies into logical categories possibly related to design, outcome, or intervention type [29]. Therefore, we used a thematic analysis to develop intervention outcome behavior change categories. These categories included nutrition and physical activity (eg, weight loss), mental health management (eg, reduction of depressive symptoms), medication adherence, general health and well-being (eg, quality of life), diabetes management (eg, hemoglobin A_{1c} management), management of other chronic diseases (eg, lowered blood pressure), and cessation/harm reduction (eg, days no drug use, reduced alcohol consumption, smoking cessation). Country income levels (eg, high and low income) were classified according to the World Bank listing [30].

Statistical Analysis

Meta-analysis

We used a random-effects model to calculate pooled effect size estimations [31], where study heterogeneity was assessed using I^2 and the Sidik-Jonkman estimator for calculating τ^2 and prediction intervals [32]. Random effects analysis was chosen over a fixed-effect model because by calculating both within- and between-study variance (τ^2), the relative weights assigned to each study are more balanced under the random-effects model than they are under a fixed-effect model, which only considers n for each study. Standardized mean differences (SMDs) were calculated as Hedges *g* for continuous measures and odds ratio for binary measures; when extracted data did not contain SDs but contained other statistical information (*t* values, *P* values, and CIs), we used those variables to calculate effect sizes following the formulae in the Cochrane Handbook [33]. Funnel plots were used to assess publication bias.

Sensitivity Analysis

Sensitivity analysis was conducted by performing subgroup analysis based on follow-up time, stand-alone app interventions, level of personalization, and pilot studies.

App Feature Analysis

We used logistic regression to examine the effect of app features detailed in Multimedia Appendix 2. We chose an outcome of an absolute SMD of 0.5 or greater as indicative of a successful app, which was modeled as the outcome variable in the logistic regression.

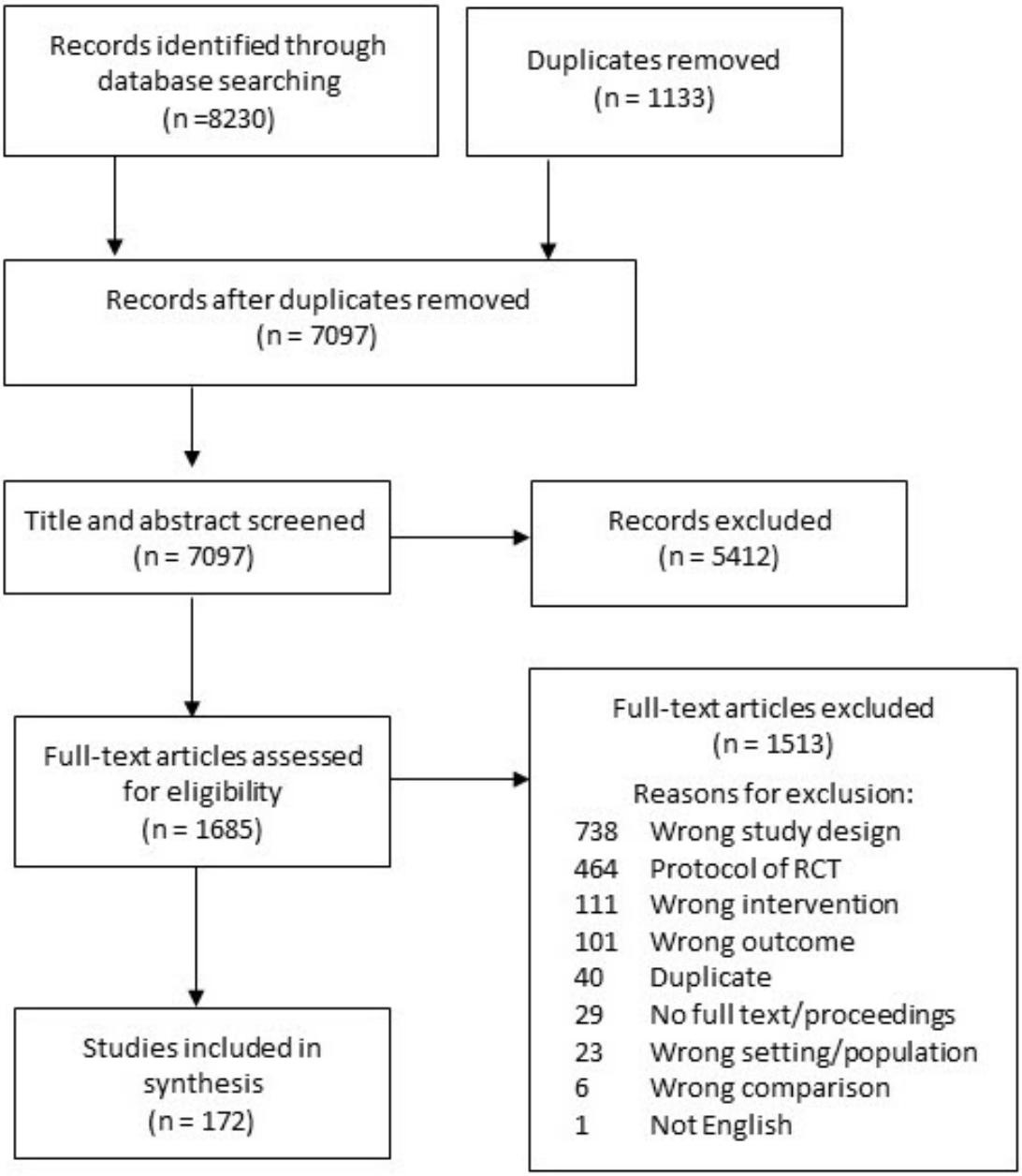
Statistical analyses were performed using R 3.6.2 (R Core Team 2019).

Results

Study Selection

From the database searches, 8230 records were initially identified. Of these, 172 met the inclusion criteria and were included in this review. See Figure 1 for the PRISMA flowchart of the study selection, including the rationale for the exclusion of full-text articles.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RCT: randomized controlled trial.



Study Characteristics

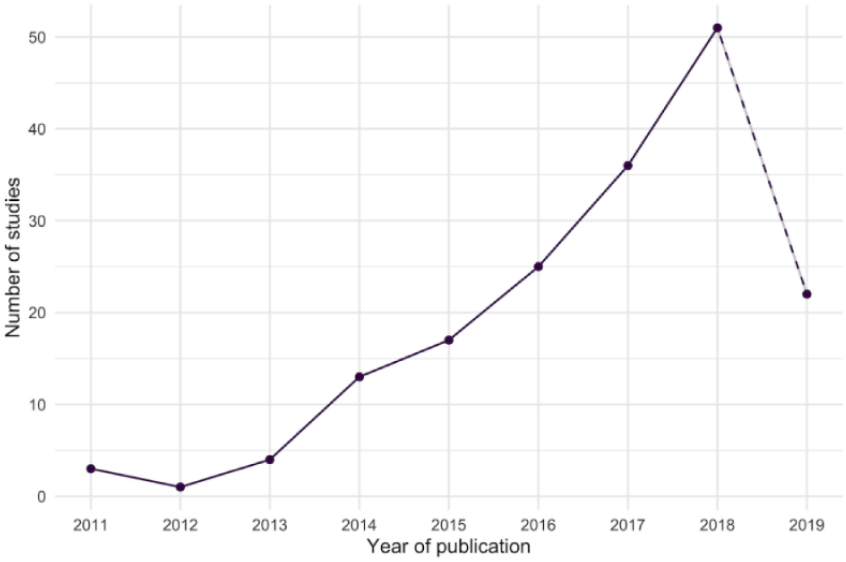
Key study characteristics are summarized in Table 1. Most studies were parallel design or cluster RCTs (169/172, 98.3%), 2 were randomized crossover, and 1 was a randomized factorial design. A total of 34.3% (59/172) of the studies were characterized as pilot RCTs. Studies were predominantly conducted in high-income countries (164/172, 94.3%), with the United States having the highest count (69/172, 46.6%). Only 1 study was conducted in a low-income country (Ghana). The number of publications per year has steadily increased since 2012 (Figure 2).

Table 1. Study characteristics (N=172), n (%).

Characteristic	Value
Reported as a randomized pilot study	
No	115 (66.1)
Yes	59 (33.9)
Region/country (number of studies per country)	
North America (USA-69, Canada-6, Mexico-1)	76 (43.7)
Europe (UK-8, Spain, Sweden, Denmark, Italy, Netherlands)	49 (28.2)
Asia (Korea-8, China-8, Japan-3, Taiwan-2)	26 (14.9)
Oceania (Australia-16, New Zealand-3)	19 (10.9)
Middle East, North Africa, Greater Arabia (Israel-2)	2 (1.2)
Sub-Saharan Africa (Ghana)	1 (0.6)
Multi-country	1 (0.6)
Participant condition	
Chronic disease (Diabetes-28, Cardiovascular Disease-13, Pulmonary disease-9, HIV-4)	55 (33.2)
Mental health (eg, depression, anxiety)	48 (27.8)
Overweight/obese/physical inactivity/diet	47 (27.2)
Substance use/abuse (eg, alcohol, nicotine, drugs)	10 (5.8)
Cancer	7 (4.1)
Neurological/musculoskeletal (eg, back pain, arthritis)	7 (4.1)
Behavior assessed (health outcome)	
Nutrition and physical activity (eg, weight loss)	61 (34.7)
Mental health management (eg, reduce depressive symptoms)	31 (17.9)
Diabetes management (eg, hemoglobin A_{1c} reduction)	22 (12.7)
General health/well-being (eg, quality of life, sleep quality)	20 (11.7)
Medication adherence	15 (8.7)
Cessation/harm reduction (eg, days no drug/alcohol use, smoking cessation)	13 (7.5)
Management of other chronic diseases (eg, reduce blood pressure)	12 (6.9)
Participant age category	
Adult	145 (83.2)
Mix of age groups	18 (10.3)
Adolescent	6 (3.5)
Pediatric	3 (1.7)
Older adult	2 (1.2)
Comparator group	
Usual/standard of care	102 (58.6)
Attention control	45 (25.9)
Waitlist control	27 (15.5)
App developed for research or clinically/commercially available	
Research	87 (50.5)
Commercial/clinical	77 (44.3)
Unclear	10 (5.7)
App intervention type	
Stand-alone intervention	108 (62.1)

Characteristic	Value
Part of a larger intervention	66 (37.9)
App+in person	22 (12.6)
App+wearable	19 (10.9)
App+2 other (eg, phone call+wearable, in person+patient portal)	15 (8.6)
App+1 other (eg, support specialist, phone call, text messaging)	10 (5.7)
Intervention frequency	
Daily/as needed	125 (71.8)
Weekly	18 (10.3)
Unclear	25 (14.4)
Other	6 (3.4)
Follow-up time (extracted last follow-up time)	
<1 month	10 (5.7)
1-6 months	135 (77.6)
7-12 months	25 (13.8)
>12 months	4 (2.3)
Reported use of behavior change theory for app development	
Yes	68 (39.1)
No/not clearly reported	106 (60.9)
Reported a power analysis	
Yes	103 (59.2)
No/not clearly reported	71 (40.8)
Reported intention-to-treat analysis	
Yes	96 (55.2)
No/not clearly reported	78 (44.8)
Level of personalization	
One-size-fits-all	74 (43.0)
Static (one-time tailoring to individual)	23 (13.4)
Dynamic (adaption occurs periodically or in real time during intervention)	75 (43.6)

Figure 2. Publication of included studies per year.



Participants

Sample sizes ranged from 6 to 14,228, with a total of 53,331 participants across the 172 studies. Overall, 83.1% (143/172) of studies targeted adults, whereas 8.1% (14/172) targeted exclusively pediatric patients or adolescents, and 0.6% (1/172) of the studies targeted older adults. The most common conditions of the participants were chronic diseases (55/172, 31.9%), mental health disorders (48/172, 27.9%), and overweight or obesity (45/172, 26.2%).

Methodology

Control groups either received no intervention (100/172, 58.1%), received attention control (eg, basic version of an app; 45/172, 26.2%), or were waitlisted (27/162, 15.7%). The majority had follow-up periods of 6 months or less (143/172, 83.1%), and only 2.3% (4/172) of studies had follow-up periods longer than 12 months. Over half reported a sample size calculation or power analysis (103/172, 59.9%) or an intention-to-treat analysis (96/172, 55.8%).

Interventions

Over half of the interventions were delivered by a stand-alone app (106/172, 61.6%), and 38.4% (66/172) used an app as a component of a larger intervention. For example, 22 evaluated an app and in-person set-up (22/172, 12.8%),19 apps were paired with a wearable (19/172, 11.0%),15 evaluated an app with 2 other interventions (15/172, 8.7%), and 5.8% (10/172) apps had one other component that was neither in-person nor wearable (eg, text messages, mHealth support specialist). In total, 67 of the studies (67/172, 39.0%) reported that their intervention was based on a behavior change theory (eg, Social Cognitive Theory). One-size-fits-all (74/172, 43.0%) and dynamic (75/172, 43.6%) were the most common levels of personalization. Intervention frequency was mostly daily or multiple times per day (123/172, 71.5%). The studies assessed interventions for a range of behavior change outcomes. The top 3 were nutrition, physical activity, or both (60/172, 34.9%), mental health disorder management (34/172, 19.8%), and diabetes management (21/172, 12.2%).

App Features

The intervention apps included a mean of 5 features (SD 2), ranging from 1 to 11 (Figure 3). The most common features included self-report adherence or self-monitoring (120/172, 69.8%), visual feedback on user data (109/172, 63.4%), and information/education (107/172, 62.2%; Figure 4). The least common features were communication messaging within app (36/172, 20.9%), app-based social support (35/172, 20.3%), and gamification (22/172, 12.8%). There were no features common to all intervention apps. The features corresponded to one or more behavior change techniques (Multimedia Appendix 2). The most common behavior change mechanisms of the app features were feedback and monitoring (corresponding to 5 out of 13 features) and shaping knowledge (corresponding to 2 out of the 13 features).

Figure 3. Number of studies with app features and frequency of app features.

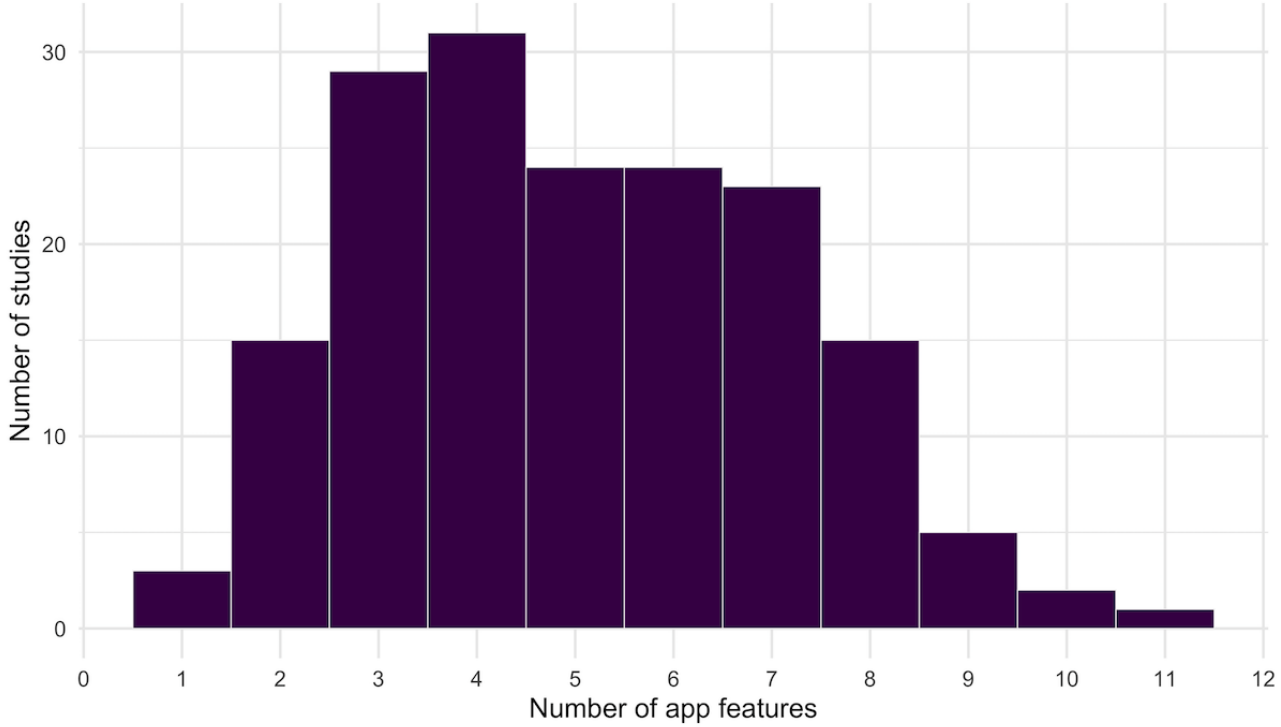
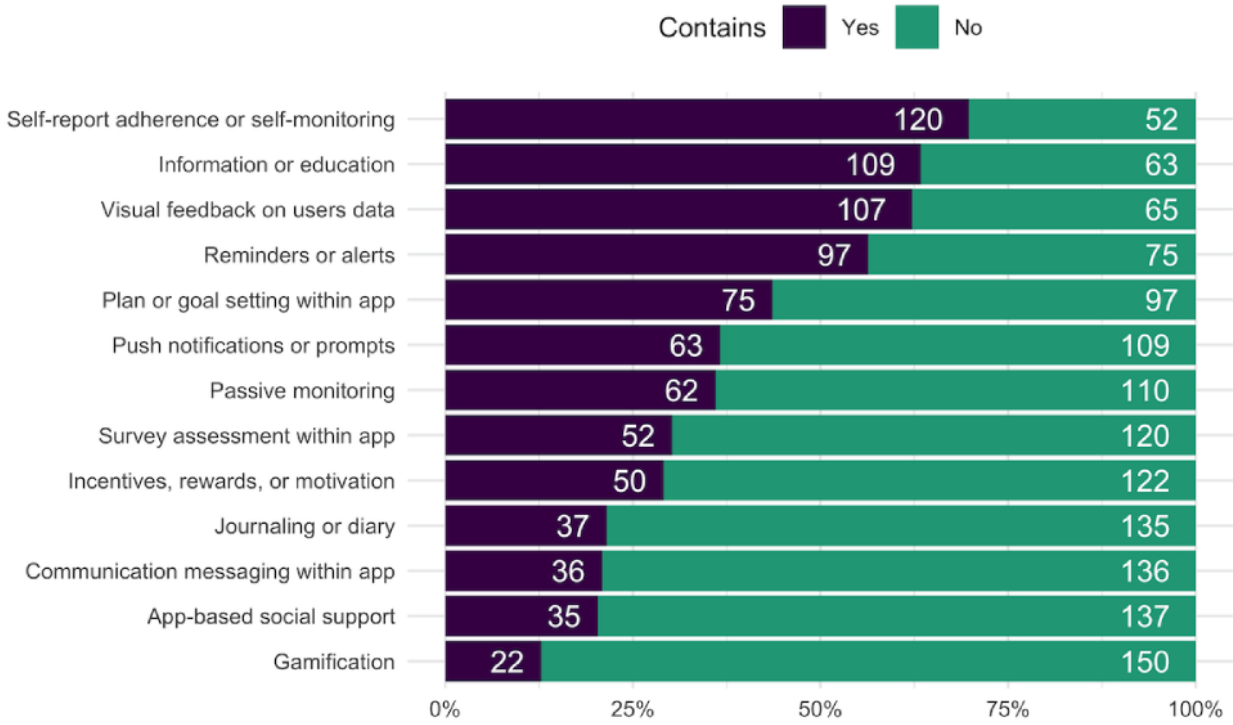


Figure 4. Frequency of app features.

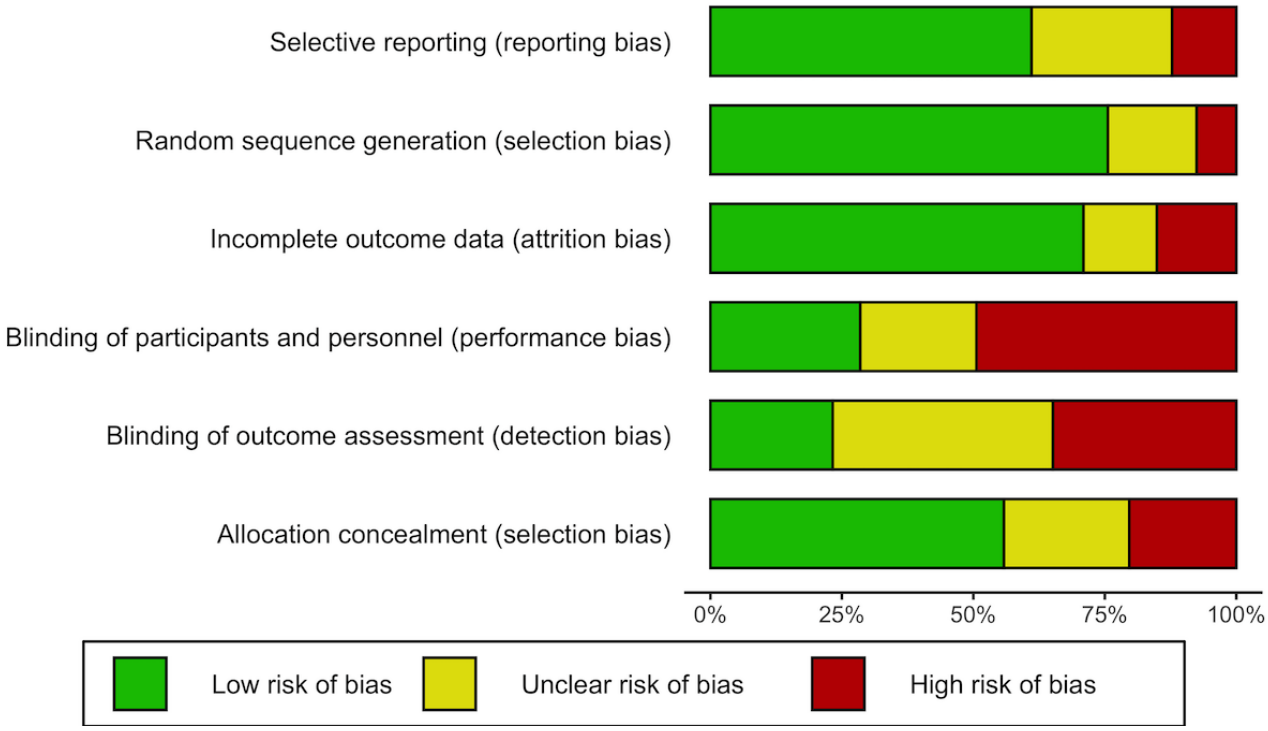


Risk of Bias

All studies were screened for risk of bias using the Cochrane Risk of Bias tool for RCTs (Figure 5). The randomization procedure (random sequence generation) for most of the studies was considered adequately described (130/172, 75.6% low risk for bias). Just over half of studies reported details of allocation concealment (96/172, 55.8% had low risk of bias), whereas blinding of participants/personnel was less well described (49/172, 28.5% had low risk of bias). Few studies reported blinding of outcomes (40/172, 23.3%). Most studies had low

risk of incomplete outcome data (122/172, 70.9%) and low risk of reporting bias (105/172, 61.0%).

Figure 5. Risk of Bias.

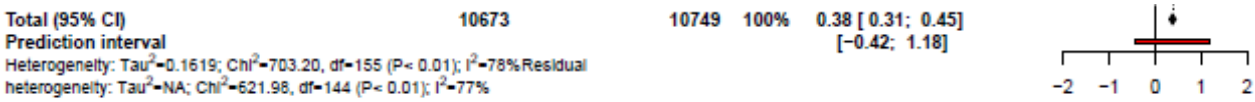


Meta-analysis

Overall Effect of Studies With Continuous Outcomes

A total of 156 studies involving 21,422 participants reported continuous health outcome data. The use of a behavior change app (either as a stand-alone or as part of a larger intervention) confers a slight or weak advantage over standard care in health interventions (SMD=0.38 [95% CI 0.31-0.45]; I^2=80%; Figure 6; Multimedia Appendices 3 and 4). Excluding pilot studies, there were a total of 105 studies involving 18,514 participants. The overall results were similar (SMD=0.35 [95% CI 0.27-0.43]; I^2=82%).

Figure 6. Effect of studies with continuous outcomes.



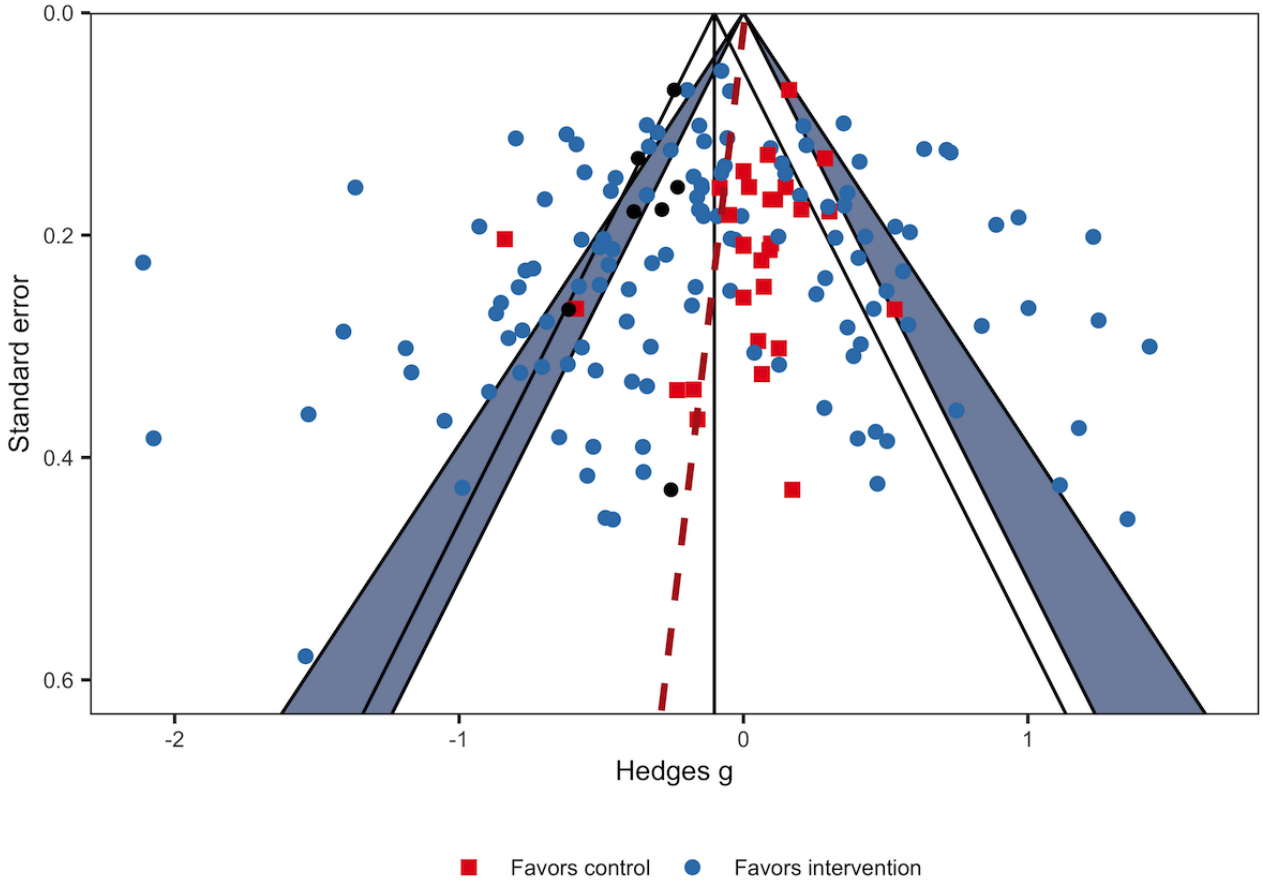
Subanalyses of Studies With Continuous Outcomes

We analyzed the impact of the app level of personalization as stand-alone or as part of larger intervention and length of follow-up time. No clear pattern of relationship emerged in addition to the positive but weak advantage over standard care in health interventions for all groupings (Multimedia Appendix 3). Two groupings (>6 months, one-size-fits-all/static, part of larger intervention; <3 months/dynamic/stand-alone) had larger effect sizes, with SMD of 0.50 and 0.59, respectively. Excluding pilot studies in the analysis did not change the results.

Publication Bias

Figure 7 presents a funnel plot assessing publication bias for continuous outcomes. The trim-and-fill process was used to adjust for funnel plot asymmetry and estimate where studies might be without publication bias (black dots in Figure 7).

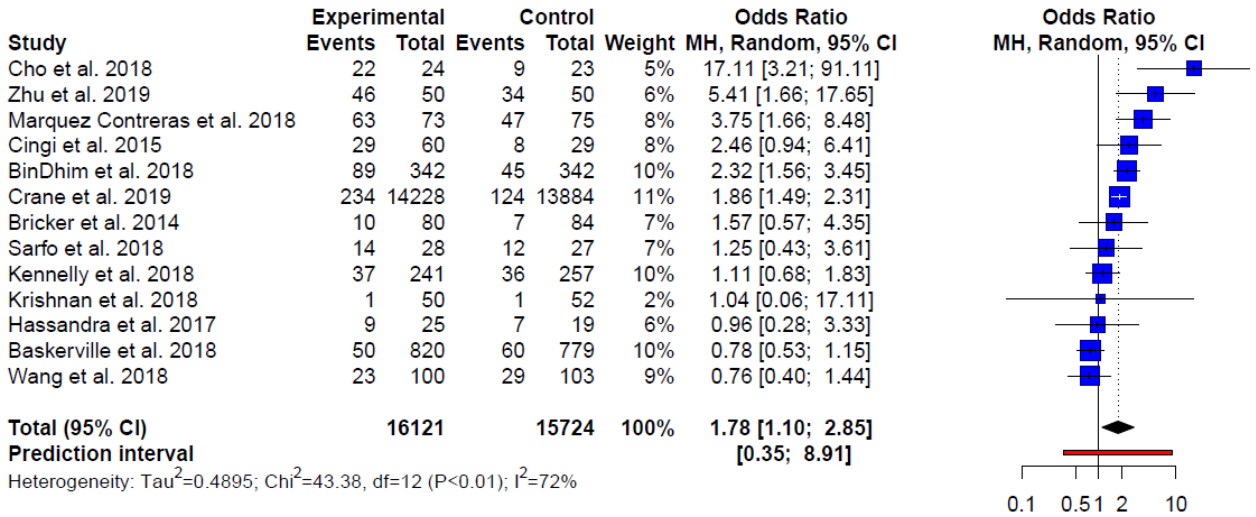
Figure 7. Funnel plot.



Overall Effect of Studies With Binary Results

A total of 13 studies involving 31,845 participants reported binary health outcome data. The use of a behavior change app (either as a stand-alone or as part of a larger intervention) confers a small advantage over standard care in health interventions with an odds ratio of 1.78 (95% CI 1.10-2.85; Figure 8).

Figure 8. Effect of studies with binary outcomes.



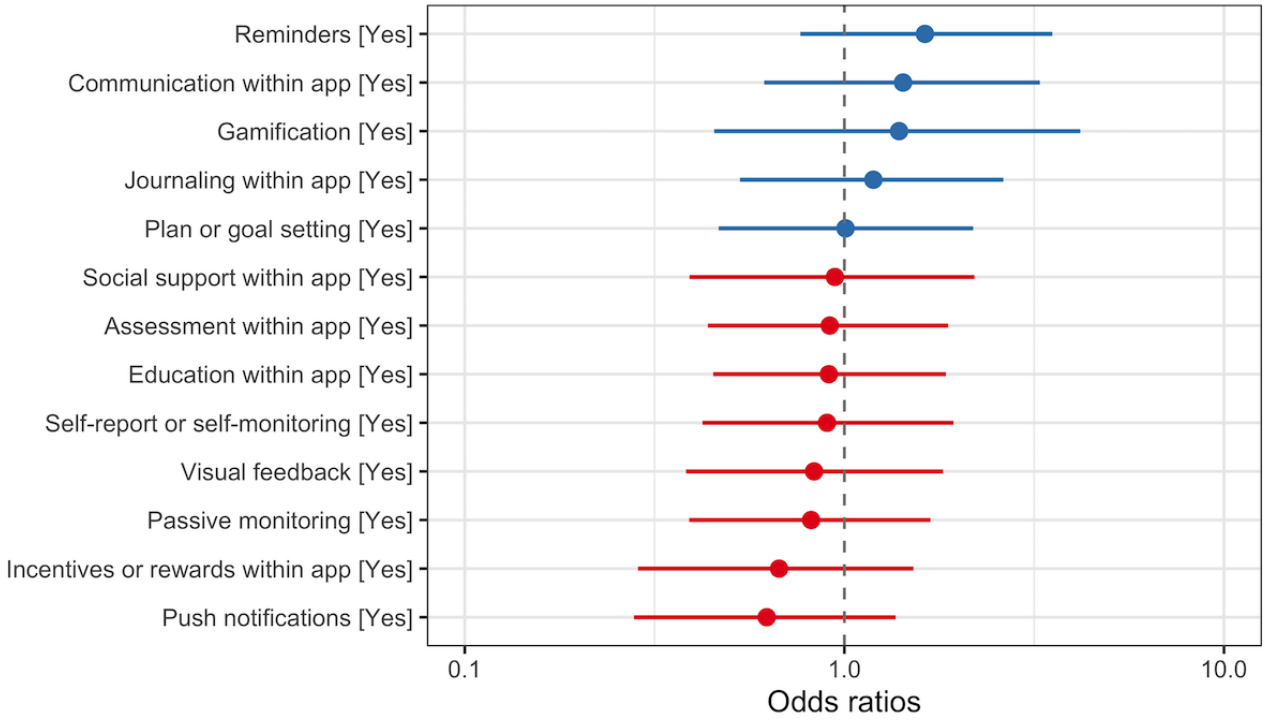
Due to the small number of this group of studies, no subanalysis was conducted. We also did not assess publication bias in this small group of studies.

Analysis of App Features of Successful Apps

Results of logistic regression showed a slight positive effect on health outcomes for the features of interactive communication,

reminders, gamification, and journaling, although no feature had significant effects on health outcomes (Figure 9).

Figure 9. Effect of app features on outcome.



Discussion

Principal Findings

In this comprehensive systematic review and meta-analysis, we synthesized the evidence from 172 studies, which included 53,331 participants, to describe the overall effect of mHealth apps used in interventions to change behavior to promote health and manage disease and also explored the impact of their features. To the best of our knowledge, this is the first meta-analysis to assess the effectiveness of health apps across a variety of diseases and health outcomes. Other related systematic reviews focused on methodologic reporting of any mHealth-based intervention (not app specific) [7], reported on specific diseases or health outcomes [12,15,18-21], explored features (not a systematic review) [34], or did not extract data for a meta-analysis [35,36].

Our results highlight a substantial growth in the number of health apps containing behavior change techniques being rigorously evaluated. In addition to the studies presented, 464 published protocols likely to meet the inclusion criteria were identified in the search, suggesting that the number of app-related RCT publications will likely continue to rise. The results from this meta-analysis suggest that app-based behavior change interventions have a positive, although weak, effect on health outcomes.

It is thought that mobile tools, such as apps, can significantly improve the provision of health care services and communication, particularly in low- and middle-income countries (LMICs) with a shortage of health care workers but where access to phones and cellular services is high even among vulnerable populations [37]. However, because the vast majority of studies included in this systematic review were conducted in high-income countries, the findings cannot be generalized to LMICs. Similarly, other systematic reviews of mHealth literature found disproportionate testing of mHealth tools, including apps, in high-income countries compared with LMICs [12,37-39].

Quality of Selected Studies

We found the current evidence to be largely composed of studies reporting rigorous methodologies that included power calculations, intention-to-treat analysis, and a low risk of bias for randomization and randomization concealment. Most of the studies were not classified as pilot RCTs, but even with RCT pilot studies removed, results were largely unchanged. Randomized trials, if sufficiently large, provide the most convincing evidence about the effects of interventions because randomization should result in both observed and unobserved baseline characteristics being similar across the compared groups [27]. The high risk of bias for intervention blinding and blinding of outcomes arises from the near impossibility for patients and health care professionals to be unaware of the use of apps and smartphones in the care process. About 40% of the studies chose to use attention control or waitlist as comparators. Having an active comparator (eg, attention control) could ensure that all participants in a trial will not be knowingly disadvantaged and may reduce bias attributed to inadequate blinding [35].

Overall Effect

Our finding of an overall positive but weak effect appears to be consistent with the few systematic reviews of app-based

interventions that included meta-analyses. Systematic reviews and meta-analyses of apps targeting diabetes [12,21,40] and asthma control [18] found effects favoring app-based treatment. The weak effect could be due, in part, to high heterogeneity among the health outcomes. However, even the targeted reviews assessing the efficacy of apps for lifestyle modification in diabetes reported high heterogeneity of up to 86% in subcategory analysis [21,40], whereas the meta-analysis of apps to support asthma management included only 3 studies [18]. Behavior change interventions are inherently complex, and it is uncommon for any 2 interventions to evaluate exactly the same intervention [21]. Moreover, outcomes may be measured in different ways and at different time points across studies, which could further blur any differences in outcomes. We expected heterogeneity given the nature of behavior change interventions and the diversity of the disease outcomes studied. Nonetheless, the overall meta-analysis results in this study should be interpreted with caution because of the high heterogeneity found. Another contributing factor to the weak positive effect could be that a few studies reported the use of a behavior change theory to guide app development, which is recommended to improve the impact of an intervention [41]. Similarly, other researchers who conducted a systematic review of health apps noted a lack of mention by study authors of the use of a behavioral model or theory for the development of apps [42,43]. Along these same lines, we did not assess the contributions of patients and health care professionals in the development of app-based interventions, even though this factor is recognized as important [44]. It is well established that usability considerations along with the use of sociotechnical design principles and a holistic approach to behavior change can impact efficacy [42]. The involvement of patients and health care professionals in app development was not consistently mentioned by the authors of the studies included in this review. Finally, publication bias could undermine even the weak effect found. Negative or null findings are less often published or reported [45], and in this review, there appears to be publication bias toward studies that showed an intervention effect. However, funnel plots are not necessarily trustworthy under high heterogeneity. Nevertheless, as health behavior change interventions are resource intensive and often difficult to scale, having an equivalent or comparable outcome may still provide wider reach, cost savings, and added convenience to the users [2].

Follow-Up Duration, Intervention Type, and Level of Personalization

We conducted sensitivity analysis on the length of follow-up time, the intervention as a stand-alone app or as part of a larger intervention, and the level of personalization. Sensitivity analyses suggested that the effects did not differ consistently based on these 3 attributes. We were concerned that the impact on effect could be skewed because the majority of the studies had follow-up times of less than 6 months. Although there was a slight trend for higher effects at lower follow-up times (less than 3 months), this was not consistent across all subanalyses groupings. Researchers have questioned the ability of app-based interventions to sustain beneficial health effects over time [40]. In particular, decline in app usage over time has been reported and could impact long-term effects [31]. In a systematic review of technology-driven behavior change techniques, researchers found that 63% were effective in the short term (<3 months), whereas only 33% were effective for long term (≥12 months) [46]. As most of the included studies targeted chronic conditions, more time may be needed to achieve a sustained behavior change that will lead to lasting changes in health outcomes. Our results are consistent with other research findings that short duration of follow-up was a limitation of many trials [7,11,13]. Another noted limitation of mHealth behavior change interventions is the lack of information on the impact of long-term intervention on patient-important primary outcomes [7]. Therefore, future intervention studies incorporating behavior change apps should increase follow-up times to enable assessment of primary outcomes as well as better describe app engagement overtime.

Regarding intervention type, it appeared that there was little impact on the overall effect if the intervention was provided as a stand-alone app or if the app was part of a larger intervention. This is unsurprising as behavior change interventions are typically complex, include multiple components, and may impact individuals in a variety of ways [7]. In the studies in which the app formed part of a larger intervention, other components included, for example, occasional in-person meetings or phone calls, a wearable device, or a combination. As the stand-alone apps had comparable results with interventions that included other components, this may suggest that apps can support a comprehensive intervention.

The level of personalization, either one-size-fits-all or static or dynamic, did not appear to alter the overall effect of outcomes. This runs counter to recommendations to offer levels of personalization rather than a one-size-fits-all functionality to improve outcome [47]. For instance, a systematic review of app features for diabetic support concluded that personalized and tailored empowerment features should be included in commercial apps for large-scale assessment of the self-management of the disease [48].

Contribution of Features

In this study, the number of features present in the app did not appear to confer an advantage. This contrasts with other researchers who reported increased efficacy with more features [12,46]. The impact of the features on health-related outcomes was inconsistent [18]. It has been proposed that an increase in the number of features could decrease the usability of an app [44]. Usability is considered critical for engagement and adherence to app interventions over time [49]. Further exploration of the complex interaction between the number of features and usability may be important for determining future app efficacy. In addition, future research could focus on better understanding of mechanisms of action of behavior change techniques within health app–based interventions to assess which combinations lead to improved health outcomes. Our aim was not to specifically detect which behavior change techniques could be the mechanism of action for the potential changes in health outcomes. The authors of app-based studies should clearly describe the behavior change techniques used in their app components and the guiding theories. Applying

taxonomy of behavior change techniques by Michie et al [24] could be used to help compare and contrast findings across studies. However, these taxonomies were not developed for coding app features; therefore, some concepts may not be fully transferable [5]. Features represent a high level of abstraction of requirements and are useful to describe the functionality of a new system without the need to drill down into too much detail or are later specific enough for implementation [50]. Research is needed to better tailor and align behavior change taxonomies for mobile app feature apps.

When individual features were assessed by health outcome effectiveness, there appeared to be a slight advantage among those apps using interactive communication, reminders, gamification, and journaling. Interactive communication has been found to be important in a number of studies. Among individuals with diabetes, including interactive communication in apps and having remote access to health professionals were associated with greater effectiveness in reducing hemoglobin A_{1C} levels [12]. Similarly, two-way interactivity has been shown to improve adherence [51]. Furthermore, smartphone-based interventions that did not include interactive support saw a decrease in app use/engagement among those with chronic respiratory diseases, diabetes, and hypertension [34]. Reminders have also been identified as a core component of mobile interventions, most notably through the use of text messaging interventions [52].

Gamification conferred an advantage even though it was the least frequent feature type. Identifying few apps that employed gamification is consistent with findings from other studies, in which authors reported that this lack of use may limit the potential to improve health outcomes [53,54]. Journaling as a feature was not mentioned in other systematic reviews of features; thus, it may be a new finding or may have been labeled differently. In any type of technology-driven intervention for diabetes management, those containing digital features that facilitated health and lifestyle education, behavior or outcome tracking, and/or web-based health coaching were most effective [46]. In this review, we identified many studies that included education, self-tracking, coaching, and goal setting; however, they did not show a clear effect on health outcomes.

Strengths and Limitations

A strength of this study is that it is a comprehensive review of rigorously tested app-delivered behavior change interventions for a variety of diseases and health outcomes. Understanding the impact of app-delivered interventions and features is important given their rapid growth in the health care market. There are also limitations to consider. First, although we attempted to focus on stand-alone app interventions, some of the studies included other elements with the apps used as one component of a larger intervention; thus, it was difficult to isolate the effects of app-based interventions. We attempted to mitigate this limitation by conducting subanalyses with stand-alone apps. Second, we did not extract data on app usage. Usage data could elucidate whether the intervention was less effective because of lack of use or exposure to the intervention. However, these data were not consistently reported in the studies. Finally, there was often a lack of clear reporting of app features, which made it challenging to extract data. If the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) guideline continues to be adopted for mHealth research, reporting quality should begin to improve [55,56].

Future Recommendations

Future mHealth studies should use standardized reporting guidelines and describe their intervention and app features with adequate detail so that results are reproducible [56]. Future research is needed to develop a taxonomy of behavior change techniques that aligns with all potential app features. These efforts have begun with focus on specific app types, such as measuring physical activity [57,58] and reducing alcohol consumption [59]. Follow-up times should be appropriate for disease and condition and likely should be longer than what we found in this study, where the majority of papers reported follow-up durations of less than 6 months. Updates to this review can be conducted to assess progress as the published protocols are completed and reported. In addition to understanding clinical effectiveness and cost-effectiveness, future work will also need to include the potential payer's perspective to lessen administrative burden, improve workflow, enhance patient and provider engagement, and improve quality of care while lowering costs [16]. Adoption or adherence to the technology should also be evaluated.

Conclusions

Rigorous studies to examine the effectiveness of behavior change app interventions on health outcomes are increasing. The results of our meta-analysis suggest that apps have a positive but weak effect on health outcomes and may be a useful adjunct in behavior change health interventions. There was insufficient evidence to make recommendations on the essential number of base features to include in apps. There is a clear need for rigorous testing of behavior change apps in LMICs where there may be added challenges of lack of human resources and access to health care services. Future research should clearly report app features, evaluate long-term effectiveness to modify health outcomes, and consider attention control comparators. In addition, negative or null findings need to be reported. Although not explored in this review, including analyses of level of app engagement is needed to better determine the actual effect of apps on outcomes and to explore the specific features that promote patient engagement with the app and adherence to the intervention.

Acknowledgments

SI was supported, in part, by the National Institute of Nursing Research, K23NR017210 (principal investigator: SI) and the National Institute of Allergies and Infectious Disease, R01AI147129 (multiple primary investigator: SI, Rubinstein). KK was

supported, in part, by the National Institutes of Nursing Research, T32NR014833 and the National Institute of Diabetes and Digestive and Kidney Diseases, T32DK007742. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We appreciate Kirsten Kincaid and Hannah Erdy for their assistance in the process of this research.

Authors' Contributions

SI developed the initial protocol. ES developed and executed the search strategies. SI, KK, TA, and YK conducted screening at all levels and data extraction. DB developed the analysis protocol and conducted the analysis and meta-analyses. All authors contributed to manuscript drafting.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE (Ovid) Search Strategy.
[DOCX File , 14 KB - mhealth_v9i1e21563_app1.docx]

Multimedia Appendix 2
App features and definitions.
[DOCX File , 25 KB - mhealth_v9i1e21563_app2.docx]

Multimedia Appendix 3
Effects of studies with continuous outcomes by level of personalization, intervention type, and length of outcome follow-up.
[PDF File (Adobe PDF File), 65 KB - mhealth_v9i1e21563_app3.pdf]

Multimedia Appendix 4
List of references cited in Multimedia Appendix 3.
[DOCX File , 45 KB - mhealth_v9i1e21563_app4.docx]

References

1. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
2. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
3. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
4. McKay FH, Slykerman S, Dunn M. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
5. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
6. Haase J, Farris KB, Dorsch MP. Mobile applications to improve medication adherence. Telemed J E Health 2017 Feb;23(2):75-79. [doi: 10.1089/tmj.2015.0227] [Medline: 27248315]
7. Oikonomidi T, Vivot A, Tran V, Riveros C, Robin E, Ravaud P. A methodologic systematic review of mobile health behavior change randomized trials. Am J Prev Med 2019 Dec;57(6):836-843. [doi: 10.1016/j.amepre.2019.07.008] [Medline: 31753266]
8. Pohl M. 325,000 mobile health apps available in 2017 – Android now the leading mHealth platform. Research2guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2018-08-20]
9. mHealth Apps Market Size Worth $236.0 Billion by 2026. Grand View Research. 2019. URL: https://www.grandviewresearch.com/press-release/global-mhealth-app-market [accessed 2020-01-19]
10. Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019;7(8):2583-2591 [FREE Full text] [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
11. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 1;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]

12. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 1;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
13. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
14. 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]
15. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
16. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676 [FREE Full text] [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
17. Fougerouse P, Yasini M, Marchand G, Aalami OO. A cross-sectional study of prominent US mobile health applications: evaluating the current landscape. AMIA Annu Symp Proc 2017;2017:715-723 [FREE Full text] [Medline: 29854137]
18. Hui C, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 1;24(3):619-632 [FREE Full text] [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
19. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [Medline: 28456072]
20. 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]
21. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. Heldenbrand S, Martin BC, Gubbins PO, Hadden K, Renna C, Shilling R, et al. Assessment of medication adherence app features, functionality, and health literacy level and the creation of a searchable web-based adherence app resource for health care professionals and patients. J Am Pharm Assoc (2003) 2016;56(3):293-302. [doi: 10.1016/j.japh.2015.12.014] [Medline: 27067551]
24. 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]
25. Ng R, Carter SR, El-Den S. The impact of mobile applications on medication adherence: a systematic review. Transl Behav Med 2019 Aug 6:- epub ahead of print. [doi: 10.1093/tbm/ibz125] [Medline: 31384950]
26. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
27. Higgins JP, López-López JA, Becker BJ, Davies SR, Dawson S, Grimshaw JM, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. BMJ Glob Health 2019;4(Suppl 1):e000858 [FREE Full text] [doi: 10.1136/bmjgh-2018-000858] [Medline: 30775014]
28. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
29. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. New Jersey, United States: Wiley; 2006.
30. World Bank Country and Lending Groups. World Bank Data Help Desk. 2020 Feb 10. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-01-20]
31. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
32. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. Stat Med 2007 Apr 30;26(9):1964-1981. [doi: 10.1002/sim.2688] [Medline: 16955539]
33. Higgins J, Li T, Deeks J. Chapter 6: Choosing effect measures and computing estimates of effect. Cochrane Training. 2020. URL: https://training.cochrane.org/handbook/current/chapter-06 [accessed 2020-12-16]

34. Donevant S, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 1;25(10):1407-1418 [FREE Full text] [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
35. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
36. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
37. Kruse C, Betancourt J, Ortiz S, Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 9;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
38. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
39. Abaza H, Marschollek M. Mhealth application areas and technology combinations. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 8;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
40. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
41. Georgsson M. Improving the uptake and continued use of mobile health technology: From behavioral change theory to contextual intervention design and the inclusion of the end-user. Online Journal of Nursing Informatics 2019;23(1) [FREE Full text]
42. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 1;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
43. Carey R, Connell LE, Johnston M, Rothman A, de Bruin M, Kelly MP, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]
44. 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. [doi: 10.2196/jmir.2968] [Medline: 24718852]
45. Page M, Higgins J, Sterne J. Chapter 13: Assessing risk of bias due to missing results in a synthesis. Cochrane Training. URL: https://training.cochrane.org/handbook/current/chapter-13#section-13-3-5-3 [accessed 2020-03-20]
46. van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
47. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
48. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237 [FREE Full text] [doi: 10.2196/12237] [Medline: 30463839]
49. Brown W, Yen P, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
50. Leffingwell D, Widrig D. Managing software requirements : a unified approach. Boston: Addison-Wesley Longman Publishing Co., Inc; 2000.
51. Wald DS, Butt S, Bestwick JP. One-way versus two-way text messaging on improving medication adherence: meta-analysis of randomized trials. Am J Med 2015 Oct;128(10):1139.e1-1139.e5. [doi: 10.1016/j.amjmed.2015.05.035] [Medline: 26087045]
52. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
53. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
54. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 4;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
55. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]

56. 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]
57. Conroy DE, Yang C, 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]
58. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
59. Garnett C, Crane D, West R, Brown J, Michie S. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. JMIR Mhealth Uhealth 2015 Jun 29;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]

Abbreviations

LMIC: low- and middle-income country
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach, L Buis; submitted 18.06.20; peer-reviewed by J Wilroy, O El-Gayar; comments to author 22.08.20; revised version received 13.10.20; accepted 28.10.20; published 11.01.21.

Please cite as:
Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E
Effectiveness of Mobile Apps to Promote Health and Manage Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2021;9(1):e21563
URL: https://mhealth.jmir.org/2021/1/e21563
doi:10.2196/21563
PMID:33427672

Original Paper

Feasibility, Usability, and Enjoyment of a Home-Based Exercise Program Delivered via an Exercise App for Musculoskeletal Health in Community-Dwelling Older Adults: Short-term Prospective Pilot Study

Robin M Daly[1], PhD; Jenny Gianoudis[1], PhD; Travis Hall[1], MClinExPhys; Niamh L Mundell[1], MClinExPhys; Ralph Maddison[1], PhD

Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia

Corresponding Author:
Robin M Daly, PhD
Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Highway, Burwood
Melbourne, 3125
Australia
Phone: 61 392446040
Email: rmdaly@deakin.edu.au

Abstract

Background: Many older adults choose and prefer to exercise at home, but to attain the greatest benefits, the correct type and dose of exercise should be prescribed and adherence maintained. Advances in digital health technologies now provide the opportunity for exercise professionals to deliver and monitor personalized, evidence-based exercise programs to anyone at any time.

Objective: The aim of this study was to evaluate the feasibility, usability, and enjoyment of a web-based exercise prescription app as a platform for exercise professionals to remotely deliver and monitor an individually tailored, home-based multicomponent exercise program (delivered through tablet computers) to older adults living independently in the community.

Methods: This was an 8-week, prospective single-arm pilot study in 20 adults aged ≥65 years living independently in the community: 10 owned a tablet computer (tablet owners) and 10 did not own tablets (tablet nonowners). All participants were prescribed a home-based, muscle strengthening, weight-bearing impact and challenging balance/mobility program (3 days/week) using a commercial exercise prescription app on a tablet computer. Study endpoints were feasibility (retention, adherence, adverse events), usability (System Usability Scale), physical activity enjoyment (Physical Activity Enjoyment Scale), changes in lower extremity function (Short Physical Performance Battery [SPPB]), and level of physical activity (questionnaire). Process measures related to the participants' experiences and perceptions of the exercise program and web-based app were also included.

Results: A total of 19 participants (mean age, 70 years) completed the study (19/20, 95%), and mean adherence to the exercise program was 84% (95% CI 70%-97%). There were 2 minor adverse events in 2 participants from 401 completed sessions. Mean weekly walking time increased by 78 minutes (95% CI 0-156, *P*=.049) and moderate-to-vigorous physical activity time by 41 minutes (95% CI –8 to 90, *P*=.09). For SPPB scores, there was a 0.3 point (95% CI –0.1 to 0.7, *P*=.17) modest sized (effect size, *d*=0.42) improvement after 8 weeks. Mean (SD) system usability was high (86 [10] with 100 best imaginable). There was no change in the overall physical activity enjoyment scores after 8 weeks, but participants reported that they enjoyed using the web-based exercise app and the exercise program (median score 4 on a 5-point Likert scale). For all measures, there were no differences between previous tablet owners and nonowners.

Conclusions: This pilot feasibility study indicates that it is safe and feasible for community-dwelling older adults to participate in a home-based, multicomponent exercise program targeting musculoskeletal health and function that was delivered and monitored remotely by exercise professionals using a tablet-based exercise prescription app.

(JMIR Mhealth Uhealth 2021;9(1):e21094) doi:10.2196/21094

KEYWORDS

home exercise; multicomponent exercise; mobile health; musculoskeletal; adherence; usability; older adults; physical activity enjoyment

Introduction

Every year, around 30% of people aged over 65 years living in the community fall at least once [1], and falls are a leading cause of fragility fractures, injury-related hospitalization, mortality, and health care costs for older adults [2,3]. By 2022, it is predicted that there will be 1 fragility fracture every 2.9 minutes (>500 per day) in Australia [4]. The current models of care for fracture risk reduction focus largely on pharmacological agents targeting bone mineral density (BMD) [5], but there is a need for multifaceted approaches that can simultaneously target multiple fall and fracture risk factors.

Exercise is widely recognized as a safe and effective approach to improve nearly all modifiable fracture risk factors, including BMD and falls risk [6,7]. Several meta-analyses of randomized controlled trials provide compelling evidence to support the benefits of exercise as a single intervention to prevent falls in community-dwelling older people [7] and multicomponent resistance-based exercise programs for improving bone health in postmenopausal women [8]. The findings from several of our previous randomized controlled trials conducted within community-based health and fitness centers have also shown that multicomponent exercise programs incorporating progressive resistance training combined with weight-bearing impact and challenging balance and mobility training are safe and effective for improving hip and lumbar spine BMD, muscle mass, strength, power, and function in healthy older adults and those with low BMD or at increased falls risk [9-12]. Despite these positive findings, geographical location and access to affordable community-based exercise programs and qualified exercise trainers and a general aversion to the gym environment are key barriers to participation reported by many older people, which has implications for intervention effectiveness [13-15]. Thus, there is a need to consider alternative models of service delivery to meet individuals' exercise needs, preferences, and financial resources more broadly.

Advances in digital health apps have provided new opportunities for health care professionals to remotely deliver and monitor evidence-based exercise programs tailored to the needs of older adults and within their own home or community environment. This is important, as a study of 240 community-dwelling adults attending osteoporosis-related programs revealed that a lack of access to exercise programs that meet their needs and preferences and limited resources, time, and trust in exercise providers were some of the key barriers to participation [13]. For falls prevention, a study in 5440 older adults indicated that a home-based strength and balance training program was preferred over other prevention strategies [16]. However, to ensure clinical effectiveness, it is important that any prescribed exercise programs adhere to current best practice guidelines and incorporate behavioral strategies to promote long-term adherence. Despite the rapid rise in the number of web-based and mobile health apps available to health care professionals to deliver exercise programs to people at home, few studies have evaluated the feasibility, usability, and enjoyment of exercise prescription apps targeting the key musculoskeletal health and function (eg, bone and fall-related) risk factors associated with fractures in older adults residing in the community.

The aim of this feasibility study was to evaluate the feasibility (retention, adherence and adverse events), usability, and enjoyment of a commercial web-based exercise prescription app (Physitrack) as a platform for exercise professionals to remotely deliver and monitor an individually tailored, home-based multicomponent exercise program (via tablet computers) for older adults living independently in the community. In addition, we explored participants' perceptions of the exercise program and Physitrack app and whether outcomes differed between previous tablet computer owners and nonowners.

Methods

Study Design

The Seniors Made Active thRough Technology (SMART) study was an 8-week community-based, prospective single-arm pilot study in which adults aged ≥65 years were prescribed a home exercise program (accessed by the PhysiApp-patient portal) by an accredited exercise physiologist (AEP) using the commercial Physitrack (clinician portal) exercise prescription app. The trial was managed through the Institute for Physical Activity and Nutrition at Deakin University, Australia, and was approved by the Deakin University Human Research Ethics Committee (HREC 2016-219).

Participants and Recruitment

Twenty relatively healthy men and women (convenience sample) aged 65 years and over living independently in the community were recruited through our research trial database, and study flyers were sent to a number of community (Rotary) clubs in the eastern suburbs of Melbourne, Victoria, Australia. In order to explore differences in the outcomes and the experiences in using mobile technology between those with tablets (tablet owners) and without tablets (tablet nonowners), we deliberately recruited 10 participants who possessed a tablet computer (and had access to Wi-Fi at home) and 10 participants who did not possess such a device. The tablet nonowners were provided with an iPad and a SIM card (and adequate data capacity) for the duration of the study and instructed on how to use it during the initial home visit.

Participants were initially screened over the telephone and included if they were able to walk without the use of an aid, willing to use an iPad (their own or one provided to them) for the execution of the exercise program, and if they were able to speak English. Participants were excluded based on the following criteria (all self-reported): (1) aged <65 years, (2) participation in resistance exercise >1 session per week for at least 20 minutes or moderate-to-vigorous intensity physical

activity for ≥150 min/week over the past 3 months, (3) recent low trauma fracture (within the past 6 months), (4) inability to stand unaided, (5) acute or terminal illness likely to compromise exercise participation, (6) unstable or ongoing cardiovascular/respiratory disorder, (7) musculoskeletal or neurological disease or functional limitations disrupting voluntary movement or that might have limited training, or (8) inability to commit to the study and its requirements. The Exercise and Sports Science Australia adult pre-exercise screening tool was used to identify any individual who may be at increased risk for any adverse event(s) due to participation in our exercise program. Participants with signs or symptoms of unstable or unmanaged disease were excluded from the study.

A total of 87 older adults expressed an interest and were screened for the study, of which 20 were included. The reasons for exclusion were as follows: 18 due to the presence of a cardiovascular, musculoskeletal, or neurological condition that could limit their ability to participate in the exercise program; 11 due to being too physically active; 10 due to expected travel during the study period; 3 due to age (<65 years); and 1 due to terminal illness. The other 24 participants were not interested or did not have the time to participate in the study upon receiving further details about the requirements.

Intervention

All participants were prescribed (by a single qualified AEP recruited to work on this study) two 4-week multicomponent home exercise programs using the commercially available web-based Physitrack exercise programming app with the accompanying PhysiApp that was accessible via their iPad/tablet. Physitrack is a cloud-based, digital platform that allows health professionals to assign exercises and programs (with training dosage) to people remotely, track progress, provide feedback in real time, and send reminders. Using this program, the AEP formulated a personalized exercise program for each participant by selecting from a battery of >3500 exercises that includes narrated videos and descriptions about how to perform each exercise. The Physitrack system allows the AEP and participants to set up automated reminders about exercise times and record exercise completion, including sets, repetitions, and rate of perceived exertion (RPE) for each exercise, as well as include feedback or messages that are sent (in real time) to the AEP (or to participants from the AEP) for monitoring and review. For each exercise, participants were prescribed a specific training dose (frequency, sets, and repetitions) and asked to report on their RPE using the 10-point scale provided in the app. Each participant's program was reviewed and progressed weekly by the AEP if needed, by reviewing the self-reported RPE and sets/repetitions for every exercise completed via the web-based Physitrack platform. The AEP also checked the Physitrack system daily for any urgent alerts/messages from participants.

Exercise prescription was individualized based on each participant's initial functional capacity determined from the baseline assessment, medical and physical activity history, as well as the AEP's clinical judgement. Each exercise program included a combination of muscle strengthening (resistance) exercises, weight-bearing impact activities, and challenging balance/mobility exercises based on the principles of the *Osteo-cise: Strong Bones for Life* program, which is an established and effective community-based osteoporosis prevention exercise program for older adults at increased risk for falls and fracture [11,12]. Participants were prescribed 8-9 targeted exercises (2-3 sets of 8 repetitions) at a moderate intensity (3-6 on the 10-point modified RPE scale) to be completed on 3 nonconsecutive days per week. All participants were provided with an exercise equipment pack (box step, dumbbells, TheraBand resistance bands, stepping cones, foam balance mat). The program was designed to be completed within 30 minutes and consisted of 2 warm-up exercises (eg, marching, side stepping, sit to stand), 2 challenging balance/functional exercises (eg, alternating lateral steps, tandem walking, single leg standing), 1 upper limb (eg, wall press up, triceps dips, overhead press), and 2 lower limb resistance exercises (eg, step ups, bodyweight squats, reverse lunge with weights), 2 weight-bearing impact exercises (eg, vertical, lateral, and multidirectional jumping or hopping), and cool down (stretching) activities.

All participants received 3 home visits from the AEP during the study. At the initial home visit, baseline assessments were completed and participants were educated on the PhysiApp and safe exercise training and prescribed their initial exercise program. Participants then received a weekly phone call for the first 2 weeks of the study to monitor progress and address any questions. Thereafter, participants were encouraged to liaise with the AEP directly via the PhysiApp, which was monitored daily. A second home visit was conducted at week 4 to review progress, record any adverse events, and update the exercise program, with the final home visit conducted at the end of the study (after week 8) to complete the follow-up assessments.

Feasibility: Retention, Adherence, and Adverse Events

Retention was recorded as the number (proportion) of participants who completed the 8-week assessment. Adherence to the exercise program, including the number of sessions completed, number of exercises, and sets and repetitions completed (all expressed as a percentage) within each session were recorded within the Physitrack system. We considered the program to be feasible if at least 90% of the participants completed the trial and if the adherence to the program was at least 66% (equivalent to 2 out of 3 sessions per week). Participants were asked to record any adverse events (including falls) directly into PhysiApp so that they could be reviewed by the AEP and research staff. Information on adverse events was also collected at home visit 2 (week 4) and 3 (after week 8). An adverse event was defined as an intervention-related event resulting in absence from or modification to the exercise intervention.

Anthropometry and Demographics

Height to the nearest 0.1 cm and body weight to the nearest 0.1 kg were measured using standard procedures. The following information was collected by the questionnaire (baseline only): date of birth, ethnic background, education, living arrangement, medical history, medication use, and history of falls. At the completion of the study, participants were also asked if they had experienced a fall(s) over the past 8 weeks.

Physical Activity

Duration (minutes per week) of walking and moderate-to-vigorous physical activity (MVPA) during the past week, which was truncated at 840 minutes, was assessed using the validated Active Australia survey [17].

Physical Function

Physical function was assessed using the standardized Short Physical Performance Battery (SPPB), which is a composite measure of 3 tasks: standing balance, habitual gait speed, and repeated (5) chair rise [18]. A score of 0 to 4 was assigned to each test and added to yield a composite score ranging from 0 to 12, with higher scores indicating better physical function. The SPPB has been shown to be valid, reliable, and sensitive to change with intraclass correlation coefficients of 0.88-0.92 for tests performed 1 week apart [19].

Physical Activity Enjoyment

At baseline and follow-up, participants completed the Physical Activity Enjoyment Scale (PACES) [20]. This 18-item questionnaire asked participants to rate "*how do you feel at the moment about the physical activity you have been doing*" using a 7-point bipolar rating scale with scores ranging from 18 to 126 points. Eleven of the 18 items are reverse-scored with higher scores representing higher levels of enjoyment.

System Usability

Usability represents the participants' experience with using the app. At the final assessment, participants completed the System Usability Scale (SUS) [21] to assess perceived usability of Physitrack. The SUS is a standard 10-item questionnaire in which responses are measured on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Questions 1, 3, 5, 7, and 9 are positive and questions 2, 4, 6, 8, and 10 are negative. A total SUS score is derived by summing the individual scores and multiplying by 2.5, which yields a score ranging between 0 (worst) and 100 (absolute best). A score >68 is considered above average usability and >80 considered high usability and a level at which participants are likely to recommend the product to peers [21].

Process Measures: Perceptions of the Program and System

Upon trial completion, process measures were collected using an author-derived questionnaire completed by participants to evaluate their experiences with and perceptions of the home exercise program and Physitrack system. Enjoyment about the exercise program and using Physitrack was assessed on a 5-point scale (1=did not enjoy at all, 5=extremely enjoyable). In addition, participants were asked open-ended questions about what they liked and disliked most about using Physitrack. The level of importance about key elements that may have helped participants maintain their motivation to continue with the program was assessed using a 5-point scale (1=extremely important, 5=not all important). Finally, participants were asked whether they would continue to use the Physitrack app to exercise at home if it was made available. All open-ended questions were analyzed using a general inductive thematic approach [22].

Statistical Analysis

As this was a pilot feasibility study [23], a convenience sample of 20 older adults was recruited with no formal sample size calculations [24]. However, the observed effect sizes (Cohen *d*) for the functional and physical activity measures were calculated using the following formula: mean posttest minus mean baseline divided by baseline standard deviation. The potential clinical meaningfulness of the results (in addition to statistical significance) was based on the magnitude of the effects: small (d=0.20), medium (d=0.50), and large (d=0.80) (and P values) [25]. Nevertheless, the results and findings from the hypothesis tests should be treated with caution, given our modest sample size.

All statistical analyses were conducted using SPSS Statistics for Windows, version 26 (SPSS Inc). Baseline characteristics between the group (tablet owners and nonowners) were compared using independent two-sided t tests for continuous variables and chi-squared tests for categorical variables. Paired sample t tests (two-sided) were used to assess within-group changes for the continuous variables and the McNemar test was used for categorical variables. Between-group differences for changes were assessed using analysis of variance or McNemar test for categorical variables. Between-group differences were calculated by subtracting the within-group changes from the baseline in each group. Within-group changes were presented as absolute changes from the baseline. All data were presented as mean (SD) or 95% CI (or median and interquartile range) and the significance was set at P<.05.

Results

Baseline Characteristics

The characteristics of the cohort are shown in Table 1, with no marked differences between the tablet owners and nonowners. The mean age of the 20 participants was 70 years (range 65-81 years); 50% (10/20) of the participants were females, 40% (8/20) were classified as overweight (BMI 25-29.9 kg/m^2), and 45% (9/20) as obese (BMI ≥30 kg/m^2), with 35% (7/20) reporting the presence of a chronic disease(s) and a median of 3 medications.

Table 1. Baseline characteristics of the cohort.

Baseline characteristics	Tablet owners (n=10)	Tablet nonowners (n=10)	Total (N=20)
Age (years), mean (SD)	70.1 (3.1)	70.8 (5.3)	70.4 (4.2)
Sex (% male), n (%)	5 (50)	5 (50)	10 (50)
Height (cm), mean (SD)	165.1 (12.4)	165.7 (9.3)	165.4 (10.7)
Weight (kg), mean (SD)	83.5 (14.3)	81.5 (20.2)	82.5 (17.1)
BMI (kg/m^2), mean (SD)	30.7 (5.0)	29.3 (5.1)	30.0 (5.0)
Ethnicity, n (%)			
Caucasian	6 (60)	3 (30)	9 (45)
Other	4 (40)	7 (70)	11 (55)
Highest level of education, n (%)			
Primary/High school	3 (30)	0 (0)	3 (15)
University or Tertiary level	7 (70)	5 (50)	12 (60)
Technical/Trade certificate	0 (0)	5 (50)	5 (25)
Living arrangement, n (%)			
Alone	0 (0)	3 (30)	3 (15)
With adult without children	9 (90)	5 (50)	14 (70)
With adult with children	0 (0)	1 (10)	1 (5)
Retirement village/hostel	1 (10)	1 (10)	2 (10)
Marital status, n (%)			
Married/De Facto	9 (90)	8 (80)	17 (85)
Separated/Divorced/Widowed	1 (10)	2 (20)	3 (15)
Number of medications, median (IQR)	3.5 (1.0-4.0)	2.5 (1.0-5.5)	3.0 (1.0-4.0)
Presence of chronic disease(s),[a] n (%)	5 (50)	2 (20)	7 (35)
Previous fall in past 12 months, n (%)	2 (10)	0 (0)	2 (10)

[a]Presence of chronic disease include self-reported hypertension, cardiovascular disease, stroke, Alzheimer disease, Parkinson disease, chronic kidney disease, liver disease, type 2 diabetes, or a neurological/brain disease.

Feasibility: Retention, Adherence, and Adverse Events

Study retention was 95% (19 of 20 participants completed the study). Mean exercise adherence over the 8 weeks was 84% (95% CI 70%-97%, median 94%) and was no different between tablet owners and nonowners (mean 95% vs 72%, *P*=.07). Mean adherence to the prescribed number of exercises per session was 81% (95% CI 68%-95%), number of sets was 82% (95% CI 68%-96%), and the number of repetitions for each exercise was 81% (95% CI 68%-95%), with no differences between the 2 groups (*P*>.05). Over the 8-week program, 1 musculoskeletal complaint (knee pain that was pre-existing) and 1 injury (strained calf muscle) was reported by 2 participants. The participant with knee pain continued to exercise with a modified program, while the second participant sought treatment and subsequently withdrew from the study. No falls were reported by any participant over the 8-week study. A total of 72 in-app messages (from 480 prescribed exercise sessions) were sent to the AEP by 9 of the 20 participants (median 6 per person).

Physical Activity

In the total cohort, mean weekly time spent walking increased on average by 78 minutes (95% CI 0-156, *P*=.049; *d*=0.66] (Table 2). Thirteen participants (68%) reported an increase in the weekly walking time (range 10-360 minutes), 2 (11%) reported no change, and 4 (21%) reported a decrease (range 30-240 minutes). For MVPA, there was a mean change of 41 minutes (95% CI –8 to 90, *P*=.09; *d*=0.35]. Nine participants (47%) reported an increase in the weekly MVPA time (range 20-240 minutes), 6 (32%) reported no change, and 4 (21%) reported a decrease (range 15-180 minutes). There were no significant between-group differences (tablet versus not tablet owners) for the change in either physical activity variable.

Table 2. Changes in physical activity and the Short Physical Performance Battery scores for the tablet owners, tablet nonowners, and all participants combined.

Parameters	Baseline (N=20), mean (SD)	Week 8 (n=19), mean (SD)	Mean change (95% CI)	Effect size (Cohen *d*)
Walking time (min/week)				
Tablet owners	154 (84)	225 (155)	67 (–32 to 166)	1.20
Tablet nonowners	155 (154)	244 (251)	89 (–50 to 227)	0.58
All	155 (121)	235 (206)	78 (0 to 156)[a]	0.66
Moderate-to-vigorous physical activity (min/week)				
Tablet owners	50 (88)	97 (164)	42 (–37 to 121)	0.54
Tablet nonowners	151 (152)	191 (217)	40 (–36 to 116)	0.26
All	100 (132)	147 (195)	41 (–8 to 90)	0.36
Short Physical Performance Battery score				
Tablet owners	11.3 (0.7)	11.7 (0.7)	0.3 (–0.3 to 1.0)	0.59
Tablet nonowners	11.7 (0.7)	11.9 (0.3)	0.2 (–0.4 to 0.8)	0.30
All	11.5 (0.7)	11.8 (0.5)	0.3 (–0.1 to 0.7)	0.42

[a]*P*=.049 within group change after 8 weeks.

Physical Function

After 8 weeks, there was a nonsignificant mean 0.3 point (95% CI –0.1 to 0.7, *P*=.17) improvement in the composite SPPB score in all participants, which represented a moderate effect (*d*=0.42) (Table 2). Five participants (26%) had an improvement of one or more points in SPPB performance, 12 (63%) had no change, and 2 (11%) experienced a reduction. There were no group differences for the change in the mean composite SPPB scores.

SUS

Based on SUS, the Physitrack app was reported to be highly usable by all participants (mean score 86, SD 10), with no group differences (Table 3). For the 10 individual questions, participants most strongly agreed that they felt confident using it and that most people would learn to use the system very quickly. Participants also strongly disagreed that the system was cumbersome and unnecessarily complex.

Table 3. Means and standard deviation scores for the System Usability Scale.

Questions[a]	Tablet owners	Tablet nonowners	All
I think I would like to use the app frequently	4.22 (0.67)	4.00 (1.05)	4.11 (0.88)
I found the system to be unnecessarily complex	1.33 (0.50)	1.40 (0.52)	1.37 (0.50)
I thought the system was easy to use	4.22 (1.30)	4.40 (0.97)	4.32 (1.11)
I think that I would need support of a technical person to be able to use the system	1.67 (0.87)	1.60 (0.84)	1.63 (0.83)
I found the various functions in the system were well integrated	4.22 (0.44)	4.10 (0.99)	4.16 (0.77)
I thought there was too much inconsistency in the system	2.00 (1.10)	1.50 (0.53)	1.74 (0.81)
I would imagine that most people would learn to use the system very quickly	4.56 (0.53)	4.30 (0.68)	4.42 (0.61)
I found the system very cumbersome to use	1.11 (0.33)	1.30 (0.48)	1.21 (0.42)
I felt very confident using the system	4.67 (0.50)	4.80 (0.42)	4.74 (0.45)
I needed to learn a lot of things before I could get going with the system	1.56 (0.73)	1.60 (1.27)	1.58 (1.02)
System Usability Scale total score (out of 100)	85.6 (7.6)	85.5 (11.8)	85.5 (9.8)

[a]Responses were scored on a 5-point Likert scale: 1=strongly disagree, 5=strongly agree.

PACES

In all participants, the mean (SD) physical activity enjoyment (PACES) scores did not change over time (baseline 71.2 [8.6] vs 8 weeks 69.2 [7.9], *P*=.45) nor differ between the groups (*P*=.07).

Process Measures: Perceptions of the Program and System

Participants reported that they enjoyed using the Physitrack app and participating in the exercise program (median score 4 out of 5), with ease of use and the narrated video demonstrations of the exercises within the app reported by participants as to

what they liked most. Overall, 71% (12/17) of the participants reported no response in terms of what they disliked most about the app but a lack of clarity around the terminology (eg, the terms reps and sets) was reported by 3 participants (17%). Participants' own motivation rather than the use of the app and interactions with the exercise trainer or potential health improvements were rated as the most important factor for them continuing with the program over the 8 weeks. Finally, 94% (16/17) of the participants reported that they would recommend the home-based program to other older people and 88% (15/17) would continue to use the Physitrack app and exercise at home if it were made available (Table 4).

Table 4. Participants' perceptions of the exercise program and web-based Physitrack exercise programming system (n=17).

Questions	Scores
How did you enjoy participating in the home-based exercise program?[a], median=4.0, mean (SD)	3.50 (0.97)
How did you enjoy using the Physitrack online exercise system?[a], median=4.0, mean (SD)	3.88 (0.86)
What did you like most about the Physitrack system (app)? n (%)	
Ease of use	7 (41)
Narrated video demonstrations of exercises	7 (41)
Ability to track and modify program, if needed	3 (18)
What did you like least about the Physitrack system (app)? n (%)	
Lack of clarity around terminology	3 (17)
Lack of flexibility if exercising on a different day	1 (6)
Just another piece of equipment (device) to use when exercising	1 (6)
No response	12 (71)
How important was each of the following to help you maintain your motivation to continue with the program?[b], median=2.0, mean (SD)	
Own motivation	1.88 (0.93)
Physitrack system	2.18 (1.01)
Exercise trainer	2.24 (1.03)
Feedback from trainer and performance assessments	2.24 (1.03)
Improvements to health	2.47 (1.07)
Do you feel that taking part in the program improved your health and fitness? n (%)	15 (88)
Would you recommend the home-based program to other older people? n (%)	16 (94)
If the Physitrack system continued to be available, would you continue to exercise at home? n (%)	15 (88)

[a]Responses scored on a 5-point scale: 1=did not enjoy at all, 5=extremely enjoyable.

[b]Responses scored on a 5-point scale: 1=extremely important, 5=not important at all.

Discussion

Summary of the Main Findings

Overall, the findings from our SMART prospective pilot feasibility study indicate that it was safe and feasible for exercise professionals to prescribe and remotely monitor a thrice weekly home-based, multicomponent exercise program targeting musculoskeletal health and function delivered via the Physitrack app for older adults living independently in the community. This feasibility was evident by the low attrition, high adherence to the exercise training, low number of adverse events, increased weekly physical activity time, and the high reported usability. No differences were observed in the outcomes between participants who were previous owners and nonowners of a tablet computer. Participants reported that they enjoyed using the Physitrack app and participating in the exercise program. Most participants stated that they would recommend the home-based exercise program to other older people and continue to use the exercise programming app if it was still available, which adds further support to the widespread usability and acceptability of our intervention and the web-based exercise programming app.

Comparison With Prior Work

Consistent with our findings, several previous interventions conducted over 2-6 months have reported that the use of web/mobile exercise prescription apps represent a safe, effective, and feasible approach to deliver tablet-based, home muscle strength and balance training for older people [26-30]. However, there were marked differences in the attrition rates (8%-47%) in some of these studies, and adherence to the exercise training ranged from 61% to 73% [27-29]. This heterogeneity is likely related to factors such as whether additional behavioral change strategies or remote support were provided or not, differences in the stability of internet connections, and duration of the

interventions. In our 8-week study, the low attrition and high exercise adherence are likely due to several components related to both the intervention and web-based exercise programming system. This includes the prescription of individualized exercise programs based on individual's functional status and health/medical history, the initial telephone calls and home visits by the AEP, the option for participants to communicate with the AEP via the app at any time to receive feedback/support, knowing that the AEP was remotely monitoring all exercise programs, and the shorter study duration. Further data to support the feasibility and usability of the exercise programming system used in our study is highlighted by the findings from a 3-week pragmatic randomized controlled trial in 305 adults being treated for a musculoskeletal condition(s) [31]. In this study, it was found that the use of the Physitrack system by physical therapists improved home exercise adherence and confidence in the ability of patients to undertake exercise at home compared to usual care (eg, written exercise instructions, printed exercise diagrams) [31].

Although our study was not designed nor powered to detect an effect of the intervention on physical function, we did observe a modest effect (d=0.42) on improving SPPB scores (mean change 0.3 points). Previous research has indicated that a change in SPPB of 0.3-0.8 points represents a minimally important (significant) change [32], with a change for 0.5 and 1.0 point classified as a small but meaningful change and substantial change, respectively [33]. In our study, it is likely that the modest changes relate to the initial functional status of our participants, who had a mean SPPB score of 11.5 out of 12. Indeed, this may explain why only 26% (n=5) of the participants experienced an improvement of one or more points on the SPPB test after 8 weeks. Nevertheless, our findings must be interpreted with caution given that there was no control (nonexercise) comparison group. For comparison, a previous 12-week tablet-based, home strength and balance training program in 44 independently living older adults found that the intervention and control groups experienced similar significant improvements in SPPB scores [26].

An important finding from our study was that participants were enthusiastic about the web-based Physitrack system and its ease of use, independent of whether they had previously owned a tablet computer and would be willing to continue to use such a system in the future. Despite these positive experiences, overall enjoyment in physical activity did not change over the 8 weeks, which may be related to the relatively short duration of the exercise program or that the questionnaire we used to monitor physical activity enjoyment may not measure the ideal constructs specific to the intervention. The lack of any marked changes in physical activity enjoyment (and SPPB performance) may also be due in part to the fact that participants recruited into the study were already habitually active (mean, 100 minutes of MVPA per week). While future studies are needed to evaluate the long-term acceptability, adherence, and clinical effectiveness of remotely prescribed web-based exercise programs using apps for older adults, we wish to highlight that we are not recommending that such approaches replace traditional community-based exercise programs but provide an alternative option that might best meet some individuals' needs, preferences, and financial resources.

An interesting observation from our study was the increase in weekly walking time following the 8-week home-based, exercise program. It is important to note that participants were not specifically instructed to engage in any additional physical activity outside of the intervention. Thus, the reason(s) for the mean 78 minutes per week increase in weekly walking time is difficult to explain. However, there is some evidence that participation in structured exercise programs is associated with an increase in nonprescribed activity and energy expenditure outside of the intervention [34]. It has been suggested that this may be related to a number of factors, including exercise-related improvements in functional capacity, gains in muscle strength, reduced levels of fatigue or feeling more energetic, improvements in exercise self-efficacy, or mood [34]. However, others have observed that adoption of structured exercise leads to no change or a decrease in habitual physical activity or energy expenditure, which has been attributed to some compensatory behavioral adaptation [35]. Given the small sample size and wide confidence interval for the change in mean weekly walking time in our study, these findings must be interpreted with caution but warrant further follow-up to understand the reason(s) why home-based exercise training may improve habitual activity levels.

Limitations

This study has a number of limitations, including the small sample size, convenience sample that limits generalizability, pretest-posttest study design, relatively short intervention duration, lack of blinding of the assessor, lack of a nonexercise control group, the use of self-reported measures of physical activity, and the inclusion of generally healthy and physically active older adults with normal functional capacity, which may also affect generalizability. The 3 home-based visits by the AEP to brief participants on the use of the app and exercise program and to conduct the functional tests is a further limitation in terms of future widespread scalability. The need for the AEP to regularly review and monitor the messages/alerts from participants and their weekly progress using the Physitrack app/platform could be considered burdensome, but the daily time commitment was typically less than 5 minutes. Finally, the addition of semistructured interviews may provide further insights into the experiences and perspectives of the participants, including potential differences between tablet computer owners and nonowners.

Conclusion

This pilot feasibility study indicates that it was safe and feasible for older adults living independently in the community to participate in a tablet-computer–delivered, home-based, multicomponent exercise program targeting musculoskeletal health and function that was developed and monitored remotely by exercise professionals using a web-based exercise prescription app.

Acknowledgments

This study was supported by an internal grant from the Institute for Physical Activity and Nutrition, Deakin University. The sponsor was not involved in the design, conduct, collection, analysis, or interpretation of data; in writing of the manuscript; or in the decision to publish the findings.

Authors' Contributions

RMD, JG, NM, and RM designed the study. JG and TH conducted this study. TH collected the data. RMD and JG performed data analysis and interpretation. JG and RMD drafted the manuscript. RMD, JG, NM, and RM revised the manuscript content and provided the final approval. RMD takes responsibility for the integrity of the data analysis.

Conflicts of Interest

None declared.

References

1. Bergen G, Stevens MR, Burns ER. Falls and Fall Injuries Among Adults Aged ≥65 Years - United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Sep 23;65(37):993-998 [FREE Full text] [doi: 10.15585/mmwr.mm6537a2] [Medline: 27656914]
2. AIHW. Hospitalisations due to falls by older people, Australia 2006-07. Injury research and statistics series no. 57. Cat. No. INJCAT 133. Canberra, AIHW. 2012. URL: https://www.aihw.gov.au/getmedia/260e721e-4d8b-4ec6-b3a9-dc9c82a723cb/12266.pdf.aspx?inline=true [accessed 2020-05-01]
3. AIHW. Trends in hospitalisations due to falls by older people, Australia 2002-03 to 2012-13. Injury research and statistics series no. 106. Cat. no. INJCAT 182. Canberra, AIHW. 2017. URL: https://www.aihw.gov.au/getmedia/5f84eadd-6f25-4429-82fc-5e9072278335/aihw-injcat-182.pdf.aspx?inline=true [accessed 2020-05-01]
4. Watts JJ, Abimanyi-Ochom J, Sanders K. Osteoporosis costing all Australians: A new burden of disease analysis-2012 to 2022, Glebe, NSW: Osteoporosis Australia. 2013. URL: https://www.osteoporosis.org.au/sites/default/files/files/Burden%20of%20Disease%20Analysis%202012-2022.pdf [accessed 2020-06-01]
5. Eastell R, Rosen CJ, Black DM, Cheung AM, Murad MH, Shoback D. Pharmacological Management of Osteoporosis in Postmenopausal Women: An Endocrine Society* Clinical Practice Guideline. J Clin Endocrinol Metab 2019 May 01;104(5):1595-1622. [doi: 10.1210/jc.2019-00221] [Medline: 30907953]
6. Daly RM, Dalla Via J, Duckham RL, Fraser SF, Helge EW. Exercise for the prevention of osteoporosis in postmenopausal women: an evidence-based guide to the optimal prescription. Braz J Phys Ther 2019;23(2):170-180 [FREE Full text] [doi: 10.1016/j.bjpt.2018.11.011] [Medline: 30503353]
7. Sherrington C, Fairhall N, Wallbank G, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community: an abridged Cochrane systematic review. Br J Sports Med 2020 Aug;54(15):885-891. [doi: 10.1136/bjsports-2019-101512] [Medline: 31792067]
8. Zhao R, Zhao M, Xu Z. The effects of differing resistance training modes on the preservation of bone mineral density in postmenopausal women: a meta-analysis. Osteoporos Int 2015 May;26(5):1605-1618. [doi: 10.1007/s00198-015-3034-0] [Medline: 25603795]
9. Kukuljan S, Nowson CA, Sanders KM, Nicholson GC, Seibel MJ, Salmon J, et al. Independent and combined effects of calcium-vitamin D3 and exercise on bone structure and strength in older men: an 18-month factorial design randomized controlled trial. J Clin Endocrinol Metab 2011 Apr;96(4):955-963. [doi: 10.1210/jc.2010-2284] [Medline: 21209030]
10. Kukuljan S, Nowson CA, Sanders K, Daly RM. Effects of resistance exercise and fortified milk on skeletal muscle mass, muscle size, and functional performance in middle-aged and older men: an 18-mo randomized controlled trial. J Appl Physiol (1985) 2009 Dec;107(6):1864-1873 [FREE Full text] [doi: 10.1152/japplphysiol.00392.2009] [Medline: 19850735]
11. Gianoudis J, Bailey CA, Ebeling PR, Nowson CA, Sanders KM, Hill K, et al. Effects of a targeted multimodal exercise program incorporating high-speed power training on falls and fracture risk factors in older adults: a community-based randomized controlled trial. J Bone Miner Res 2014 Jan;29(1):182-191 [FREE Full text] [doi: 10.1002/jbmr.2014] [Medline: 23775701]
12. Daly RM, Gianoudis J, Kersh ME, Bailey CA, Ebeling PR, Krug R, et al. Effects of a 12-Month Supervised, Community-Based, Multimodal Exercise Program Followed by a 6-Month Research-to-Practice Transition on Bone Mineral Density, Trabecular Microarchitecture, and Physical Function in Older Adults: A Randomized Controlled Trial. J Bone Miner Res 2020 Mar;35(3):419-429. [doi: 10.1002/jbmr.3865] [Medline: 31498937]
13. Ziebart C, McArthur C, Lee L, Papaioannou A, Laprade J, Cheung AM, et al. "Left to my own devices, I don't know": using theory and patient-reported barriers to move from physical activity recommendations to practice. Osteoporos Int 2018 May;29(5):1081-1091. [doi: 10.1007/s00198-018-4390-3] [Medline: 29441402]
14. Cavill NA, Foster CE. Enablers and barriers to older people's participation in strength and balance activities: A review of reviews. JFSF 2018 Jun;03(02):105-113. [doi: 10.22540/jfsf-03-105]

15. Timmons JF, Griffin C, Cogan KE, Matthews J, Egan B. Exercise Maintenance in Older Adults 1 Year After Completion of a Supervised Training Intervention. J Am Geriatr Soc 2020 Jan;68(1):163-169. [doi: 10.1111/jgs.16209] [Medline: 31647585]
16. Yardley L, Kirby S, Ben-Shlomo Y, Gilbert R, Whitehead S, Todd C. How likely are older people to take up different falls prevention activities? Prev Med 2008 Nov;47(5):554-558. [doi: 10.1016/j.ypmed.2008.09.001] [Medline: 18817810]
17. Australian Institute of Health and Welfare. The Active Australia Survey: A guide and manual for implementation, analysis and reporting. 2003. URL: https://www.aihw.gov.au/getmedia/ff25c134-5df2-45ba-b4e1-6c214ed157e6/aas.pdf.aspx?inline=true [accessed 2020-06-01]
18. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000 Apr;55(4):M221-M231. [doi: 10.1093/gerona/55.4.m221] [Medline: 10811152]
19. Ostir GV, Volpato S, Fried LP, Chaves P, Guralnik JM. Reliability and sensitivity to change assessed for a summary measure of lower body function. Journal of Clinical Epidemiology 2002 Sep;55(9):916-921. [doi: 10.1016/s0895-4356(02)00436-5]
20. Kendzierski D, DeCarlo K. Physical activity enjoyment scale: two validation studies. Int J Sport Exerc Psychol 1991;13:50-64. [doi: 10.1123/jsep.13.1.50]
21. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Abbade LPF, Abbade JF, Thabane L. Introducing the CONSORT extension to pilot trials: enhancing the design, conduct and reporting of pilot or feasibility trials. J Venom Anim Toxins Incl Trop Dis 2018;24:4 [FREE Full text] [doi: 10.1186/s40409-018-0142-2] [Medline: 29434634]
24. 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]
25. Cohen J. Statistical power analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates; 1988.
26. van Het Reve E, Silveira P, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: part 2 of a phase II preclinical exploratory trial. J Med Internet Res 2014 Jun 25;16(6):e159 [FREE Full text] [doi: 10.2196/jmir.3055] [Medline: 24966165]
27. Silveira P, van de Langenberg R, van Het Reve E, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase II preclinical exploratory trial. J Med Internet Res 2013 Aug 12;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]
28. Geraedts HAE, Zijlstra W, Zhang W, Spoorenberg SLW, Báez M, Far IK, et al. A Home-Based Exercise Program Driven by Tablet Application and Mobility Monitoring for Frail Older Adults: Feasibility and Practical Implications. Prev Chronic Dis 2017 Feb 02;14:E12 [FREE Full text] [doi: 10.5888/pcd14.160227] [Medline: 28152361]
29. Nikitina S, Didino D, Baez M, Casati F. Feasibility of Virtual Tablet-Based Group Exercise Among Older Adults in Siberia: Findings From Two Pilot Trials. JMIR Mhealth Uhealth 2018 Feb 27;6(2):e40 [FREE Full text] [doi: 10.2196/mhealth.7531] [Medline: 29487045]
30. Hong J, Kong H, Yoon H. Web-Based Telepresence Exercise Program for Community-Dwelling Elderly Women With a High Risk of Falling: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 28;6(5):e132 [FREE Full text] [doi: 10.2196/mhealth.9563] [Medline: 29807877]
31. Bennell KL, Marshall CJ, Dobson F, Kasza J, Lonsdale C, Hinman RS. Does a Web-Based Exercise Programming System Improve Home Exercise Adherence for People With Musculoskeletal Conditions?: A Randomized Controlled Trial. Am J Phys Med Rehabil 2019 Oct;98(10):850-858. [doi: 10.1097/PHM.0000000000001204] [Medline: 31021823]
32. Kwon S, Perera S, Pahor M, Katula JA, King AC, Groessl EJ, et al. What is a meaningful change in physical performance? Findings from a clinical trial in older adults (the LIFE-P study). J Nutr Health Aging 2009 Jun;13(6):538-544. [doi: 10.1007/s12603-009-0104-z] [Medline: 19536422]
33. Perera S, Mody S, Woodman R, Studenski S. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006;54:743-749. [doi: 10.1111/j.1532-5415.2006.00701.x]
34. Drenowatz C, Grieve GL, DeMello MM. Change in energy expenditure and physical activity in response to aerobic and resistance exercise programs. Springerplus 2015;4:798 [FREE Full text] [doi: 10.1186/s40064-015-1594-2] [Medline: 26702387]
35. Melanson EL. The effect of exercise on non-exercise physical activity and sedentary behavior in adults. Obes Rev 2017 Feb;18 Suppl 1:40-49 [FREE Full text] [doi: 10.1111/obr.12507] [Medline: 28164451]

Abbreviations

AEP: accredited exercise physiologist

BMD: bone mineral density
MVPA: moderate-to-vigorous physical activity
PACES: physical activity enjoyment scale
RPE: rate of perceived exertion
SMART: Seniors Made Active thRough Technology
SPPB: short physical performance battery
SUS: system usability scale

Edited by L Buis; submitted 05.06.20; peer-reviewed by N Gell, H Hawley-Hague; comments to author 04.08.20; revised version received 25.08.20; accepted 29.10.20; published 13.01.21.

Please cite as:
Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R
Feasibility, Usability, and Enjoyment of a Home-Based Exercise Program Delivered via an Exercise App for Musculoskeletal Health in Community-Dwelling Older Adults: Short-term Prospective Pilot Study
JMIR Mhealth Uhealth 2021;9(1):e21094
URL: http://mhealth.jmir.org/2021/1/e21094/
doi:10.2196/21094
PMID:33439147

Original Paper

The Human Factor in Automated Image-Based Nutrition Apps: Analysis of Common Mistakes Using the goFOOD Lite App

Maria F Vasiloglou[1*], MSc; Klazine van der Horst[2*], PhD; Thomai Stathopoulou[1], MSc; Michael P Jaeggi[2], BSc; Giulia S Tedde[2], BSc; Ya Lu[1], MSc; Stavroula Mougiakakou[1], PhD

[1]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

[2]Department of Health Professions, Bern University of Applied Sciences, Bern, Switzerland

[*]these authors contributed equally

Corresponding Author:
Stavroula Mougiakakou, PhD
ARTORG Center for Biomedical Engineering Research
University of Bern
Murtenstrasse 50
Bern, 3008
Switzerland
Phone: 41 6327592
Email: stavroula.mougiakakou@artorg.unibe.ch

Abstract

Background: Technological advancements have enabled nutrient estimation by smartphone apps such as goFOOD. This is an artificial intelligence–based smartphone system, which uses food images or video captured by the user as input and then translates these into estimates of nutrient content. The quality of the data is highly dependent on the images the user records. This can lead to a major loss of data and impaired quality. Instead of removing these data from the study, in-depth analysis is needed to explore common mistakes and to use them for further improvement of automated apps for nutrition assessment.

Objective: The aim of this study is to analyze common mistakes made by participants using the goFOOD Lite app, a version of goFOOD, which was designed for food-logging, but without providing results to the users, to improve both the instructions provided and the automated functionalities of the app.

Methods: The 48 study participants were given face-to-face instructions for goFOOD Lite and were asked to record 2 pictures (1 recording) before and 2 pictures (1 recording) after the daily consumption of each food or beverage, using a reference card as a fiducial marker. All pictures that were discarded for processing due to mistakes were analyzed to record the main mistakes made by users.

Results: Of the 468 recordings of nonpackaged food items captured by the app, 60 (12.8%) had to be discarded due to errors in the capturing procedure. The principal problems were as follows: wrong fiducial marker or improper marker use (19 recordings), plate issues such as a noncompatible or nonvisible plate (8 recordings), a combination of various issues (17 recordings), and other reasons such as obstacles (hand) in front of the camera or matching recording pairs (16 recordings).

Conclusions: No other study has focused on the principal problems in the use of automatic apps for assessing nutritional intake. This study shows that it is important to provide study participants with detailed instructions if high-quality data are to be obtained. Future developments could focus on making it easier to recognize food on various plates from its color or shape and on exploring alternatives to using fiducial markers. It is also essential for future studies to understand the training needed by the participants as well as to enhance the app's user-friendliness and to develop automatic image checks based on participant feedback.

(JMIR Mhealth Uhealth 2021;9(1):e24467) doi:10.2196/24467

KEYWORDS

mHealth; dietary assessment; smartphone; apps; human mistakes; mobile phone

Introduction

Background

Around 12% of mobile health (mHealth) apps have a *wellness* focus on nutrition and diet. These apps often enable users to capture their food intake and receive recommendations for a healthy diet [1]. Sharp et al [2] reviewed comparative studies on methods of recording diets; they found that most participants preferred procedures that included dietary assessment methods based on a mobile phone rather than conventional reference methods [2].

These new technologies have several advantages over conventional methods. First, they do not rely on respondents' memory, but they can provide users with automatically processed data and real-time advice [3]. Second, they use portable devices and have better social acceptance than conventional methods of dietary assessment [4]. Researchers also benefit from smartphone app-based methods, as apps can decrease workload, reduce printing and postage costs, lower the risk of transcription errors [5], and optimize space and security aspects required for paper file storage [6]. Owing to their feasibility and cost-effectiveness, innovative mobile phone–based tools may be superior to conventional tools in large-scale setups [7].

Image-based apps can be divided into 2 broad categories. The first one includes the majority of the existing apps, which are either (1) manual, in which no artificial intelligence (AI) component is integrated and the user inserts manually both the type and portion size of food or drink, or (2) semiautomatic, in which some type of AI features are integrated, for example, automatic food recognition, but the portion size estimation is manually provided by the user [8]. The second category includes systems that are fully automatic based on AI approaches [8]. Systems of the first category usually require the user to manually enter the food item, while often they use either barcode scanners to recognize packaged food labels [9] or algorithms for the automatic recognition of food items from images. Once food items have been recognized, the user is typically asked to enter portion size or volume by hand, so that the system can convert this input into nutrient information. However, tools in this category are usually not validated or certified, the number of food categories they support is limited, and it is not always clear which nutrient database is used for nutrition information [10]. Another drawback of this category is that individuals may inaccurately estimate the portion size [11]. This is a significant problem and accounts for nearly 50% of the mistakes in the food records of dietary assessment apps [12]. The systems in the second category (totally based on AI) use food image or video input [13-17] automatically and in real time to (1) identify and segment the different food items, (2) recognize each type of food item, and (3) create a 3D model of all individual food items. The conversion of food images or videos to calories or macronutrient content is supported by food composition databases (eg, United Sates Department of Agriculture [USDA] nutrient database and Swiss food composition database) [8]. The primary limitation of this category is that some food types, such as mixed foods (eg, lasagna) or beverages, are challenging to analyze [13], as they usually contain ingredients that are not clearly identifiable or that are placed in containers, which makes it difficult to assess volume.

Owing to the use and continuous improvement of mHealth and AI technologies, it is likely that they will replace paper-and-pencil methods altogether. Currently, nutritional scientists studying dietary assessment by apps mainly focus on comparing innovative with conventional methods for dietary assessment, whereas computer scientists mainly focus on optimizing the algorithms. However, to achieve high-quality data, both nutrition and computer scientists also need to focus on the behavioral aspects of data acquisition. In an international survey conducted among health care professionals (n=1001) in 6 continents, they mentioned that to recommend a *Nutrition and Diet* app to their patients or clients, they would prefer an app that is easy to use (87.1%), validated (68.1%), supports automatic food recording (56.5%), and automatically outputs nutrient estimations (52.4%) [18]. However, accurate data for food quality assessments can only be based on the correct capture of meal images, which is of vital importance if smartphones are to be used as a reliable source for food records. Thus, the quality of the captured data can severely affect the quality of the assessment, and it is indicated that correct data capture is a critical factor if the app is to be properly used and is to provide the most accurate results. Rather than removing erroneous data from studies, in-depth analysis should help explore common mistakes and thus further improve automated apps to assess nutritional intake.

Objectives

Along these lines, the aim of this study is to explore and evaluate common user errors made when using the goFOOD Lite app for collecting dietary intake data, considering that the collected data are used for automatic food recognition and nutrient estimation (goFOOD). Thus, mistakes can potentially influence the automation of the process. The results will help improve the instructions given, adjust the app to the user needs, and enhance the overall automatic functionalities of the app.

Methods

Recruitment and Screening Procedure

Sample Inclusion and Exclusion Criteria

Eligible volunteers for the study included adults (18 years and older) from the general population with self-reported adequate literacy in information communication technology, that is, they knew how to use a smartphone, who also provided written informed consent before the start of the study. Dietitians, nutritionists, and students in the fields of nutrition and dietetics were excluded to avoid bias related to profession. No prior familiarity with the app was needed to be eligible to participate. The participants did not perform any nutrient estimation; hence, they were not required to be experienced in this subject.

Participants

A convenience sample was recruited following the snowball method, that is, starting with acquaintances of members of the team at Bern University of Applied Sciences and the Artificial

Organ (ARTORG) Research Center of Biomedical Engineering (University of Bern). In addition, students who had been informed of our study through our communication campaign via promotional flyers at clinics, the campus of Bern University of Applied Science, and social media were enrolled.

goFOOD System

goFOOD is an Android system that supports both images and video as an input to automatically determine the type, volume, calories, and macronutrient content (carbohydrates, fat, and protein) of a meal by using AI and computer vision–based analysis of images acquired by the users' smartphones [19].

The version of interest in this study is the one using 2 meal images as input, captured from different viewing angle, and a reference card that must be placed next to the meal during the image capturing. The image processing module of the system consists of the following 3 main stages: (1) food segmentation, (2) food item recognition, and (3) volume estimation. Deep neural networks are applied to process the captured food images, and this performs food segmentation and recognition [20,21], whereas a 3D reconstruction–based algorithm estimates food volume [17]. The meal's calorie and macronutrient content are then calculated on the basis of each food category, volume, and food composition database [22,23]. goFOOD supports 319 fine-grained food categories and has been validated technically.

goFOOD Lite and Participants' Actions or Walkthrough Process

For the purpose of this study, a simplified version of goFOOD, called goFOOD Lite, was developed. With goFOOD Lite, the users can record their meal (food or beverage), but—unlike in the original goFOOD app—the Lite version does not provide any estimated results (eg, nutrient content and portion size) to the users as would have been the case for collecting dietary intake on the population level. With this app, participants could record their food intake by taking 2 photographs at specific angles of their meal before and after eating. This version of the app informs the user of the correct positioning (angle) of the phone for photo and input. As in other apps, goFOOD Lite uses a specially designed reference card as a fiducial marker that must be placed next to the recorded item to ensure that the 3D volume estimation is accurate. Moreover, for the images to be valid, the following criteria must be met:

1. The recorded foods are best positioned within an elliptical plate, either neutral (white) in color or with high color contrast to the background. Though this is not an absolute requirement, the participant is urged to comply, as this facilitates subsequent processing, that is, nutrient estimation.
2. The recorded item must be fully visible in the image. If it is a plated meal, then the entire plate must be in the image. If it is a nonplated meal, then the entire items must be within the image. If it is a beverage, then the entire glass or bottle containing the beverage must be within the image.
3. A special reference card (the size of a credit card) used as a fiducial marker is provided to the participant. This card must be used for all recordings of foods or beverages, as it is of vital importance for the estimation of food volume and the subsequent nutritional analysis. The card must be placed on the same table or surface as the items being recorded and has to be fully visible in all images from its top side, with the most colors and textures. Loyalty cards of large supermarket chains can also be used instead of the designated reference card.
4. The relative position of the recorded item and the card must not change between the 2 different lateral images.
5. The foods and beverages must be recorded separately.

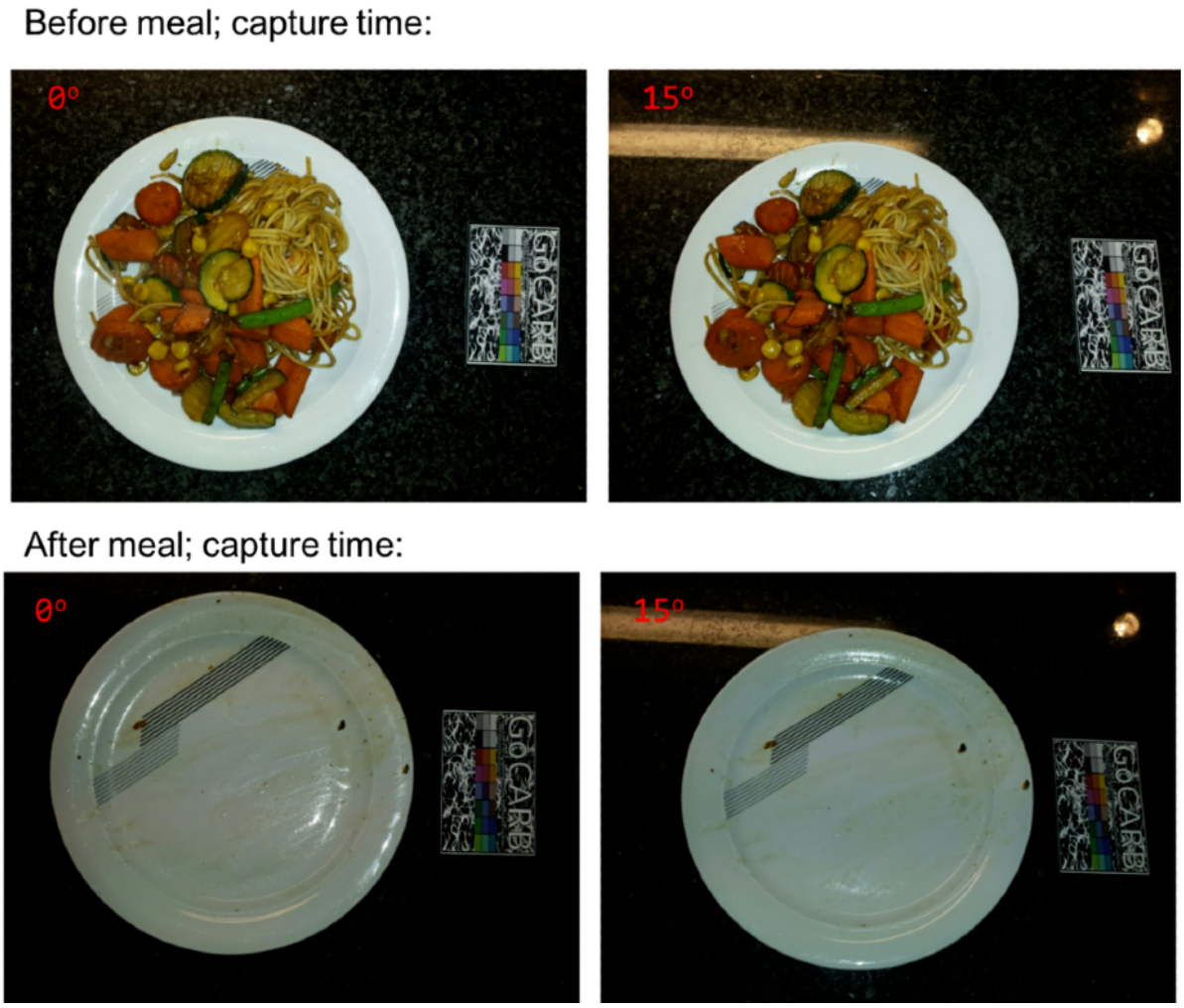
Every consumed food (plated, nonplated, or packaged) and beverage needs to be recorded. More specifically, the participants were asked to record their food and beverages before and after consumption. One recording comprises 2 images of the corresponding item or items, that is, the food or beverage. These 2 images are captured from different viewing angles, as indicated and guided by the app (0° and 15° to the surface or table). The recording also contains the creation time and date. An example of 2 correct recordings is shown in Figure 1.

The specific procedure for each recording was as follows:

1. The user indicates if the recording is for a food or a beverage.
2. The user indicates if the recording occurs before or after consumption.
3. The user captures 2 images at specific angles. The app has a feature that guides the user toward the correct angle, and if this angle is not met, then it is not possible to capture the image.
4. The app attempts to transmit the recording to the server.
    a. If no internet connection is available, then the user is prompted with an informative message urging them to ensure that the phone has internet access and to attempt to record again. The recording is not completed and is not transmitted or stored.
    b. If a working internet connection is available, the recording is transmitted and stored on the server. The recording is completed.

If the procedure does not reach step 4b, then it is not considered complete and is not stored.

Figure 1. Example of 2 correct food recordings: 1 recording before and 1 recording after consumption.



Study Procedure

Study participants were asked to use the goFOOD Lite app to record all foods and beverages consumed during a period of 24 hours.

Participants took part in 1 of the 3 different instruction days to be informed about the aim of the study, how the goFOOD Lite app works, and to sign consent forms. The procedure and criteria presented in the previous section (*goFOOD Lite and Participants' Actions or Walkthrough Process*) were clearly explained to the participants as well as the correct side of the reference card; this is part of the recording procedure but has not yet been specifically mentioned.

The participants were guided through the app's functions by watching several demos and by trying the app themselves and asking questions. Those unable to be present on the designated introduction days received personal instructions from one of the scientists responsible for data collection. At the end of each instruction day, all participants were asked to sign a consent form. No compensation was provided to the participants. Every participant was provided with the following:

- An Android smartphone with a preinstalled goFOOD Lite app and a functioning 4G internet connection
- A designated reference card
- Instructions for returning the phone at the end of the 24-hour test period
- Written instructions and a demonstration video to ensure the proper usage of the app

The participants had to provide basic personal data (address, email, and telephone number) on the instruction days, as this was required for sending the smartphones and the delivery of the log-in data and the instructional video.

The list of instructions provided to the participants, as a summary, is as follows:

1. Produce a recording of every food or drink you consumed within a period of 24 hours.
2. Produce a recording before and after consuming the food or drink.
3. Make sure that the phone has internet access before attempting any recording.
4. Follow the 4 procedural steps presented to you on the instruction day and given to you with the written instructions.
5. Try to follow criterion number 1 (presented in *goFOOD Lite and Participants' Actions or Walkthrough Process* section).
6. Make sure that criteria 2-5 are met; otherwise, the recording will not be valid.

Furthermore, an email including the participant's log-in data was sent by our team to each participant. In addition, there was no further communication unless the participant faced technical difficulties, in which case, they were instructed to contact the team for support.

Statistical Analysis

All the images stored on the server were evaluated and categorized into one of the following mistake categories: (1) missing recording, that is, only before or only after consumption recording; (2) packaged food mistake; (3) plate mistake; (4) fiducial marker mistake; (5) combination; and so on. The chi-square test of independence was performed to examine the relation between participants who made mistakes and their sex, age, number of days since the instruction day, and whether or not the participant pursued a technical profession. RStudio (version 1.0.153, 2009-2017 RStudio, Inc) was used for data processing. Statistical significance was set at P=.04. Descriptive analysis, defined as mean (SD), was performed.

Results

Self-Reported Basic Characteristics

The study began with 50 participants, but 2 participants dropped out due to technical difficulties. The study then included 48 participants (27 men and 21 women) with a mean age of 34.2 years (SD 11.7). All participants were Caucasian German-speaking Europeans living in Switzerland. The average BMI (kg/m²) was 22.7 (SD 2.9), with 76% (38/50) of the study subjects lying within the normal range (BMI 18.5-24.9), 16% (8/50) being overweight (BMI 25.0-29.9), 2% (1/50) being obese (BMI≥30), and 6% (3/50) being underweight (BMI<18.5). The self-reported characteristics of the study participants are shown in Table 1.

Table 1. Characteristics of the participants (n=48).

Characteristics	Participants, n (%)
Age (years)	
18-29	18 (37.5)
30-49	24 (50.0)
>50	6 (12.5)
Sex	
Female	21 (43.8)
Male	27 (56.2)
Ethnicity or race	
Caucasian (European)	48 (100.0)
Spoken language	
German	48 (100.0)
Profession	
Student	9 (18.8)
Employed	38 (79.2)
Retired	1 (2.1)

Food Pictures

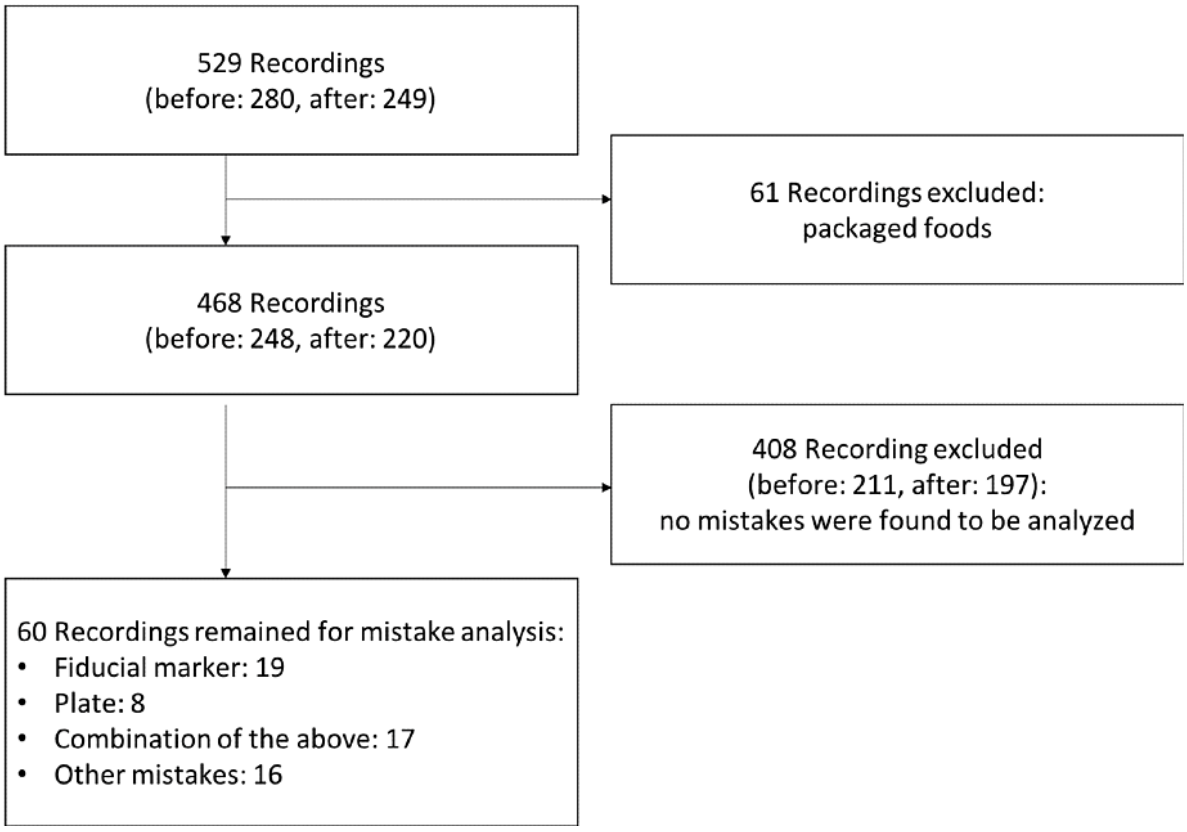
A total of 529 food recordings were captured by the app. Of these, 9.6% (51/529) were single recordings, that is, they contained images from only before or only after the meal. However, these 51 recordings were not discarded from the automatic analysis. The remaining 478 recordings formed 239 before or after meal pairs. Moreover, 61 of the initial 529 recordings contained packaged food and were excluded from the automatic analysis, as the system did not yet support a barcode scanner when the study was conducted.

Of the 468 nonpackaged food recordings, 60 (12.8%) contained mistakes and were further categorized. More details on the subcategories of the mistakes are given in the next section (*Characteristics of Errors*). In Figure 2, examples of correct and usable photos are provided. In detail, 4 recordings are shown, 2 from before the meal (left images) and 2 from after the meal (right images). Each recording is represented by 1 of the 2 (different angles) images captured. All 4 of the pictures shown in Figure 2 have a second image, captured from a different angle, but as the second angle is identical in terms of the food items and plates shown, we only provide 1 of them here.

Figure 3 describes the process of data filtering to exclude photos that could not be processed due to errors or development stage. As mentioned in the *Methods* section, no frequent communication was planned from the members of our team with the participants, unless they faced technical difficulties.

Figure 2. Samples of usable recordings.

Figure 3. Flowchart of the data filtering process of the images obtained by goFOOD Lite (1 recording=2 photos captured at a 0° and 15° angle from the table or surface).



Characteristics of Errors

This section describes the characterization and categorization of the encountered errors.

Fiducial Marker Errors (n=19)

To process the pictures correctly, the app requires a fiducial marker, in our case, an object the size of a credit card. In our study, the participants had to use a dedicated card created by our lab or commercial supermarket loyalty cards. The instructions given indicated that the card must (1) be placed next to the plate or food or beverage, (2) be placed on the same table or surface as the plate or food or beverage, (3) be fully visible, (4) not be moved between the 2 angle photos of 1 recording (0° and 15°), and (5) be placed with a specific side facing up (the correct side was clearly indicated). Some pictures included nonfully visible or no cards (n=9). Other issues included the use of the correct card but at the wrong side (n=2); pictures where the card was placed on top of the plate (n=2); and 6 pictures that contained mistakes outside of the aforementioned categories, such as incorrect cards or cards being moved between the 2 angle photos of 1 recording (n=6). Examples are shown in Figure 4.

Figure 4. Examples of images: (a) fiducial marker (reference card) on top of the plate and (b) no fiducial marker.

Plate Not Fully Visible (n=8)

Although participants were informed that capturing the entire plate is required for the app to function, there were nonetheless instances where the plate was not fully visible in the picture (n=8). Some examples are provided in Figure 5.

Figure 5. Examples of images where the plate is not fully visible.

Combinations of Errors (n=17)

A combination of the 2 previous error categories as well as with other issues are provided in this section, and examples of which are depicted in Figure 6. Following a list of the combination of errors is provided:

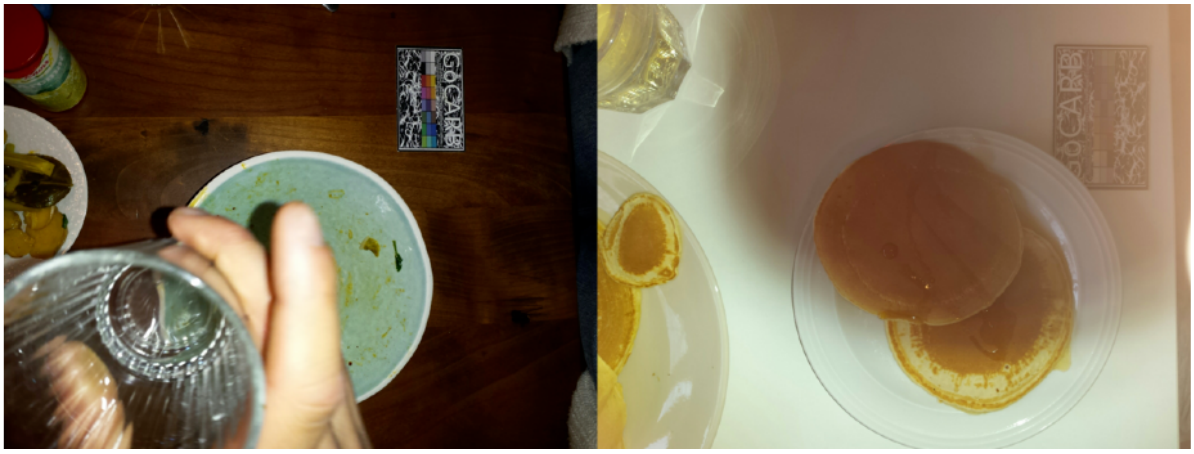
- Problems with the plate plus an obstacle (n=2)
- Problems with both the plate and the card not being fully visible (n=6)
- Card not fully visible and duplicate entry of the same image (n=2)
- A picture where the card was moved between the first and second pictures and where it was not fully visible (n=1)
- A picture where the food item changes between the 2 angle pictures of the same recording and the card is placed on the plate (n=1)
- Use of an incorrect type of card combination with a plate that was not fully visible plate (n=4).

Figure 6. Example of combinations of errors (plate not fully visible and fiducial marker—reference card—not fully visible).

Other Mistakes in Photo Entries (n=16)

Some images included an obstacle hindering the visibility of the recorded meal, such as a person's hand (n=3). In one instance, the 2 pictures from different angles that are required for 1 recording contained different items (image at 0° was food and image at 15° was drink; n=1). Other mistakes that contained errors on files (n=2) or training testing images (n=7) were also excluded. Examples are shown in Figure 7.

Figure 7. Examples of images with mistakes of (a) hand-hindering visibility and (b) shadow-hindering visibility.



Other Issues: Nonerrors

In this section, we wish to present certain issues that cannot be considered user errors, as they have never been strictly instructed, but can affect the smooth functionality of the app. Specific issues relate to the limitations of the software or hardware. Participants were informed that the ideal scenario included elliptical white plates, but they were not strictly instructed to use plates with these characteristics. As a result, issues that posed a challenge from an algorithmic point of view included the use of plates with a nonelliptical shape, for example, rectangular, highly patterned plates, transparent plates, or those that created reflections. The highly patterned plates make it more difficult for the system to detect and recognize the food. Similarly, a plate that is not elliptical may impair overall accuracy, as in such cases, the system is required to estimate a corresponding elliptical plate. Examples of the aforementioned issues are presented in Figure 8.

Certain other issues are related to possible differences in the needs for instruction, as perceived by the study organizers and the end users. At the time of the study, no barcode scanner was integrated into the app. Therefore, no specific instructions were given in this respect. However, users were instructed to record all consumed items if they were part of their daily diet. By examining the recordings of packaged foods, we discovered a common problem with the images. In many cases, the before and/or after recordings contained photos showing only the packaging of the food and not its content, which is the actual food. Such an example is shown in Figure 9. This is highly problematic, as the system currently recognizes food and is not capable of recognizing brands and/or read barcodes. However, even if the system had been able to recognize the type of food from the barcode, these recordings would still be problematic, as it would be impossible to estimate the quantity. In this case, the recording of the meal both before and after consumption would be useless.

The same issue was encountered in certain beverage recordings. A number of these contain recordings of opaque bottles, where the contents, that is, the actual beverage, as well as the quantity were not visible.

Figure 8. Examples of other issues: (a) wrong plate, object inside, reflections created and (b) wrong plate, highly textured.

Figure 9. Example of mistake made while capturing packaged foods (actual food not visible).

Demographic Characteristics of the Participants Making Errors

Our analysis of the errors revealed that 52% (25/48) participants made at least one error, whereas 4 participants made 4 errors and 1 participant made 6 errors. Of the participants who made errors, 13 were women and 12 were men. By age, 72% (13/18) of people aged 18-29 years made mistakes, 42% (10/24) of those aged 30-39 years made mistakes, and finally 33% (2/6) of those aged over 50 years made mistakes. However, when we checked for associations between those who made mistakes in relation to their sex (P=.65), age (P=.38), technical knowledge (P=.22), or days passed from the instruction day (P=.65), no statistically significant differences were detected.

Discussion

Principal Findings

This study analyzed human errors and other challenges caused by human factors when assessing food intake using the AI-based system goFOOD. Analysis of user errors should provide useful material for improving the app and with it, the quality and reliability of app-based diet recording. Most existing studies focus only on the validity or compare the performance of innovative and conventional methods, for example, apps versus

manual food logs for tracking food intake and dietary assessment. In those studies, participants are generally only asked about their opinion of the user experience and/or user-friendliness of such apps; however, the reasons for any errors made in recording food intake of image-based apps are usually not investigated. In our study, 12.8% (60/468) of the captured images of nonpackaged food items would have to be discarded. Moreover, 52% (25/48) of our participants made at least one mistake in photographing their food with the app, which underlines the need for improvements in automatic methods for collecting data on dietary intake as well as in the instructions and points of emphasis provided to the user.

The main errors were primarily associated with incorrect or improper use of the fiducial marker; problems with the plates used (such as incompatible or only partially visible plates); and missing recordings, that is, discordant pairs of recordings before and after consumption. Other problems were related to testing recordings, obstruction of the camera with a hand, and a combination of errors. The results obtained from the chi-square tests indicate that errors made with the app were not associated with age, technical knowledge, sex, and time since instruction. Although our sample size is small, we could suggest that advanced technological literacy does not play a role in fully understanding how to meet the image criteria required for the app to function. Although, according to chi-square tests, we found that older people tend to make fewer mistakes, in fact, we cannot draw strong conclusions related to age, as our sample of older participants is very small. Some assumptions include that older participants were possibly more cautious after receiving instructions for using the app, as they informed us that they wanted to be conscientious in their contribution to this study. Although young people are generally more familiar with smartphones, this contradictory result may be due to overestimation of their own technological literacy or difficulty when it comes to paying attention to guidelines. Furthermore, our findings imply that immediate initiation of food and beverage consumption recording after the instruction day does not influence the number of mistakes or the correct use. As a consequence, further studies can consider the use of an image-based log app any time after the instruction days.

Comparison With Prior Work

Other studies have been conducted where the user aspect of a nutrition app was mentioned, but without fully analyzing or providing a goal for using these observations. A study with adolescents (n=18) tested their amenability in free-living conditions with limited parental input. They were asked to use the FRapp to record their dietary intake. This app uses inputs such as images, text, voice, barcodes, and a selection of recently consumed food sets. After indicating consumed food sets, the user must insert the type and amount of food consumed. The authors suggested that only a minority of participants followed all directions [24]. In another study, the participants used the mpFR app to acquire images both before and after consumption of meals and beverages. This app uses semiautomatic food recognition, automatic volume estimation, and semiautomatic nutrient estimations [25]. An additional app study using mpFR was conducted. This included adolescents and their parents, and participants were requested to capture before and after images of their food and beverages [26]. Several findings of these studies are discussed in the following paragraphs.

Before and After Images

In our study, 9.6% (51/529) of the images obtained from the app had only one recording from either before or after the meal. Our results are in line with other studies in which participants were asked to capture 2 separate meals (each before and after consumption) to enter in a dietary assessment app. Participants included both before and after pictures in 80% of entries for the first meal and 84% of entries for the second meal [25]. In a study with both adults and adolescents, it was found that adults were more likely than adolescents to capture images both before and after the meal and to include all the foods and beverages consumed [26]. Furthermore, in a study by Casperson et al [24], participants had difficulties in remembering to capture a postmeal image, which implies that this requirement is problematic. To help remind people to take the postconsumption image of a meal, it might be beneficial to have personalized reminders such as timing functions accompanied by sounds or vibrations or pop-up messages, as this could possibly reduce the omission of *after consumption* images.

Fiducial Marker

Despite taking earlier user feedback into account in our study design, a total of 19 errors related to the fiducial marker still occurred. In our study, the number of recordings that did not include the entire fiducial marker was small (1.7%). However, other researchers have reported errors with fiducial markers due to participants having difficulty with the size of the card. Daugherty et al [26] asked adults (n=57) and adolescents (n=78) to record 1 or 2 meals under uncontrolled conditions, and they noted that the fiducial marker required (checkerboard square) was too large and, as a result, was sometimes partially covered by a plate or utensil (98/156). With regard to the type of fiducial marker preferred, the majority of adults (91%) and adolescents (67%) mentioned that a credit card–sized object would be easier to use and carry. The respective percentages for a USB-sized fiducial marker were 42% (30/78) for adults and 67% (38/57) for adolescents.

Another analysis from the same study with the same adolescent sample (n=70) indicated that only 23% to 29% of participants did not include the entire fiducial marker in both their before and after images [25]. According to the researchers, these mistakes arose because some participants were too short to capture the entire meal correctly. Moreover, they pointed out that even with repeated use of the app, no significant change was noted in the number of participants who included the entire fiducial marker. This mistake can be avoided or alleviated by adding a screen notification reminder on the use of the fiducial marker [22,26]. Another possible solution could be an automatic detector, which will verify in real time whether or not the marker is within the camera frame and thus inform the user. However, with the new generation of smartphones equipped with 2 or more rear cameras, the fiducial marker can be eliminated, as 2 images can be captured simultaneously [19,27].

Other Issues

With regard to the pictures that contained only part of the plates, as well as the problems mentioned for packaged foods and beverages, we theorize that the users had misunderstood how the volume estimation works or were absolutely unaware that it even occurs. The most probable scenario is that they did not critically assess the importance of capturing the entire content, but only the packaging, or even part of it.

For the case of partially captured plates, we suppose that in our study, the users misunderstood the way the system processes the images, which requires the actual estimated food or beverage to be visible, so that its volume can be estimated. The same issue comes up quite often with the beverage recordings, where there are several recordings showing simply an opaque bottle whose content is unknown. We also believe that the requirements for the fiducial marker and that the entire plate should be in the photo might have posed a challenge to some users.

In future studies, the need to include the entire plate should be clearly stressed. In addition, it would be wise to present to the participants a more detailed structure of the system, explaining the step of volume estimation, so that the picture must contain everything the users are about to consume or have already consumed. Furthermore, it might be helpful if the app asked whether the user had given the answer, and if so, allow the user to modify it.

Strengths

One of the main strengths of our study design was the ability to test the app in real-life situations. Furthermore, this is one of the few studies that analyzes errors made by the users rather than focusing on aspects related to diet or accuracy in nutrition calculation. Moreover, our participants came from different educational backgrounds (students, workers, etc) and represented a wide range of ages, with different levels of technological familiarity. Another advantage was that the participants were not asked to alter their eating patterns to participate. For example, we did not advise them to exclude from their diet any packaged foods, although our system was not ready to process such data (work in progress) when the fieldwork was done. In line with our goal of keeping the procedure as noninvasive as possible and capturing a *normal* day of eating, we opted not to send participants any individualized text push messages assisting them with the image entry process. In fact, we only sent one email communication halfway through the study after participants reported difficulties and compliance issues with using the app to our team of computer scientists. Another strength of our study is the low dropout rate (4%), which may have resulted from the continuous support by our team on any technical problems reported.

Furthermore, it could be considered a strength that objective data were used to capture entry errors because all data containing mistakes were excluded by the trained experts instead of only the participants' subjective opinions. Contrary to another study [21], our participants were not encouraged to act without any limitations, and, for example, we did not let them take as many pictures as they wanted, as we saw this as a subjective opinion. Future research needs to combine both objective and subjective aspects to achieve a complete result. In the next stage, these studies could be compared in terms of acceptance rate, difficulties, and preferences.

Limitations

A key limitation of our study design was that the participants recorded their intake for only one 24-hour period. As a result, the evaluation of the progression of entry errors—whether they increased or decreased over time and with increased use of the app—was not possible. It would have been especially interesting to see if errors decreased after the midpoint email communication addressing participants' concerns over entry errors. Observing users over a longer period could help understand app acceptance over time, patterns of use, and changes to usage error rates. Another limitation of our study was the requirement of using a separate phone, as opposed to participants' own phones to take the pictures. Using an additional phone could have led to a higher number of entry errors, owing to different and potentially unfamiliar user interfaces. Carrying another phone might also have caused discomfort. We also cannot omit the possibility that some foods and beverages were not captured, most probably those consumed in between meals (eg, ready-to-eat snacks such as fruits or beverages such as water). However, it is not known if meals were skipped or if participants chose not to record some snacks or simply forgot them. As stated earlier, the app is not yet ready to analyze packaged food and beverages because this feature is currently under development. However, a barcode scanner could easily be integrated into the app with the assistance of an appropriate web or mobile stored database. Manual selection of the consumed beverage and the respective portion size from a list of beverages could also be easily integrated into goFOOD in a further stage of the app's development. Furthermore, technology savviness may have been an obstacle for older participants, and thus, their participation in our study was limited. As mentioned earlier, unfortunately, we did not have a representative sample of older participants, and thus, we cannot draw strong conclusions regarding age. Finally, another limitation of this study is that the sample consisted only of Caucasian, German-speaking people living in Switzerland, and thus, the results cannot be generalized to a wider population.

Suggestions for Future Work

Our state-of-the-art study is novel because no other studies have focused on errors in data derived from the use of automatic apps for nutrition assessment. This study showed that adequate instructions may be needed to learn how to correctly use image-based apps and that general technology literacy may not be enough when it comes to these types of apps. Future improvements to the app could focus on improving the recognition of food on various types of plates, that is, with different colors and shapes, as well as exploring alternatives to the use of fiducial markers. Moreover, it is of vital importance for future studies that we improve our understanding of the users' training needs as well as enhance the app's user-friendliness and develop an automatic image check feature based on participant feedback. Furthermore, future studies with older participants are necessary to obtain concrete results on

how age affects the number or severity of mistakes that are made using an image-based dietary assessment app.

This section outlines possible improvements to the app to reduce user errors. Improved navigation through the app, including some training material, could help the users to train and test themselves before the actual trial and thus eliminate basic mistakes. As an example, one participant mentioned “Maybe I took a photo of the meal too many different times or I did not take the correct picture because the app got stuck.” In this case, a system that checks and informs the user whether they have captured the picture correctly would be helpful. Another useful feature for the app might be the possibility of deleting an entry, especially in combination with a prompt from the app asking the user to try again because the “picture was not saved since the picture was not taken properly.” Text messages that also verify good lighting conditions would improve the image capturing process and usability of the app. A video tutorial at a variable pace could support those less apt or confident in using these kinds of apps. Furthermore, a section with frequently asked questions could help users troubleshoot on their own. Likewise, the integration of text messages at different stages in the data entry process could assist users and reduce errors. The suggestions offered here could be translated into different languages to ensure that users speaking different languages can fully understand the app prompts. Moreover, the app currently runs on smartphones using the Android operating system. The app could also be developed for the iOS system, and with this, the vast majority of smartphone users would be covered.

Finally, by analyzing user errors, we have learned the importance of integrating users in the design and development process of the goFOOD app. This makes sense, given that end users and health care professionals who benefit from such apps should have their needs considered throughout the entire development process with tools and techniques, such as extensive surveys and periodic trial tests to facilitate this process. Thus, research is essential to outline how and when these apps may most efficiently aid those needs [18].

Conclusions

To the best of our knowledge, this is the first research study that objectively analyzes user errors in the automation of food and nutritional recognition apps in real-life conditions. Error analysis thus yields novel results, identifying many forms of human errors in different steps in the process of entering meal information into the app. The analysis of the mistakes and omissions from this study is fundamental, as the knowledge obtained can be used to optimize the different aspects of the app and to accelerate the procedure for entering meals and shed light on areas of the app and user experience that require improvements. goFOOD was designed to be a functional app that can help the process of nutritional assessment by assisting health care professionals in their everyday practice. More specifically, our hope is that nutrition apps such as goFOOD could work both as a food log (goFOOD Lite) and as a dietary assessment app, thus reducing the time and effort required by conventional methods for assessing nutrition. Moreover, the exchange of data between the user and the dietitian will facilitate their coordination in tracking food intake. Last but not least, researchers who work in the field of acquisition of nutrition data or who plan epidemiological and clinical studies will benefit from our analysis, as they can learn which data with specific characteristics (human errors) should be omitted as well as improve their understanding of mitigation controls that they can integrate to improve study planning and data quality.

Acknowledgments

The authors would like to thank the volunteers of the study.

Authors' Contributions

MV, KH, and SM designed the study; SM is responsible for the goFOOD project; MJ and GT gathered the data; MV, TS, MJ, GT, and YL conducted the study; MV and TS conducted the data analysis; MV wrote the manuscript; KH, TS, YL, and SM made critical manuscript revisions; and SM had primary responsibility for the final content. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Evans R. IMS institute for healthcare informatics. Nurs Stand 2014 Apr 09;28(32):30-30. [doi: 10.7748/ns2014.04.28.32.30.s36]
2. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. Nutrition 2014;30(11-12):1257-1266. [doi: 10.1016/j.nut.2014.02.020] [Medline: 24976425]
3. Food and Agriculture Organization of the United Nations. Dietary Assessment: A Resource Guide to Method Selection and Application in Low Resource Settings. Rome: Food and Agriculture Organization; 2018.
4. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 09;17(1):5 [FREE Full text] [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]

5. Della Torre SB, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1):76 [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
6. Bejar LM, Sharp BN, García-Perea MD. The e-EPIDEMIOLOGY mobile phone app for dietary intake assessment: comparison with a food frequency questionnaire. JMIR Res Protoc 2016 Nov 02;5(4):e208 [FREE Full text] [doi: 10.2196/resprot.5782] [Medline: 27806922]
7. Conrad J, Nöthlings U. Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. Proc Nutr Soc 2017 Dec;76(3):213-219. [doi: 10.1017/S0029665116003025] [Medline: 28162110]
8. Reber E, Gomes F, Vasiloglou MF, Schuetz P, Stanga Z. Nutritional risk screening and assessment. J Clin Med 2019 Jul 20;8(7):1065 [FREE Full text] [doi: 10.3390/jcm8071065] [Medline: 31330781]
9. Archundia Herrera MC, Chan CB. Narrative review of new methods for assessing food and energy intake. Nutrients 2018 Aug 10;10(8) [FREE Full text] [doi: 10.3390/nu10081064] [Medline: 30103401]
10. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
11. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't VP. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. Br J Nutr 2009 Jul;101 Suppl 2:S73-S85. [doi: 10.1017/S0007114509990602] [Medline: 19594967]
12. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. Nutrition 2005 Jun;21(6):672-677. [doi: 10.1016/j.nut.2004.11.006] [Medline: 15925290]
13. Vasiloglou M, Mougiakakou S, Aubry E, Bokelmann A, Fricker R, Gomes F, et al. A Comparative Study on Carbohydrate Estimation: GoCARB vs. Dietitians. Nutrients 2018 Jun 07;10(6):741 [FREE Full text] [doi: 10.3390/nu10060741] [Medline: 29880772]
14. Bally L, Dehais J, Nakas CT, Anthimopoulos M, Laimer M, Rhyner D, et al. Carbohydrate estimation supported by the GoCARB system in individuals with type 1 diabetes: a randomized prospective pilot study. Diabetes Care 2017 Feb;40(2):e6-e7. [doi: 10.2337/dc16-2173] [Medline: 27899490]
15. Anthimopoulos M, Dehais J, Shevchik S, Ransford BH, Duke D, Diem P, et al. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. J Diabetes Sci Technol 2015 May;9(3):507-515 [FREE Full text] [doi: 10.1177/1932296815580159] [Medline: 25883163]
16. Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey RH, et al. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. J Med Internet Res 2016 May 11;18(5):e101 [FREE Full text] [doi: 10.2196/jmir.5567] [Medline: 27170498]
17. Dehais J, Anthimopoulos M, Shevchik S, Mougiakakou S. Two-view 3D reconstruction for food volume estimation. IEEE Trans Multimedia 2017 May;19(5):1090-1099. [doi: 10.1109/tmm.2016.2642792]
18. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. What Healthcare Professionals Think of "Nutrition & Diet" Apps: An International Survey. Nutrients 2020 Jul 24;12(8):2214 [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
19. Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, et al. goFOOD: an artificial intelligence system for dietary assessment. Sensors (Basel) 2020 Jul 31;20(15) [FREE Full text] [doi: 10.3390/s20154283] [Medline: 32752007]
20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Jul 30 Presented at: Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV, USA URL: https://ieeexplore.ieee.org/document/7780677 [doi: 10.1109/CVPR.2016.308]
21. He K. Mask R-CNN. In: Proceedings of IEEE International Conference on Computer Vision (ICCV). 2017 Presented at: IEEE International Conference on Computer Vision (ICCV); December25, 2017; Venice, Italy p. 2980-2988 URL: https://ieeexplore.ieee.org/document/8237584 [doi: 10.1109/iccv.2017.322]
22. Standard Reference 11 to 28. US Department of agriculture. 2016. URL: https://www.ars.usda.gov/Services/docs.htm?docid=8964 [accessed 2020-12-22]
23. Nutritionix Database. URL: https://www.nutritionix.com/database [accessed 2020-07-30]
24. Casperson SL, Sieling J, Moon J, Johnson L, Roemmich JN, Whigham L. A mobile phone food record app to digitally capture dietary intake for adolescents in a free-living environment: usability study. JMIR Mhealth Uhealth 2015;3(1):e30 [FREE Full text] [doi: 10.2196/mhealth.3324] [Medline: 25775506]
25. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. J Am Diet Assoc 2010 Jan;110(1):74-79 [FREE Full text] [doi: 10.1016/j.jada.2009.10.010] [Medline: 20102830]
26. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
27. Yoshikazu A, Takumi E, Jaehyeong C, Keiji Y. DepthCalorieCam: A Mobile Application for Volume-Based FoodCalorie Estimation using Depth Cameras. In: MADiMa '19. 2019 Presented at: 5th International Workshop on Multimedia Assisted

Dietary Management; October 2019; Nice, France URL: https://dl.acm.org/doi/10.1145/3347448.3357172 [doi: 10.1145/3347448.3357172]

Abbreviations

AI: artificial intelligence
mHealth: mobile Health

Edited by L Buis; submitted 21.09.20; peer-reviewed by I Spanakis, S Kriventsov; comments to author 08.10.20; revised version received 02.11.20; accepted 30.11.20; published 13.01.21.

Please cite as:
Vasiloglou MF, van der Horst K, Stathopoulou T, Jaeggi MP, Tedde GS, Lu Y, Mougiakakou S
The Human Factor in Automated Image-Based Nutrition Apps: Analysis of Common Mistakes Using the goFOOD Lite App
JMIR Mhealth Uhealth 2021;9(1):e24467
URL: http://mhealth.jmir.org/2021/1/e24467/
doi:10.2196/24467
PMID:33439139

Original Paper

A Self-Help App for Syrian Refugees With Posttraumatic Stress (Sanadak): Randomized Controlled Trial

Susanne Röhr[1,2*], PhD; Franziska U Jung[1*], PhD; Alexander Pabst[1], PhD; Thomas Grochtdreis[3], PhD; Judith Dams[3], PhD; Michaela Nagl[4], PhD; Anna Renner[4], MA; Rahel Hoffmann[4], MSc; Hans-Helmut König[3], MD, PhD; Anette Kersting[4], MD, PhD; Steffi G Riedel-Heller[1], MPH, MD, PhD

[1]Institute of Social Medicine, Occupational Health and Public Health, Medical Faculty, University of Leipzig, Leipzig, Germany

[2]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

[3]Department of Health Economics and Health Services Research, Hamburg Center for Health Economics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[4]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Leipzig, Leipzig, Germany

[*]these authors contributed equally

Corresponding Author:
Susanne Röhr, PhD
Institute of Social Medicine, Occupational Health and Public Health
Medical Faculty, University of Leipzig
Philipp-Rosenthal-Strasse 55
Leipzig
Germany
Phone: 49 341 97 24568
Email: susanne.roehr@medizin.uni-leipzig.de

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2021/3/e28336/

Abstract

Background: Syrian refugees residing in Germany often develop posttraumatic stress as a result of the Syrian civil war, their escape, and postmigration stressors. At the same time, there is a lack of adequate treatment options. The smartphone-based app Sanadak was developed to provide cognitive behavioral therapy–based self-help in the Arabic language for Syrian refugees with posttraumatic stress.

Objective: The aim of this study was to evaluate the effectiveness and cost-effectiveness of the app.

Methods: In a randomized controlled trial, eligible individuals were randomly allocated to the intervention group (IG; app use) or control group (CG; psychoeducational reading material). Data were collected during structured face-to-face interviews at 3 assessments (preintervention/baseline, postintervention/after 4 weeks, follow-up/after 4 months). Using adjusted mixed-effects linear regression models, changes in posttraumatic stress and secondary outcomes were investigated as intention-to-treat (ITT) and per-protocol (PP) analysis. Cost-effectiveness was evaluated based on adjusted mean total costs, quality-adjusted life years (QALYs), and cost-effectiveness acceptability curves using the net benefit approach.

Results: Of 170 screened individuals (aged 18 to 65 years), 133 were eligible and randomized to the IG (n=65) and CG (n=68). Although there was a pre-post reduction in posttraumatic stress, ITT showed no significant differences between the IG and CG after 4 weeks (Posttraumatic Diagnostic Scale for DSM-5, Diff –0.90, 95% CI –0.24 to 0.47; *P*=.52) and after 4 months (Diff –0.39, 95% CI –3.24 to 2.46; *P*=.79). The same was true for PP. Regarding secondary outcomes, ITT indicated a treatment effect for self-stigma: after 4 weeks (Self-Stigma of Mental Illness Scale/SSMIS–stereotype agreement: *d*=0.86, 95% CI 0.46 to 1.25; stereotype application: *d*=0.60, 95% CI 0.22 to 0.99) and after 4 months (*d*=0.52, 95% CI 0.12 to 0.92; *d*=0.50, 95% CI 0.10 to 0.90), the IG showed significantly lower values in self-stigma than the CG. ITT showed no significant group differences in total costs and QALYs. The probability of cost-effectiveness was 81% for a willingness-to-pay of €0 per additional QALY but decreased with increasing willingness-to-pay.

Conclusions: Sanadak was not more effective in reducing mild to moderate posttraumatic stress in Syrian refugees than the control condition nor was it likely to be cost-effective. Therefore, Sanadak is not suitable as a standalone treatment. However, as the app usability was very good, no harms detected, and stigma significantly reduced, Sanadak has potential as a bridging aid within a stepped and collaborative care approach.

Trial Registration: German Clinical Trials Register DRKS00013782; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00013782

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-019-2110-y

(JMIR Mhealth Uhealth 2021;9(1):e24807) doi:10.2196/24807

KEYWORDS

app; cost-utility analysis; mHealth; posttraumatic stress; PTSD; quality-adjusted life years; randomized controlled trial; refugees; stimga; Syrian refugees; usability

Introduction

According to the United Nations Refugee Agency in 2019, the unprecedented number of 26 million individuals worldwide have been seeking shelter as refugees [1]. As a result of displacement and adverse associated experiences such as torture, trauma, and loss, refugees have an increased risk of mental ill-health [2]. In Germany, Syrians who have escaped the ongoing civil war since 2010-2011 represent the largest group among refugees. Studies have shown that Syrian refugees were typically exposed to potentially traumatizing events, increasing their vulnerability to posttraumatic stress and comorbid mental health outcomes [2]. The most frequently reported disorders associated with war and escape are posttraumatic stress disorder (PTSD) and major depression, often accompanied by somatization [3]. For example, among 518 adult Syrian refugees in Germany, 75.3% reported having witnessed and/or experienced traumatic events and 11.4% had symptoms of PTSD [2]. Furthermore, moderate to severe depression was present in 14.5% and moderate to severe generalized anxiety in 13.5% of Syrian adult refugees [2].

Current guidelines on PTSD treatment by the German Association of the Scientific Medical Societies indicate that trauma adaptive psychotherapy should be offered in a timely manner [1]. However, particularly for refugees, several barriers to treatment exist, including language and cultural barriers, legal and health insurance regulations in regard to asylum, and lack of psychoeducation [2]. In addition, proactive uptake of mental health care is low among refugees [3].

Research suggests that treatment of PTSD is effective [4]; however, in the context of the high prevalence of PTSD symptomatology among refugees, health care systems in host countries may often not have enough resources to cover the need for treatment. Therefore, eHealth interventions have been suggested as a means to close the treatment gap [5]. Moreover, as PTSD is associated with high costs of outpatient treatment, nonphysician outpatient contacts, and psychiatric contacts [6], app-based interventions could be a cost-effective alternative. Therefore, Sanadak, a smartphone-based interactive low-threshold self-help app in the Arabic language, has been developed based on evidence-driven cognitive behavioral therapy for PTSD [7]. During the development of the app, typical themes and needs of refugees as well as cultural specifics were incorporated. Therefore, focus groups were conducted to assess relevant aspects (eg, concepts of disease and disease management), which have been found to be highly recommended in comparison with traditional mental health interventions [8,9]. The content of the Sanadak app is multimodal (ie, it includes psychoeducational information to increase knowledge and awareness of PTSD and related mental health issues and self-help techniques and skills training for symptom management). In addition, a short self-test on posttraumatic symptom severity was implemented to allow for automated tailored feedback regarding progress at any time. Interactive materials, such as animated videos and audios as well as games and exercises are provided to maximize usability. Further information is detailed in the study protocol [7]. To the best of our knowledge, there is currently no comparable multimodal app intervention available for this target group that has been evaluated in terms of its effectiveness and cost-effectiveness in a randomized controlled trial (RCT). The primary aim was to evaluate the app's effectiveness in reducing posttraumatic stress symptoms, which we hypothesized to be superior to the control condition.

Methods

Study Context Information

The study design is detailed in the study protocol, published elsewhere [7]. The trial was registered with the German Clinical Trials Register [DRKS00013782] on July 6, 2018. Study results are reported according to the Consolidated Standards of Reporting Trials (CONSORT) statement [10]. The app was developed by frühlingsproduktionen, a creator of eHealth interventions based in Berlin, Germany, on behalf of the study principal investigator. Sanadak app versions 1.4.0 and 1.5.0 were evaluated, first released on October 29, 2018 (no major changes between versions).

Trial Design

After screening for eligibility, study participants were randomly allocated (1:1) to the intervention group (IG) or control group (CG), which received a psychoeducational brochure. In order to test short- as well as medium-term treatment effects, 3 face-to-face interviews were scheduled with the study participants: baseline (T0: pre), immediately after the intervention (T1: post, 4 weeks after baseline), and 4 months after baseline (T2: follow-up).

Ethics and Guidelines

The trial was approved by the ethics committee of the Medical Faculty of the University of Leipzig, Germany (ID: 111-17-ek) and adheres to the Declaration of Helsinki and the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines for good clinical practice. All participants were informed about the study aims, including clarification about data protection measures and data security according to latest legal standards. Participation was only allowed after written informed consent.

Participants

By using a multistrategic approach to recruit Syrian refugees residing in the urban areas of Leipzig, Halle/Saale, and Dresden in Germany, potential study participants were attracted. This included bilingual posters and brochures, building contact with specific multipliers and institutions, a snowball sampling approach, and use of social media and personal contacts of the native Arabic-speaking study personnel. Recruitment strategies have been described in detail elsewhere [11]. Eligibility according to prespecified inclusion and exclusion criteria was checked during a face-to-face screening. Inclusion criteria comprised being Syrian refugee residing in Germany, aged 18 to 65 years, experiencing at least one traumatic event and subsequent mild to moderate posttraumatic stress symptom severity (score of 11 to 59) on the Posttraumatic Diagnostic Scale for DSM-5 (PDS-5) [12], and owning a compatible device to use the app (Android/iOS). Exclusion criteria included posttraumatic stress symptomatology outside of the range mentioned above; severe depressive symptoms (Patient Health Questionnaire [PHQ-9] ≥20) [13]; acute suicidal tendencies (Depressive Symptom Inventory–Suicidality Subscale [DSI-SS] ≥3) [14]; current psychotherapy, psychiatric treatment, and/or psychotropic medication; or pregnancy. If individuals were not eligible due to severity of symptoms, they received psychoeducational material on mental health care and contact information of regional initiatives that offer face-to-face support. A detailed report of the recruitment and baseline characteristics has been published previously [15].

Interventions

Participants in the IG had the opportunity to use the self-help app via person-specific log-in data (to avoid group contamination) for 4 weeks on demand. They were advised to use the app regularly and work through the modules. Participants in the CG received psychoeducational reading material in the Arabic language covering traumatization and posttraumatic stress (identical to the information delivered by the app).

Assessments and Outcomes

In addition to sociodemographic characteristics such as age, gender, and family status, information on residence status, employment, religious beliefs, and escape-related information were collected during standardized interviews at baseline.

The primary outcome was posttraumatic stress, measured by the PDS-5 [12]. Secondary outcomes included symptoms of depression (PHQ-9) [13]; generalized anxiety (Generalized Anxiety Disorder [GAD-7]) [16]; somatic symptoms (Physical Health Questionnaire [PHQ-15]) [17,18]; general self-efficacy (GSE) [19]; self-stigma (Self-Stigma of Mental Illness Scale–Short Form [SSMIS-SF], stereotype awareness [SSMIS-AW], stereotype agreement [SSMIS-AG], stereotype application [SSMIS-AP], and harm to self-esteem [SSMIS-HS]) [20]; resilience (Resilience Scale [RS-13]) [21]; social isolation (short form of the Lubben Social Network Scale [LSNS-6]) [22]; social support (ENRICHD Social Support Inventory [ESSI]) [23]; health-related quality of life (EuroQoL 5-Dimension 5-Level [EQ-5D-5L] and visual analog scale [EQ-VAS]) [24,25]; and posttraumatic growth (Posttraumatic Growth Inventory [PGI]) [26]. If not available, instruments were translated from German into the Arabic language using the Translation, Review, Adjudication, Pretesting, and Documentation procedure [27] involving native Arabic-speaking experts. All outcomes were assessed at T0, T1, and T2.

For the cost-effectiveness analysis, health service utilization was assessed retrospectively over 4 months at T0 and T2, using an adapted and shortened version of the German Client Socio-Demographic and Service Receipt Inventory [28]. The questionnaire covered inpatient care, rehabilitation, and outpatient physician and nonphysician services. Costs were calculated from a health care payer perspective in euros (€) for the year 2019. Monetary valuation of used health care services was conducted using standardized unit costs within the German health care system [29]. Intervention costs consisted of the costs for technical support for the Sanadak app during follow-up. Health effects were quantified by quality-adjusted life years (QALYs) observed during the 4-month follow-up period calculated by linearly interpolating EQ-5D-5L index scores from baseline to follow-up [25,30].

In the IG, we furthermore assessed information on app usability (System Usability Scale [SUS] [31]). Furthermore, deidentified metadata of the app use (ie, duration of app use during evaluation period, in minutes) stored in the app's log files were collected.

Last, a standardized assessment at T1 and T2 was implemented to monitor potential harms due to trial participation. Harms/negative effects were defined as adverse events (AE) and severe adverse events (SAE). AE comprised an increase in target symptoms (posttraumatic and depressive symptomatology, suicidality), occurrence of novel psychological or physical symptoms, any negative events, all of which may or may not be associated with trial participation in both IG and CG. SAE were defined as events that require some form of high-intensity treatment (ie, deliberate self-harm, suicide attempt, life-threatening events, nonelective or extended hospitalization, an event causing chronic or severe disability) or fatality, including suicide.

Sample Size

The sample size was calculated based on recent evaluations regarding the efficacy of telemedical-based treatment of PTSD symptoms [32]. Given a moderate between-group effect at follow-up 1 (Cohen d=0.5), a significance level of α=.05 (1-sided), and a statistical power of 1–β=0.80, optimal sample size for estimating a significant treatment effect is n=102. Considering attrition due to different circumstances (eg, change in residence status, trial termination by the participant), a

dropout rate of approximately 20% may be expected, suggesting a sufficient baseline sample size of n=128 participants.

Randomization and Masking

After participants were found to be eligible for participation, they were randomly assigned to IG or CG using a 1:1 ratio using randomized permuted blocks of 6, stratified by age and sex, which ensured both balance in sample size across groups and control of important covariates. An external, independent statistician generated the randomization block lists with a respective computer program (blockrand package written for R [R Foundation for Statistical Computing]).

The study coordinator (SR), responsible for individual group allocation, remained blind to the randomization list strata identity. Moreover, the data analyst (AP), who conducted the primary analysis concerning the hypothesized group differences (IG vs CG) in primary and secondary outcome measures (see above), was blind to group assignment.

Data Management, Data Protection, and Quality Control

With regard to interview assessments, deidentified data entry took place immediately after each interview using the statistical software SPSS Statistics version 24 (IBM Corp), generating a password-protected and locally stored file with access granted only to study personnel who had signed data protection wavers. Concurrently, data completeness and consistency checks were conducted to ensure data integrity. With regard to data on app use, anonymous log-file data were collected during participant interaction with the app, stored at the app developer's secured server, and collectively transferred upon completion of all intervention periods to the study investigator using Secure Sockets Layer technology to ensure data encryption and protection. In fact, specific arrangements regarding data protection (eg, no real names for program log-in) were prespecified in a data protection concept. The data protection concept detailed data handling of all collected data (ie, interview data at screening, T0, T1, and T2 and metadata of the app use) with the purpose to ensure compliance with the General Data Protection Regulation (GDPR) of the European Union. The GDPR specifies the lawful processing of personal data. The data protection concept was composed with and approved by an external lawyer specializing in data protection.

As an additional measure of data quality, auditing took place in the form of independent source data verification, performed by commissioned external statisticians. Specifically, 5% of the questionnaires at T0, T1, and T2 (source data) were randomly drawn and inspected regarding their degree of matching with the electronic data file.

Statistical Analysis

Quality checks on baseline data revealed missing information on both outcome variables and covariates. Frequency of missing values was low (ie, 5/133 cases or less) for all variables but high for education (24/133 cases) and summed up to 27.8% for the set of baseline characteristics. Since sensitivity analysis showed that missing values were not completely at random, a complete case analysis was inappropriate. Therefore, we multiple-imputed missing baseline data using the algorithm of chained equations implemented in Stata (StataCorp LLC) with all sociodemographic variables and baseline assessments of outcome variables as predictors. The resulting pooled estimates of 25 imputed datasets were used for all analyses.

Primary analysis of trial data was intention-to-treat (ITT) as outlined in the guidelines of the CONSORT statement and its supplement for reporting eHealth trials [33]. In order to evaluate treatment effect, multilevel mixed-effects linear regression models were used, since all outcomes were approximately normally distributed. The models included an indicator of treatment group, time, and an interaction between treatment group and time as fixed effects and comprised a random intercept to control for potential heterogeneity within participants over time. All models were further adjusted for the baseline outcome score, as well as for age, gender, education (low/middle/high according to the Comparative Analysis of Social Mobility in Industrial Nations classification of education [34]), marital status, living situation, residence permit, employment, income, religious beliefs, and duration of residence in Germany as these covariates were considered to be prognostic in relation to the outcomes.

For sensitivity analysis, we performed a per-protocol (PP) analysis by excluding participants who had not used the intervention as indicated by deidentified log-in data. Mixed-effects linear regression models for both primary and secondary outcomes were also used on that restricted sample following the same procedure outlined above. The above described analyses were repeated with regard to subgroups (age: 18 to 29 years, 30 to 39 years, 40+ years; gender: male, female; level of education: low, medium, high; frequency of app use: 1 to 42.5 minutes, 42.5+ minutes; posttraumatic stress symptom severity: low, moderate).

To analyze the cost-effectiveness of Sanadak compared with receiving psychoeducational reading material on trauma and PTSD only, adjusted total costs, QALYs, and differences in total costs and QALYs between the IG and CG were calculated using mixed-effects linear regression models with robust standard errors adjusted for age, sex, costs, EQ-5D-5L indices, and PDS-5 indices at baseline. Furthermore, cost-effectiveness acceptability curves of Sanadak based on the net benefit approach were calculated accordingly by multilevel mixed-effects linear regressions based on the net monetary benefit for different willingness-to pay (WTP) thresholds [35].

Descriptive data are presented as number of observations with percentages or means with corresponding standard deviations or 95% confidence intervals. Results of the mixed-effects regression models are presented as adjusted mean differences and 95% confidence intervals in primary and secondary outcome scores between the treatment groups at both follow-ups. We also report standard effect sizes of treatment on all study outcomes at follow-up (Cohen *d*) using adjusted mean scores from the mixed models. All analyses were performed using the Stata 16.1 SE software package (StataCorp LLC). A *P* value of <.05 was considered statistically significant.

Results

Participants

Participants were recruited between October 2018 and December 2019, and follow-up was completed in April 2020. Figure 1 shows the flow of participants through the trial. Altogether, 170 individuals were assessed for eligibility, of which 37 were excluded as they did not meet inclusion criteria (n=32, 86.5%) or as they declined further participation (n=5, 13.5%). The eligible 133 individuals were randomly allocated either to the IG or CG. Table 1 shows baseline characteristics of the participants. The IG sample consisted of 65 participants with a mean age of 33.0 (SD 11.0, range 18 to 65) years; 66.2% (43/65) were male. The average age of the control group (n=68) was 33.7 (SD 11.4, range 18 to 65) years; 57.4% (39/68) were male. No significant differences were found for key sociodemographic variables or other baseline variables between the two groups, except for higher unemployment in the CG, indicating that the randomization was successful.

Figure 1. Flowchart of sample selection and attrition in the Sanadak trial.

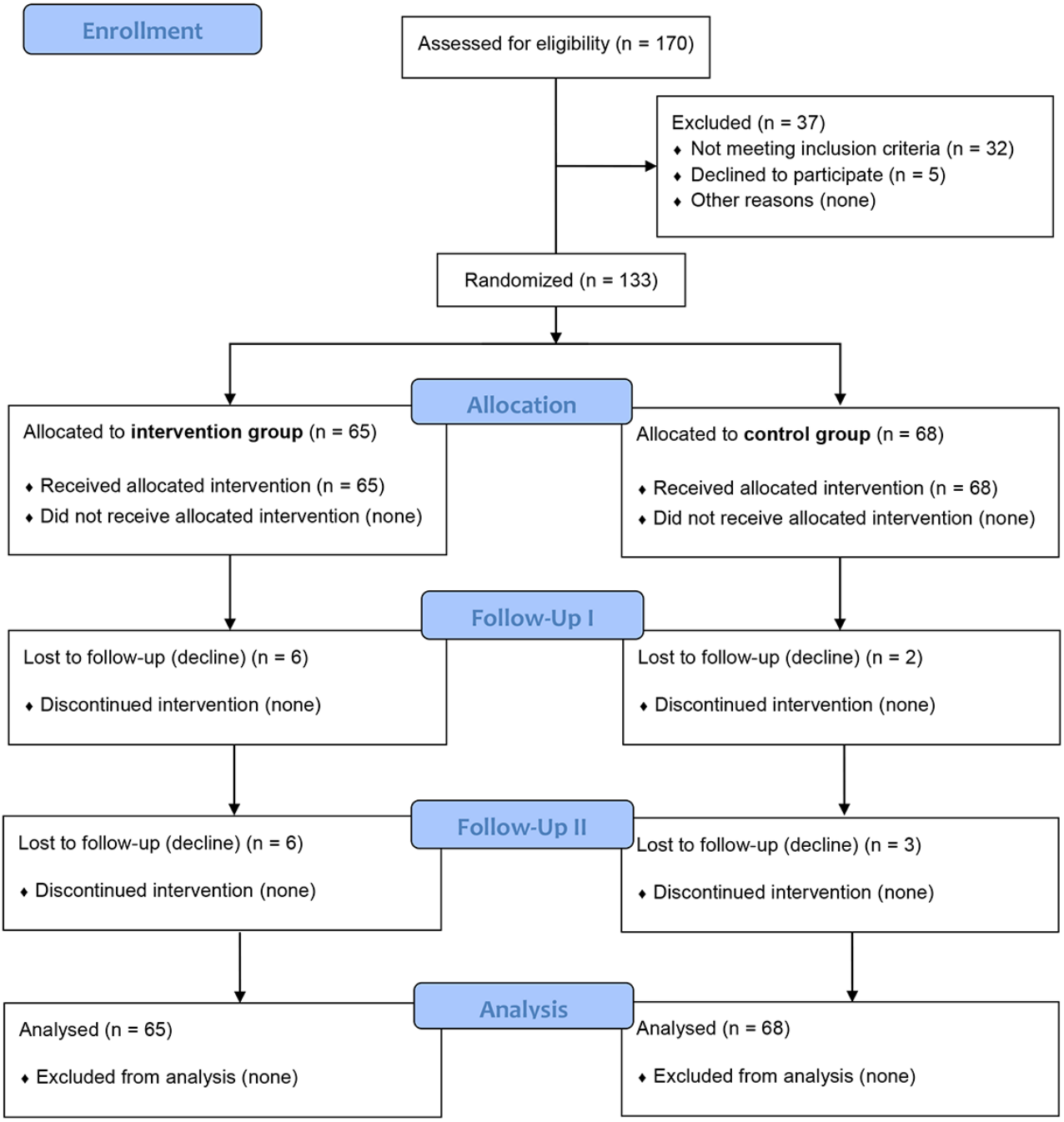


Table 1. Baseline characteristics of study participants of the Sanadak trial (n=133).

Characteristics	CG[a] (n=68)	IG[b] (n=65)
Male	39 (57.4)	43 (66.2)
Age in years, mean (SD)	33.67 (11.4)	32.98 (11.0)
Marital status, n (%)		
Married	30 (44.1)	21 (32.3)
Unmarried/divorced/widowed	38 (55.9)	44 (67.7)
Living situation, n (%)		
Alone	15 (22.1)	18 (27.7)
With family	38 (55.9)	35 (53.9)
With others	15 (22.1)	12 (18.5)
Education, n (%)		
Low	14 (20.6)	17 (26.2)
Medium	30 (44.1)	21 (32.3)
High	24 (35.3)	27 (41.5)
Employment, n (%)		
Employed	13 (19.1)	23 (35.4)
Unemployed	55 (80.9)	42 (64.6)
Income per month (€), n (%)		
<500	13 (19.1)	14 (21.5)
500-1000	40 (58.8)	33 (50.8)
>1000	15 (22.2)	18 (27.7)
Residence permit, n (%)		
Entitlement to political asylum	39 (57.4)	35 (53.9)
Refugee protection	16 (23.5)	12 (18.5)
Other (subsidiary/visa/reunification)	13 (19.1)	18 (27.7)
Religious beliefs, mean (SD)	15.01 (5.7)	14.95 (4.9)
Duration of residence in years, mean (SD)	3.53 (1.3)	3.49 (1.1)

[a]CG: control group.

[b]IG: intervention group.

Attrition

A total of 6.0% (8/133) of the participants were lost to follow-up at the end of 4 weeks (T1). In all cases, participants refused further participation. Another 6.8% (9/133) of the sample did not complete assessments at T2. Two of those refused further participation; in 7 cases, repeated contact attempts failed. Total attrition was 12.8% (17/133). There were no significant differences between participants that did and did not complete any of the follow-up interviews nor were there significant differences between conditions in participating at T1 and T2.

Effectiveness of the Intervention

Table 2 shows average scores for the primary (PDS-5) and secondary outcome variables (PHQ-9, GAD-7, PHQ-15, EQ-5D-5L, EQ-VAS, GSE, SSMIS-AW, SSMIS-AG, SSMIS-AP, SSMIS-HS, LNSN, ESSI, and PGI) at baseline and follow-up. Table 3 shows results of the analysis of the ITT sample for all outcomes. There were no significant differences in PTSD symptoms between the IG and CG after 4 weeks (PDS-5, Diff –0.90, 95% CI –0.24 to 0.47; *P*=.52) and after 4 months (PDS-5, Diff –0.39, 95% CI –3.24 to 2.46; *P*=.79; Figure 2). The same was true for secondary outcomes except self-stigma (Figure 3). Table 4 shows results of the PP analysis for all outcomes that scarcely differed from those in Table 3. Again, significant group differences in PTSD symptoms between the two groups were found neither after 4 weeks (PDS-5, Diff 0.10, 95% CI –2.77 to 2.98; *P*=.95) nor after 4 months (PDS-5, Diff 0.56, 95% CI –3.56 to 2.45; *P*=.72). A significant difference was found for self-stigma (SSMIS-AG and SSMIS-AP). After 4 weeks (SSMIS-AG, *d*=0.86, 95% CI 0.46 to 1.25; SSMIS-AP, *d*=0.60, 95% CI 0.22 to 0.99) and after 4 months (SSMIS-AG, *d*=0.52, 95% CI 0.12 to 0.92; SSMIS-AP, *d*=0.50, 95% CI 0.10 to 0.90), the IG showed significantly lower values in self-stigma than the CG (Figure 4).

Table 2. Average scores of primary and secondary outcome variables at baseline and follow-up.

Outcome	CG[a] (n=68)			IG[b] (n=65)		
	n	Mean	95% CI	n	Mean	95% CI
Primary outcome						
PDS-5[c]						
Baseline	68	24.43	21.43-27.42	65	23.18	20.51-25.86
4 weeks	65	21.75	18.48-25.03	59	19.66	16.61-22.72
4 months	61	17.15	14.43-19.87	53	15.79	12.91-18.68
Secondary outcome						
PHQ-9[d]						
Baseline	68	9.34	7.96-10.71	65	9.15	7.97-10.34
4 weeks	66	8.52	7.21-9.82	59	7.90	6.69-9.11
4 months	61	7.41	4.91-7.38	53	6.79	5.48-8.11
GAD-7[e]						
Baseline	68	8.84	7.50-10.17	65	8.23	7.15-9.31
4 weeks	66	6.98	5.73-8.24	59	6.56	5.43-7.69
4 months	62	6.15	4.91-7.38	52	5.77	4.74-6.80
PHQ-15[f]						
Baseline	68	8.54	7.31-9.78	65	8.80	7.49-10.11
4 weeks	66	7.44	6.22-8.66	59	6.93	5.71-8.15
4 months	60	6.37	5.13-7.60	53	6.30	4.93-7.67
ED-5D-5L[g]						
Baseline	68	0.80	0.74-0.85	65	0.86	0.82-0.89
4 weeks	66	0.86	0.82-0.91	59	0.88	0.85-0.92
4 months	63	0.88	0.83-0.92	53	0.90	0.88-0.93
Health status						
Baseline	68	72.53	67.47-77.59	65	74.23	70.03-78.43
4 weeks	66	72.98	68.85-77.12	59	78.53	74.74-82.31
4 months	63	75.92	71.47-80.37	53	74.77	70.58-78.96
GSE[h]						
Baseline	68	28.01	27.04-28.99	65	26.78	25.49-28.08
4 weeks	65	27.88	26.78-28.97	59	26.39	25.05-27.73
4 months	63	28.57	27.49-29.66	53	26.45	25.25-27.66
SSMIS-AW[i]						
Baseline	68	28.63	26.62-30.65	65	27.82	26.18-29.45
4 weeks	65	28.49	26.71-30.28	59	27.88	25.78-29.98
4 months	63	27.43	25.61-29.25	53	28.09	26.13-30.06
SSMIS-AG[j]						
Baseline	68	17.69	16.12-19.26	65	20.05	18.37-21.72
4 weeks	64	17.89	16.36-19.42	59	16.32	14.68-17.97
4 months	62	17.92	16.18-19.66	53	18.23	16.46-19.99
SSMIS-AP[k]						

Outcome	CG[a] (n=68)			IG[b] (n=65)		
	n	Mean	95% CI	n	Mean	95% CI
Baseline	68	15.31	13.82-16.80	65	17.85	16.16-19.53
4 weeks	66	15.73	13.90-17.56	59	13.95	12.40-15.50
4 months	63	16.13	14.40-17.85	53	15.70	13.73-17.67
SSMIS-HS[l]						
Baseline	68	18.19	15.88-20.51	65	19.22	16.83-21.60
4 weeks	65	17.65	15.29-20.01	58	16.40	14.26-18.53
4 months	62	21.50	18.99-24.01	53	20.3	18.32-22.29
LSNS-6[m]						
Baseline	68	15.21	13.96-16.46	65	14.95	13.60-16.31
4 weeks	66	13.38	12.16-14.60	59	14.14	12.72-15.56
4 months	63	13.56	12.45-14.67	53	14.13	12.83-15.44
ESSI[n]						
Baseline	68	17.66	16.39-18.93	65	18.37	17.35-19.38
4 weeks	66	17.20	15.91-18.48	58	18.43	17.09-19.77
4 months	63	17.52	16.40-18.65	52	18.65	17.44-19.87
PGI[o]						
Baseline	68	24.94	23.40-26.48	65	22.75	21.10-24.40
4 weeks	65	23.69	21.87-25.52	59	22.07	19.85-24.29
4 months	61	24.34	22.38-26.31	53	20.89	18.99-22.79
Total costs						
Baseline	68	507.95	159.21-856.69	65	349.46	224.10-474.82
4 months	57	551.85	255.59-848.11	50	306.88	201.76-412.01

[a]CG: control group.

[b]IG: intervention group.

[c]PDS-5: Posttraumatic Diagnostic Scale for DSM-5.

[d]PHQ-9: Patient Health Questionnaire.

[e]GAD-7: Generalized Anxiety Disorder.

[f]PHQ-15: Physical Health Questionnaire.

[g]ED-5D-5L: EuroQoL 5-Dimension 5-Level.

[h]GSE: General Self-Efficacy.

[i]SSMIS-AW: Self-Stigma of Mental Illness Scale–stereotype awareness.

[j]SSMIS-AG: Self-Stigma of Mental Illness Scale–stereotype agreement.

[k]SSMIS-AP: Self-Stigma of Mental Illness Scale–stereotype application.

[l]SSMIS-HS: Self-Stigma of Mental Illness Scale–harm to self-esteem.

[m]LSNS-6: Lubben Social Network Scale.

[n]ESSI: ENRICHD Social Support Inventory.

[o]PGI: Posttraumatic Growth Inventory.

Table 3. Results of the Sanadak trial based on intention-to-treat analysis.

Outcome	n	Diff[a]	95% CI	*P* value	*d*	95% CI
Primary outcome						
PDS-5[b]						
Baseline	—	—	—	—	—	—
4 weeks	124	–0.901	–3.624 to 1.823	.52	0.118	–0.235 to 0.471
4 months	114	–0.390	–3.238 to 2.457	.79	0.051	–0.317 to 0.419
Secondary outcome						
PHQ-9[c]						
Baseline	—	—	—	—	—	—
4 weeks	125	–0.176	–1.343 to 0.991	.77	0.054	–0.298 to 0.405
4 months	114	–0.255	–1.480 to 0.970	.68	0.078	–0.291 to 0.446
GAD-7[d]						
Baseline	—	—	—	—	—	—
4 weeks	125	0.317	–0.752 to 1.386	.56	–0.106	–0.458 to 0.246
4 months	114	0.541	–0.582 to 1.664	.35	–0.180	–0.549 to 0.190
PHQ-15[e]						
Baseline	—	—	—	—	—	—
4 weeks	125	–0.463	–1.373 to 0.447	.32	0.181	–0.171 to 0.533
4 months	113	0.080	–0.879 to 1.039	.87	–0.031	–0.401 to 0.338
ED-5D-5L[f]						
Baseline	—	—	—	—	—	—
4 weeks	125	–0.027	–0.071 to 0.017	.23	0.219	–0.133 to 0.571
4 months	116	–0.019	–0.065 to 0.026	.40	0.158	–0.208 to 0.524
Health status						
Baseline	—	—	—	—	—	—
4 weeks	125	5.254	0.830 to 9.679	.02	–0.421	–0.776 to –0.065
4 months	116	–1.165	–5.783 to 3.452	.62	0.093	–0.272 to 0.459
GSE[g]						
Baseline	—	—	—	—	—	—
4 weeks	124	5.254	0.830 to 9.679	.44	0.140	–0.213 to 0.493
4 months	116	–0.905	–2.043 to 0.233	.12	0.295	–0.073 to 0.661
SSMIS-AW[h]						
Baseline	—	—	—	—	—	—
4 weeks	124	0.203	–1.901 to 2.308	.85	–0.034	–0.387 to 0.318
4 months	116	1.563	0.624 to 3.749	.16	–0.265	0.631 to 0.103
SSMIS-AG[i]						
Baseline	—	—	—	—	—	—
4 weeks	122	–3.587	–5.355 to –1.819	<.001	0.732	0.364 to 1.097
4 months	114	–1.813	–3.655 to 0.028	.05	0.369	–0.003 to 0.739
SSMIS-AP[j]						
Baseline	—	—	—	—	—	—

Outcome	n	Diff[a]	95% CI	*P* value	*d*	95% CI
4 weeks	125	–3.658	–5.543 to –1.773	<.001	0.691	0.329 to 1.052
4 months	116	–2.743	–4.712 to 0.774	.01	0.517	0.145 to 0.888
SSMIS-HS[k]						
Baseline	—	—	—	—	—	—
4 weeks	123	–1.370	–4.069 to 1.329	.32	0.182	–0.173 to 0.536
4 months	115	–1.938	–4.739 to 0.863	.18	0.257	–0.112 to 0.625
LSNS-6[l]						
Baseline	—	—	—	—	—	—
4 weeks	125	0.804	–0.471 to 2.079	.22	–0.224	–0.576 to 0.128
4 months	116	0.310	–1.016 to 1.636	.65	–0.087	–0.452 to 0.279
ESSI[m]						
Baseline	—	—	—	—	—	—
4 weeks	124	0.520	–0.741 to 1.781	.42	–0.147	–0.500 to 0.206
4 months	115	0.110	–1.213 to 1.432	.87	–0.031	–0.398 to 0.336
PGI[n]						
Baseline	—	—	—	—	—	—
4 weeks	122	0.233	–1.517 to 2.037	.80	–0.047	–0.402 to 0.309
4 months	111	–1.217	–3.110 to 0.676	.21	0.243	–0.131 to 0.617

[a]Diff: difference.
[b]PDS-5: Posttraumatic Diagnostic Scale for DSM-5.
[c]PHQ-9: Patient Health Questionnaire.
[d]GAD-7: Generalized Anxiety Disorder.
[e]PHQ-15: Physical Health Questionnaire.
[f]ED-5D-5L: EuroQoL 5-Dimension 5-Level.
[g]GSE: General Self-Efficacy.
[h]SSMIS-AW: Self-Stigma of Mental Illness Scale–stereotype awareness.
[i]SSMIS-AG: Self-Stigma of Mental Illness Scale–stereotype agreement.
[j]SSMIS-AP: Self-Stigma of Mental Illness Scale–stereotype application.
[k]SSMIS-HS: Self-Stigma of Mental Illness Scale–harm to self-esteem.
[l]LSNS-6: Lubben Social Network Scale.
[m]ESSI: ENRICHD Social Support Inventory.
[n]PGI: Posttraumatic Growth Inventory.

Figure 2. Adjusted predictions of primary outcome of posttraumatic stress symptoms (total score of the Posttraumatic Diagnostic Scale for DSM-5) at baseline and 4 weeks (T1) and 4 months (T2) after baseline in the intervention group and control group in the Sanadak trial, an Arabic language self-help app for Syrian refugees with posttraumatic stress, using mixed-effects linear regression models.

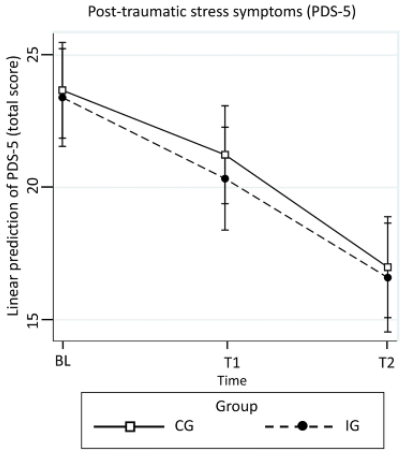


Figure 3. Adjusted predictions of secondary outcomes at baseline and 4 weeks (T1) and 4 months (T2) after baseline in the intervention group and control group in the Sanadak trial, an Arabic language self-help app for Syrian refugees with posttraumatic stress, using mixed-effects linear regression models. EQ-5D-5L: EuroQoL 5-Dimensions Questionnaire; EQ-VAS: EuroQoL Visual Analog Scale; ESSI: ENRICHD Social Support Inventory; GAD: General Anxiety Disorder; GSE: General Self-Efficacy; LSNS: Lubben Social Network Scale; PGI: Posttraumatic Growth Inventory; PHQ-9/-15: Patient Health Questionnaire.

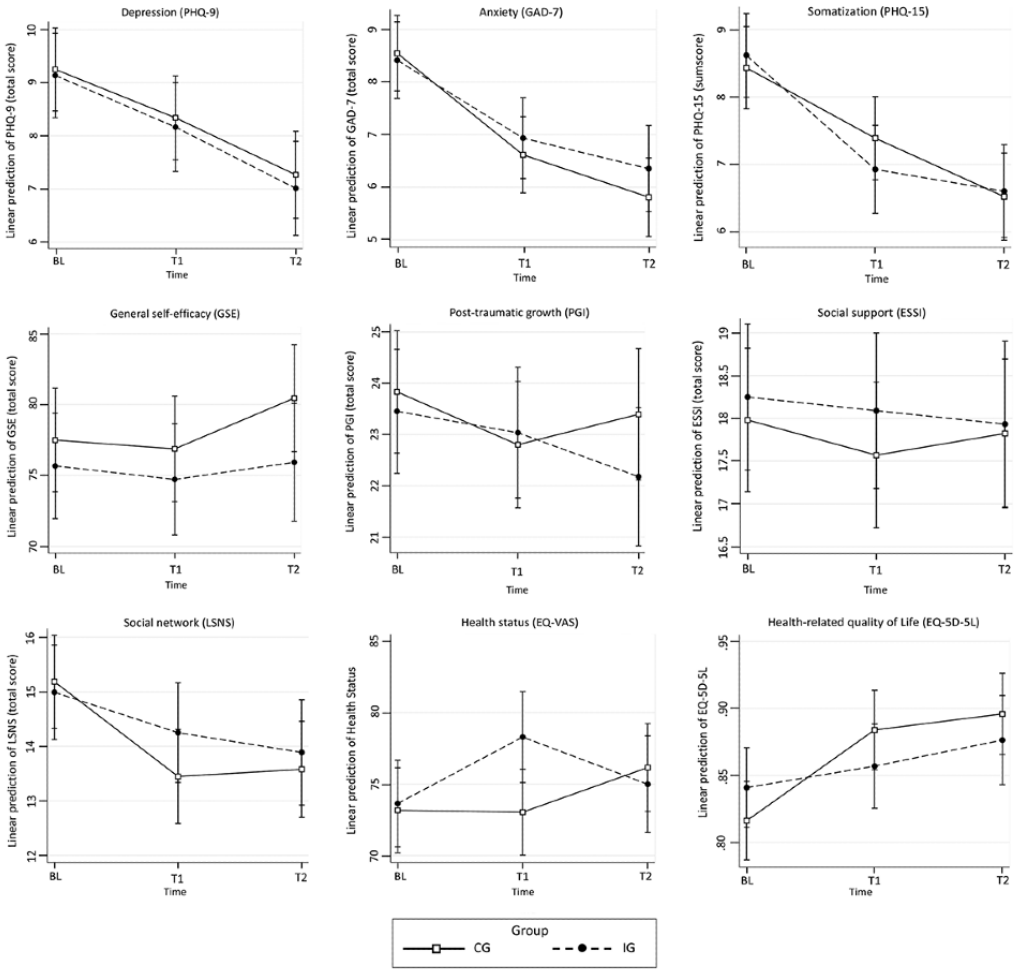


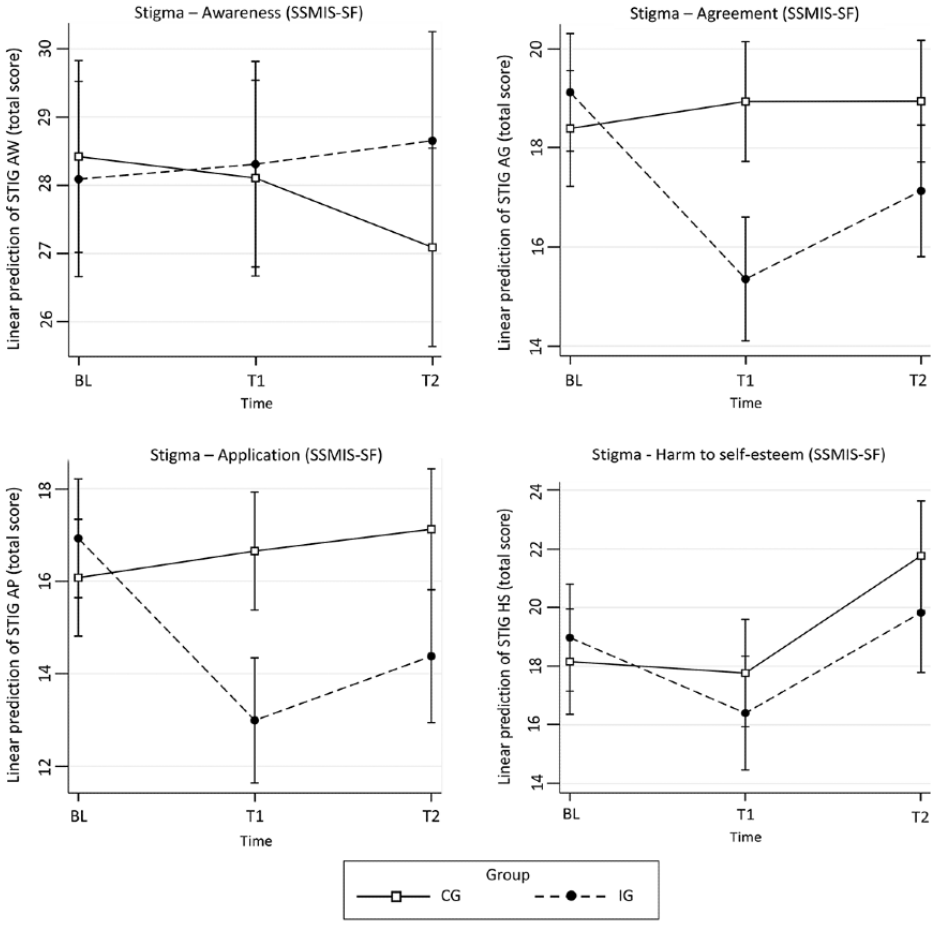
Table 4. Results of the Sanadak trial based on per-protocol analysis.

Outcome	n	Diff[a]	95% CI	*P* value	*d*	95% CI
Primary outcome						
PDS-5[b]						
Baseline	—	—	—	—	—	—
4 weeks	111	0.102	–2.774 to 2.978	.95	–0.010	–0.391 to 0.364
4 months	102	0.556	–3.560 to 2.448	.72	0.074	–0.322 to 0.470
Secondary outcome						
PHQ-9[c]						
Baseline	—	—	—	—	—	—
4 weeks	112	–0.194	–1.450 to 1.062	.76	0.059	–0.318 to 0.435
4 months	102	–0.503	–1.822 to 0.816	.46	0.153	–0.243 to 0.549
GAD-7[d]						
Baseline	—	—	—	—	—	—
4 weeks	112	0.377	–0.800 to 1.554	.53	–0.120	–0.499 to 0.254
4 months	102	0.822	–0.417 to 2.061	.19	–0.270	–0.667 to 0.131
PHQ-15[e]						
Baseline	—	—	—	—	—	—
4 weeks	112	–0.04	–1.027 to 0.947	.94	0.015	–0.361 to 0.392
4 months	101	0.424	–0.616 to 1.465	.42	–0.170	–0.562 to 0.234
ED-5D-5L[f]						
Baseline	—	—	—	—	—	—
4 weeks	112	–0.024	–0.073 to 0.025	.34	0.189	–0.189 to 0.556
4 months	104	–0.021	–0.071 to 0.030	.43	0.163	–0.163 to 0.556
Health status						
Baseline	—	—	—	—	—	—
4 weeks	112	4.464	–0.421 to 9.350	.07	–0.350	–0.727 to 0.032
4 months	104	–0.228	–5.338 to 4.881	.93	0.017	–0.376 to 0.411
GSE[g]						
Baseline	—	—	—	—	—	—
4 weeks	111	–0.581	–1.733 to 0.572	.32	0.194	–0.185 to 0.572
4 months	104	–0.862	–2.060 to 0.336	.16	0.288	–0.108 to 0.682
SSMIS-AW[h]						
Baseline	—	—	—	—	—	—
4 weeks	111	0.158	–2.088 to 2.405	.89	–0.030	–0.405 to 0.351
4 months	104	0.635	–1.704 to 2.974	.60	–0.110	–0.502 to 0.286
SSMIS-AG[i]						
Baseline	—	—	—	—	—	—
4 weeks	109	–3.996	–5.795 to –2.197	<.001	0.858	0.460 to 1.254
4 months	102	–2.435	–4.313 to –0.557	.01	0.522	0.119 to 0.923
SSMIS-AP[j]						
Baseline	—	—	—	—	—	—

Outcome	n	Diff[a]	95% CI	*P* value	*d*	95% CI
4 weeks	112	–3.201	–5.245 to –1.158	.01	0.601	0.215 to 0.985
4 months	104	–2.649	–4.787 to –0.510	.02	0.497	0.097 to 0.895
SSMIS-HS[k]						
Baseline	—	—	—	—	—	—
4 weeks	110	–0.446	–3.391 to 2.500	.77	0.058	–0.322 to 0.438
4 months	103	–1.877	–4.935 to 1.181	.23	0.246	–0.151 to 0.641
LSNS-6[l]						
Baseline	—	—	—	—	—	—
4 weeks	112	0.864	–0.509 to 2.233	.22	–0.240	–0.618 to 0.137
4 months	104	0.642	–0.785 to 2.070	.38	–0.180	–0.574 to 0.214
ESSI[m]						
Baseline	—	—	—	—	—	—
4 weeks	111	0.783	–0.613 to 2.179	.27	–0.220	–0.595 to 0.165
4 months	103	0.059	–1.411 to 1.528	.94	–0.020	–0.412 to 0.380
PGI[n]						
Baseline	—	—	—	—	—	—
4 weeks	109	0.46	–1.455 to 2.374	.64	–0.090	–0.473 to 0.287
4 months	99	–1.455 to 2.374	–2.090 to 1.927	.94	0.017	–0.383 to 0.416

[a]Diff: difference.
[b]PDS-5: Posttraumatic Diagnostic Scale for DSM-5.
[c]PHQ-9: Patient Health Questionnaire.
[d]GAD-7: Generalized Anxiety Disorder.
[e]PHQ-15: Physical Health Questionnaire.
[f]ED-5D-5L: EuroQoL 5-Dimension 5-Level.
[g]GSE: General Self-Efficacy.
[h]SSMIS-AW: Self-Stigma of Mental Illness Scale–stereotype awareness.
[i]SSMIS-AG: Self-Stigma of Mental Illness Scale–stereotype agreement.
[j]SSMIS-AP: Self-Stigma of Mental Illness Scale–stereotype application.
[k]SSMIS-HS: Self-Stigma of Mental Illness Scale–harm to self-esteem.
[l]LSNS-6: Lubben Social Network Scale.
[m]ESSI: ENRICHD Social Support Inventory.
[n]PGI: Posttraumatic Growth Inventory.

Figure 4. Adjusted predictions of measures of self-stigma (Self-Stigma of Mental Illness Scale–Short Form; AW: stereotype awareness, AG: stereotype agreement, AP: stereotype application, HS: harm to self-esteem) at baseline and 4 weeks (T1) and 4 months (T2) after baseline in the intervention group and control group in the Sanadak trial, an Arabic language self-help app for Syrian refugees with posttraumatic stress, using mixed-effects linear regression models.



Ancillary Analyses

None of the subgroup analyses in regard to age groups, gender, education, app use frequency, and posttraumatic stress symptom severity indicated any significant effect apart from reduced self-stigma as seen in the main analysis (results not further shown).

Cost-Effectiveness of the Intervention

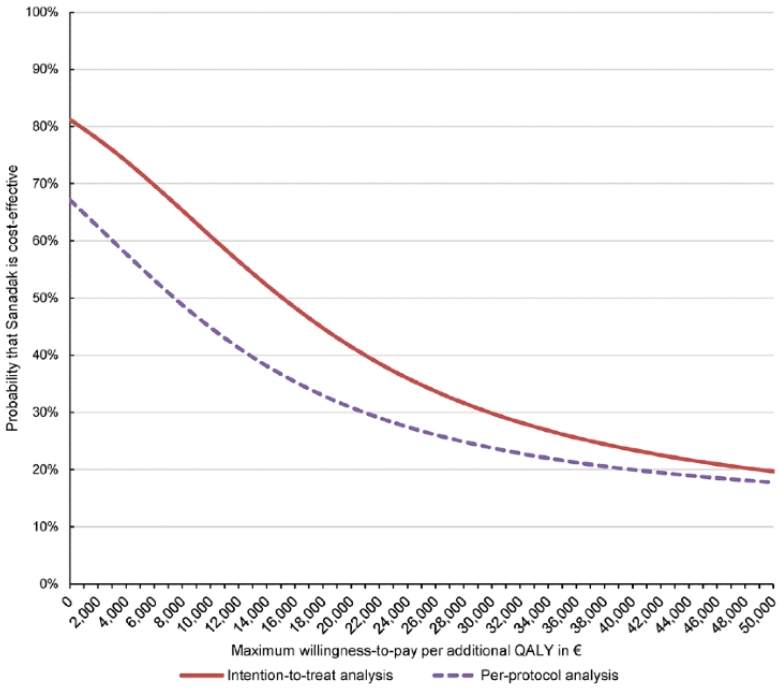
During the 4-month follow-up, mean adjusted total costs were €384 in the IG (including €52 intervention costs) and €484 in the CG (Table 5), with the difference not being statistically significant in the ITT analysis (−€100, *P*=.38). Mean adjusted QALYs were 0.29 in the IG and 0.29 in the CG, with the difference not being statistically significant (−.004, *P*=.35; Table 5). In the PP analysis, the differences in mean adjusted total costs (−€52, *P*=.11) and in mean adjusted QALYs (−.005, *P*=.34) between the IG and CG were not statistically significant. The probability for cost-effectiveness at a WTP of €0 per additional QALY was 81% (67%) in the intention-to-treat analysis (per protocol analysis). For a WTP of €50.000 per additional QALY, the probability for cost-effectiveness was 20% (18%; Figure 5).

Table 5. Adjusted mean costs (by cost category) and quality-adjusted life years (QALYs), and differences between intervention group and control group in mean costs (by cost category) and QALYs during 4-month follow-up based on intention-to-treat analysis and per-protocol analysis[a].

Cost category/measure of health effect	N	IG[b], mean (SE[c])	CG[d], mean (SE)	Diff[e] IG–CG	*P* value
ITT[f] ($)					
Inpatient care and rehabilitation	116	148 (52)	250 (93)	−103 (87)	.24
Outpatient physician services	108	180 (31)	214 (36)	−33 (46)	.47
Outpatient nonphysician services	114	13 (7)	20 (7)	−7 (10)	.50
Total costs	107	384 (67)	484 (111)	−100 (112)	.38
QALYs[g]	116	0.290 (0.004)	0.294 (0.004)	−0.004 (0.005)	.35
PP[h] ($)					
Inpatient care and rehabilitation	104	180 (64)	258 (93)	−78 (88)	.38
Outpatient physician services	96	195 (40)	205 (34)	−10 (50)	.84
Outpatient nonphysician services	102	19 (10)	18 (6)	0 (11)	.98
Total costs	95	433 (83)	484 (111)	−52 (117)	.66
QALYs	104	0.288 (0.004)	0.293 (0.004)	−0.005 (0.005)	.34

[a]Adjusted for age, sex, costs, EQ-5D-5L index and PDS-5 index at baseline using mixed-effects linear regression models with robust standard errors.
[b]IG: intervention group.
[c]SE: standard error.
[d]CG: control group.
[e]Diff: difference.
[f]ITT: intention to treat.
[g]QALY: quality-adjusted life year.
[h]PP: per protocol.

Figure 5. Adjusted cost-effectiveness acceptability curves for an additional quality-adjusted life year based on intention-to-treat analysis and per-protocol analysis (adjusted for age, sex, costs, EQ-5D-5L index, and PDS-5 index at baseline by multilevel mixed-effects linear regression with robust standard errors with research sites as random effect). QALY: quality-adjusted life year.



App Usability

At T1, participants in the IG were asked to rate the usability (SUS) of the app; 89.2% (58/65) of participants completed the SUS, indicating a mean usability score of 78.9 (SD 12.6; range 50.0 to 97.5; possible scoring range 0 to 100).

Harms

Across the intervention and follow-up period, one AE occurred with relation to the trial participation. The participant self-reported increased anxiety at T1 which, however, was not quantified by the GAD-7 scores comparing baseline and T1. The participant wanted to remain in the trial. Follow-up at T2 showed that the AE had resolved. No SAE occurred.

Source Data Verification

Across all assessment waves, 36 paper-and-pencil questionnaires (source data) were randomly drawn and compared with the electronic data file. For each questionnaire, up to 16 case report forms were inspected. Altogether, 7837 items were checked and 56 errors identified. This cumulated in a total error rate of 0.71%.

Discussion

Principal Findings

The smartphone-based app Sanadak, a self-help intervention in the Arabic language, was primarily developed for reducing posttraumatic stress in Syrian refugees residing in Germany as there is a gap of timely and culturally appropriate treatment. Its effectiveness was evaluated in an RCT with 133 Syrian refugees aged 18 to 65 years. Although symptom severity decreased, Sanadak was not superior in reducing mild to moderate posttraumatic stress in Syrian refugees in the short-term (4 weeks) and midterm (4 months) compared with the CG, who received psychoeducational reading material on trauma and PTSD. Likewise, the app showed no effectiveness regarding secondary outcomes, including depressive symptoms, anxiety, somatization, posttraumatic growth, general self-efficacy, social support, social isolation, health-related quality of life, and health state. However, there was a significant treatment effect for aspects of self-stigma: stereotype agreement as well as stereotype application reduced significantly in the short-term and midterm in the IG compared with the CG. Moreover, Sanadak is unlikely to be cost-effective. The app showed a small but insignificant reduction in total costs compared with the CG. As the number of QALYs tended to be (insignificantly) higher in the CG than in the IG, the probability of cost-effectiveness decreased with increasing willingness to pay per additional QALY, reaching only 20% at a WTP of €50.000. In a model-based economic evaluation, self-help without support also showed a small reduction in total costs of £235 (€199) compared with no treatment and was unlikely to be cost-effective compared with other treatment options, with a probability of cost-effectiveness of only 42% for a WTP of £20.000 (€23.624) per additional QALY [36].

Although evidence-based treatments are available for mild to severe PTSD and cognitive-behavioral therapies are among the recommended treatments [37], findings on the efficacy of app-based delivered interventions remain inconclusive and rely on a small number of studies [38-40]. In fact, a recent meta-analysis that identified 5 RCTs evaluating app-based interventions for PTSD symptoms found that the use of such apps is associated with reductions in PTSD symptoms, but that there was little evidence to suggest that apps were more effective than control conditions (usually waitlist) [40]. Another recent meta-analysis with 2 RCTs and 4 pre-post studies that focused on self-help apps for subthreshold or full PTSD likewise concluded that respective apps were not more effective than waitlist controls despite positive pre-post effects [41]. This is in line with the results of our trial, although our study differed regarding the target population. While the cited studies targeted either community-dwelling individuals or military/veteran populations, we were focusing on Syrian refugees residing in Germany. We could not identify similar app-based interventions targeting (Syrian) refugees with posttraumatic stress that have been previously evaluated in an RCT. There are a few apps targeting various mental health outcomes in (Syrian) refugees. However, they have either not been evaluated (eg, ALMHAR: self-help app for psychoeducation about emotions regulation [42]) or evaluation is ongoing (Smartphone Mediated Intervention for Learning Emotional Regulation of Sadness/SMILERS: self-help app for depressive symptoms [43]; BALSAM: self-help app for psychoeducation about emotions regulation as part of the Mental health in refugees and asylum seekers/MEHIRA project [44]; Step-by-Step: self-help app for depressive symptoms [45]). Evaluation results of the latter three apps will provide further valuable information on the effectiveness of app-based delivered mental health interventions in refugees. In general, regardless of targeted groups, there is no shortage of apps aiming to address posttraumatic stress. A systematic review by Sander and colleagues [46] identified 69 apps in Google Play and the iOS App Store. The overall app quality was found to be medium, and the authors only distinguished one app (1.4%) that had been scientifically evaluated in an RCT.

The literature provides more evidence regarding the effectiveness of interventions for posttraumatic stress in refugees delivered via internet or face-to-face. For example, in a web-based psychotherapy for war-traumatized Arab patients (focused on Iraq), posttraumatic stress was significantly reduced from baseline to posttreatment in the IG compared with the CG and effects sustained at 3-month follow-up [47]. Regarding face-to-face interventions, systematic reviews and meta-analyses found cognitive behavioral therapy, eye movement desensitization and reprocessing treatment and narrative exposure therapy [48,49] as well as psychosocial interventions [50] beneficial for certain populations of refugees; however, this evidence largely stems from settings in high-income countries, with only low-quality evidence from humanitarian settings in low- and middle-income countries [51].

The question is why treatment for posttraumatic stress may be less effective if delivered via apps compared with internet-based or face-to-face treatments. Due to the small number of studies, this remains to be investigated. However, it is likely that brief intervention periods, the self-management approach without clinical expert guidance or support, and the use of self-report

measures are explaining factors [40]. For example, we advised using the app as much and as regularly as possible over the 4-week intervention period; however, the logged mean time of app use was 42.5 minutes, with 22 IG participants (32.3%) having never used the app beyond the onboarding module (psychoeducation about posttraumatic stress). This requires thinking about strategies on how to increase engagement with the app, potentially using push notification reminders (not implemented in Sanadak due to data protection measures), motivational calls by study personnel, or detailed working schedules over a certain period of time on how and when to use the app. Potentially, 4 weeks is too short to obtain significant treatment effects relative to a control condition. We did not find differential results in regard to sociodemographic subgroups (age, gender, education, income, employment status), which further supports that lack of effectiveness is rather associated with the intervention delivery mode and use.

On a positive note, the SUS usability score of Sanadak was found to be above average, indicating very good usability. There were no harms associated with the use of Sanadak, and pre-post comparisons indicated a significant reduction in posttraumatic stress, even though not more than in the control condition. Moreover, Sanadak significantly reduced self-stigma regarding mental health relative to the CG. This aspect should not be underestimated. Mental health stigma has been named a key barrier to help seeking in refugees with posttraumatic stress [52]. By reducing mental health stigma, traumatized refugees in need of help may be more willing to take up respective treatment. As such, Sanadak may serve as a pathway to conventional face-to-face treatments.

Strengths and Limitations

The study has several strengths: a robust RCT design to determine effectiveness, adequate power, midterm follow-up assessment and low attrition, monitoring for potential harms as well as rigorous data quality control and imputation strategies for missing values, which decrease risk of bias in determining trial results. There are also limitations to consider. First, the multistrategic recruitment that heavily relied on snowball sampling techniques may limit generalizability of our findings to other traumatized populations. Moreover, the control condition which saw the provision of psychoeducational reading material in the Arabic language may not have been ideal in determining the app's treatment effect as building awareness on posttraumatic stress symptoms in the CG may been associated with symptom severity reduction. Potentially, a mere waiting list condition would have led to differential results. Moreover, assessments heavily relied on self-report assessments. There were no in-depth clinically diagnostic interviews to assess outcomes. As service use was assessed retrospectively over 4 months using an adapted and shortened version of a service receipt inventory, it is possible that health care services utilization was not completely assessed and a potentially delayed effect of Sanadak on total health care costs could not been identified.

Conclusions

Sanadak, an interactive self-help app in the Arabic language, showed reductions in posttraumatic stress in Syrian refugees residing in Germany in a pre-post comparison but was not superior to the control condition. This is in line with similar studies evaluating apps for posttraumatic stress, although targeting different populations. In addition, Sanadak is unlikely to be cost-effective. Future studies that investigate reasons for the limited effectiveness of respective apps are warranted to allow for improved app development and delivery modes. However, there were no harms associated with the use of Sanadak, and importantly, we found a significant and sustained treatment effect for reducing self-stigma. Consequently, Sanadak cannot be recommended as a standalone treatment for posttraumatic stress in refugee populations, but trial results indicate potential usefulness as a bridging aid, particularly as the usability was found to be very good, for the uptake of more effective internet-based or face-to-face treatments within a stepped and collaborative care approach.

Acknowledgments

The authors would like to thank all study participants for their time and for speaking so openly. Members of the HELP@APP Study Group: SGRH (principal investigator), AK (coprincipal investigator), HHK (coprincipal investigator), TG, RH, FUJ, MN, AR, SR, Maram Alhamarneh, Mais Alhamwi, Ghadir Ali, Zain Assaad, Ismail Ayoub, Jonathan Bär, Franziska Berg, Franziska Elias, Johannes Golchert, David Jäckle, Yamen Khamis, Waad Layka, Alia Kheirbek, Chiara Manthey, Amera Mojahed, Anna Plexnies, and Adrian Schultz. This publication is part of the study HELP@APP: Development and Evaluation of a Self-Help App for Traumatized Syrian Refugees and was funded by means of the Innovationsfonds of the German Federal Joint Committee (Gemeinsamer Bundesausschuss grant 01VSF16033). The funding source had no involvement in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Authors' Contributions

SRH, AK, HHK, and MN conceptualized the study. SR, FJ, AP, TG, JD, MN, AR, RH, HHK, AK, and SRH reviewed the literature. MN, AR, RH, AK, SR, and SRH developed the app. SR, SRH, AK, HHK, and MN designed the study. SR and FJ collected the data. SR, FJ, AP, and TG analyzed the data. SR and FJ drafted the manuscript. AP, TG, JD, MN, AR, RH, HHK, AK, and SRH revised the manuscript for important intellectual content. All authors had full access to all the data, contributed to data interpretation, and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2089 KB - mhealth_v9i1e24807_app1.pdf]

References

1. Flatten G, Gast U. S3-LEITLINIE Posttraumatische Belastungsstörung ICD-10: F43.1. Trauma & Gewalt 2011;5(3):202-210.
2. Renner A, Hoffmann R, Nagl M, Roehr S, Jung F, Grochtdreis T, et al. Syrian refugees in Germany: perspectives on mental health and coping strategies. J Psychosom Res 2020 Feb;129:109906. [doi: 10.1016/j.jpsychores.2019.109906] [Medline: 31884301]
3. Jefee-Bahloul H, Bajbouj M, Alabdullah J, Hassan G, Barkil-Oteo A. Mental health in Europe's Syrian refugee crisis. Lancet Psychiatry 2016 Apr;3(4):315-317. [doi: 10.1016/S2215-0366(16)00014-6] [Medline: 26868309]
4. Nosè M, Ballette F, Bighelli I, Turrini G, Purgato M, Tol W, et al. Psychosocial interventions for post-traumatic stress disorder in refugees and asylum seekers resettled in high-income countries: systematic review and meta-analysis. PLoS One 2017;12(2):e0171030 [FREE Full text] [doi: 10.1371/journal.pone.0171030] [Medline: 28151992]
5. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
6. von der Warth R, Dams J, Grochtdreis T, König H. Economic evaluations and cost analyses in posttraumatic stress disorder: a systematic review. Eur J Psychotraumatol 2020 May 29;11(1):1753940. [doi: 10.1080/20008198.2020.1753940]
7. Golchert J, Roehr S, Berg F, Grochtdreis T, Hoffmann R, Jung F, et al. HELP@APP: development and evaluation of a self-help app for traumatized Syrian refugees in Germany—a study protocol of a randomized controlled trial. BMC Psychiatry 2019 Apr 30;19(1):131 [FREE Full text] [doi: 10.1186/s12888-019-2110-y] [Medline: 31039786]
8. Shah SFA, Hess JM, Goodkind JR. Family separation and the impact of digital technology on the mental health of refugee families in the United States: qualitative study. J Med Internet Res 2019 Sep 03;21(9):e14171 [FREE Full text] [doi: 10.2196/14171] [Medline: 31482853]
9. Griner D, Smith TB. Culturally adapted mental health intervention: a meta-analytic review. Psychotherapy (Chic) 2006;43(4):531-548. [doi: 10.1037/0033-3204.43.4.531] [Medline: 22122142]
10. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Consolidated Standards of Reporting Trials Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):e1-e37 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.03.004] [Medline: 20346624]
11. Roehr S, Wittmann F, Jung F, Hoffmann R, Renner A, Dams J, et al. [Strategies to recruit refugees for intervention studies: lessons learned from the Sanadak trial]. Psychother Psychosom Med Psychol 2019 Dec;69(12):484-489. [doi: 10.1055/a-1030-4570] [Medline: 31801164]
12. Foa EB, McLean CP, Zang Y, Zhong J, Powers MB, Kauffman BY, et al. Psychometric properties of the Posttraumatic Diagnostic Scale for DSM-5 (PDS-5). Psychol Assess 2016 Oct;28(10):1166-1171. [doi: 10.1037/pas0000258] [Medline: 26691504]
13. 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] [Medline: 11556941]
14. von Glischinski M, Teismann T, Prinz S, Gebauer JE, Hirschfeld G. Depressive symptom inventory suicidality subscale: optimal cut points for clinical and non-clinical samples. Clin Psychol Psychother 2016 Nov;23(6):543-549. [doi: 10.1002/cpp.2007] [Medline: 26857479]
15. Röhr S, Jung FU, Renner A, Plexnies A, Hoffmann R, Dams J, et al. Recruitment and baseline characteristics of participants in the “Sanadak” trial: a self-help app for Syrian refugees with post-traumatic stress. Int J Environ Res Public Health 2020 Oct 18;17(20):7578 [FREE Full text] [doi: 10.3390/ijerph17207578] [Medline: 33081070]
16. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
17. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
18. Kocalevent R, Hinz A, Brähler E. Standardization of a screening instrument (PHQ-15) for somatization syndromes in the general population. BMC Psychiatry 2013 Mar 20;13:91 [FREE Full text] [doi: 10.1186/1471-244X-13-91] [Medline: 23514436]
19. Schwarzer R, Jerusalem M. Skalen zur Erfassung von Lehrer- und Schülermerkmalen. In: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin; 1999.

20. Corrigan PW, Michaels PJ, Vega E, Gause M, Watson AC, Rüsch N. Self-stigma of mental illness scale–short form: reliability and validity. Psychiatry Res 2012 Aug 30;199(1):65-69 [FREE Full text] [doi: 10.1016/j.psychres.2012.04.009] [Medline: 22578819]
21. Leppert K, Koch B, Brähler E, Strauss B. Die Resilienzskala (RS)—Überprüfung der Langform RS-25 und einer Kurzform RS-13. Klinische Diagnostik und Evaluation 2008:226-243.
22. Lubben JE. Assessing social networks among elderly populations. Family Community Health 1988;11(3):42-52. [doi: 10.1097/00003727-198811000-00008]
23. Kendel F, Spaderna H, Sieverding M, Dunkel A, Lehmkuhl E, Hetzer R, et al. Eine deutsche Adaptation des ENRICHD Social Support Inventory (ESSI). Diagnostica 2011 Apr;57(2):99-106. [doi: 10.1026/0012-1924/a000030]
24. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
25. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
26. Tedeschi RG, Calhoun LG. The Posttraumatic Growth Inventory: measuring the positive legacy of trauma. J Trauma Stress 1996 Jul;9(3):455-471. [doi: 10.1007/BF02103658] [Medline: 8827649]
27. Harkness JA, Villar A, Edwards B. Translation, adaptation, and design. In: Harkness JA, Braun M, Edwards B, Johnson TB, Lyberg L, Mohler PP, et al, editors. Survey Methods in Multinational, Multiregional, and Multicultural Contexts. Hoboken: John Wiley & Sons, Inc; 2010:117-140.
28. Roick C, Kilian R, Matschinger H, Bernert S, Mory C, Angermeyer MC. [German adaptation of the client sociodemographic and service receipt inventory—an instrument for the cost of mental health care]. Psychiatr Prax 2001 Oct;28 Suppl 2:S84-S90. [doi: 10.1055/s-2001-17790] [Medline: 11605129]
29. Bock J, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen 2015 Jan;77(1):53-61. [doi: 10.1055/s-0034-1374621] [Medline: 25025287]
30. Ludwig K, Graf von der Schulenburg J, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics 2018 Jun;36(6):663-674 [FREE Full text] [doi: 10.1007/s40273-018-0615-8] [Medline: 29460066]
31. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry, Vol 189. London: Taylor and Francis; 1996:4-7.
32. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
33. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
34. Brauns H, Scherer S, Steinmann S. The CASMIN educational classification in international comparative research. In: Hoffmeyer-Zlotnik JHP, Wolf C, editors. Advances in Cross-National Comparison. Boston: Springer; 2003:221-244.
35. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programme. 3rd Edition. Oxford: Oxford University Press; 2005.
36. Mavranezouli I, Megnin-Viggars O, Grey N, Bhutani G, Leach J, Daly C, et al. Cost-effectiveness of psychological treatments for post-traumatic stress disorder in adults. PLoS One 2020;15(4):e0232245 [FREE Full text] [doi: 10.1371/journal.pone.0232245] [Medline: 32353011]
37. Foa E, Keane T, Friedman M, Cohen J. Effective Treatments for PTSD. New York: Guilford Publications; 2020.
38. Rodriguez-Paras C, Tippey K, Brown E, Sasangohar F, Creech S, Kum H, et al. Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. JMIR Mhealth Uhealth 2017 Oct 26;5(10):e156 [FREE Full text] [doi: 10.2196/mhealth.7318] [Medline: 29074470]
39. Haun JN, Ballistrea LM, Melillo C, Standifer M, Kip K, Paykel J, et al. A mobile and web-based self-directed complementary and integrative health program for veterans and their partners (mission reconnect): protocol for a mixed-methods randomized controlled trial. JMIR Res Protoc 2019 May 13;8(5):e13666 [FREE Full text] [doi: 10.2196/13666] [Medline: 31094345]
40. Wickersham A, Petrides PM, Williamson V, Leightley D. Efficacy of mobile application interventions for the treatment of post-traumatic stress disorder: A systematic review. Digit Health 2019;5:2055207619842986 [FREE Full text] [doi: 10.1177/2055207619842986] [Medline: 31019722]
41. Goreis A, Felnhofer A, Kafka JX, Probst T, Kothgassner OD. Efficacy of self-management smartphone-based apps for post-traumatic stress disorder symptoms: a systematic review and meta-analysis. Front Neurosci 2020;14:3 [FREE Full text] [doi: 10.3389/fnins.2020.00003] [Medline: 32038153]
42. Mental health aid for refugees. URL: http://almhar.org/ [accessed 2020-12-28]
43. Ilajnafsy Selbsthilfe app. 2017. URL: https://ilajnafsy.bzfo.de/portal/de/onlinetherapien/selbsthilfe-app/ [accessed 2020-10-30]
44. Böge K, Karnouk C, Hahn E, Schneider F, Habel U, Banaschewski T, et al. Mental health in refugees and asylum seekers (MEHIRA): study design and methodology of a prospective multicentre randomized controlled trail investigating the effects

of a stepped and collaborative care model. Eur Arch Psychiatry Clin Neurosci 2020 Feb;270(1):95-106. [doi: 10.1007/s00406-019-00991-5] [Medline: 30796528]
45. Burchert S, Alkneme MS, Bird M, Carswell K, Cuijpers P, Hansen P, et al. User-centered app adaptation of a low-intensity e-mental health intervention for Syrian refugees. Front Psychiatry 2018;9:663 [FREE Full text] [doi: 10.3389/fpsyt.2018.00663] [Medline: 30740065]
46. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. "Help for trauma from the app stores?" A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
47. Knaevelsrud C, Brand J, Lange A, Ruwaard J, Wagner B. Web-based psychotherapy for posttraumatic stress disorder in war-traumatized Arab patients: randomized controlled trial. J Med Internet Res 2015 Mar 20;17(3):e71 [FREE Full text] [doi: 10.2196/jmir.3582] [Medline: 25799024]
48. Thompson CT, Vidgen A, Roberts NP. Psychological interventions for post-traumatic stress disorder in refugees and asylum seekers: a systematic review and meta-analysis. Clin Psychol Rev 2018 Jul;63:66-79. [doi: 10.1016/j.cpr.2018.06.006] [Medline: 29936342]
49. Slobodin O, de Jong JTVM. Mental health interventions for traumatized asylum seekers and refugees: what do we know about their efficacy? Int J Soc Psychiatry 2015 Feb;61(1):17-26. [doi: 10.1177/0020764014535752] [Medline: 24869847]
50. Turrini G, Purgato M, Acarturk C, Anttila M, Au T, Ballette F, et al. Efficacy and acceptability of psychosocial interventions in asylum seekers and refugees: systematic review and meta-analysis. Epidemiol Psychiatr Sci 2019 Aug;28(4):376-388 [FREE Full text] [doi: 10.1017/S2045796019000027] [Medline: 30739625]
51. Purgato M, Gastaldon C, Papola D, van Ommeren M, Barbui C, Tol WA. Psychological therapies for the treatment of mental disorders in low- and middle-income countries affected by humanitarian crises. Cochrane Database Syst Rev 2018 Jul 05;7:CD011849 [FREE Full text] [doi: 10.1002/14651858.CD011849.pub2] [Medline: 29975811]
52. Nickerson A, Byrow Y, Pajak R, McMahon T, Bryant RA, Christensen H, et al. “Tell Your Story”: a randomized controlled trial of an online intervention to reduce mental health stigma and increase help-seeking in refugee men with posttraumatic stress. Psychol Med 2020 Apr;50(5):781-792. [doi: 10.1017/S0033291719000606] [Medline: 30973115]

Abbreviations

AE: adverse events
CG: control group
CONSORT: Consolidated Standards of Reporting Trials
DSI-SS: Depressive Symptom Inventory–Suicidality Subscale
EQ-5D-5L: EuroQoL 5-Dimension 5-Level
EQ-VAS: EuroQoL visual analog scale
ESSI: ENRICHD Social Support Inventory
GAD-7: Generalized Anxiety Disorder
GDPR: General Data Protection Regulation
GSE: General Self-Efficacy
IG: intervention group
ITT: intention-to-treat
LSNS-6: Lubben Social Network Scale
PDS-5: Posttraumatic Diagnostic Scale for DSM-5
PGI: Posttraumatic Growth Inventory
PHQ-9: Patient Health Questionnaire
PHQ-15: Physical Health Questionnaire
PP: per-protocol
PTSD: posttraumatic stress disorder
QALY: quality-adjusted life year
RCT: randomized controlled trial
RS-13: Resilience Scale
SAE: severe adverse events
SSMIS-AG: stereotype agreement
SSMIS-AW: stereotype awareness
SSMIS-AP: stereotype application
SSMIS-HS: harm to self-esteem
SSMIS-SF: Self-Stigma of Mental Illness Scale–Short Form
SUS: System Usability Scale
T0: baseline
T1: 4 weeks after baseline

T2: 4 months after baseline
WTP: willingness-to pay

Edited by G Eysenbach; submitted 06.10.20; peer-reviewed by Z Ma, I Schiering; comments to author 29.10.20; revised version received 02.11.20; accepted 03.11.20; published 13.01.21.

Please cite as:
Röhr S, Jung FU, Pabst A, Grochtdreis T, Dams J, Nagl M, Renner A, Hoffmann R, König HH, Kersting A, Riedel-Heller SG
A Self-Help App for Syrian Refugees With Posttraumatic Stress (Sanadak): Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(1):e24807
URL: http://mhealth.jmir.org/2021/1/e24807/
doi:10.2196/24807
PMID:33439140

Original Paper

A Mobile App–Based Intervention Program for Nonprofessional Caregivers to Promote Positive Mental Health: Randomized Controlled Trial

Carme Ferré-Grau[1*], RN, BSc, MSN, PhD; Laia Raigal-Aran[1*], RN, BSc, MSN, PhD; Jael Lorca-Cabrera[2], RN, MSc; Teresa Lluch-Canut[3], RN, BSc, MSc, PhD; Maria Ferré-Bergadà[4], BSc, PhD; Mar Lleixá-Fortuño[5], RN, BSc, MSc, PhD; Montserrat Puig-Llobet[3], RN, MSc, PhD; Maria Dolores Miguel-Ruiz[6], RN, MSc, PhD; Núria Albacar-Riobóo[2], RN, MSc, PhD

[1]Department of Nursing, Universitat Rovira i Virgili, Tarragona, Spain

[2]Department of Nursing, Universitat Rovira i Virgili, Tortosa, Spain

[3]Department of Public Health, Mental Health and Maternal-Child Nursing, School of Nursing, University of Barcelona, Barcelona, Spain

[4]Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, Spain

[5]Territorial Health Services of Terres de l'Ebre, Catalan Health Institute, Tortosa, Spain

[6]Department of Mental Health, Campus Docent Sant Joan de Déu-–Fundació Privada, University of Barcelona, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Laia Raigal-Aran, RN, BSc, MSN, PhD
Department of Nursing
Universitat Rovira i Virgili
35
Catalunya Avenue
Tarragona, 43002
Spain
Phone: 34 977558000
Email: laia.raigal@urv.cat

Abstract

Background: While nonprofessional caregivers often experience a sense of fulfillment when they provide care, there is also a significant risk of emotional and physical burnout. Consequently, this can negatively affect both the caregiver and the person being cared for. Intervention programs can help empower nonprofessional caregivers of people with chronic diseases and develop solutions to decrease the physical and psychological consequences resulting from caregiving. However, most clinically tested intervention programs for nonprofessional caregivers require face-to-face training, and many caregivers encounter obstacles that hinder their participation in such programs. Consequently, it is necessary to design internet-based intervention programs for nonprofessional caregivers that address their needs and test the efficacy of the programs.

Objective: The aim of this study was to evaluate the effectiveness of a smartphone app–based intervention program to increase positive mental health for nonprofessional caregivers.

Methods: This study was a randomized controlled trial of 3 months' duration. A total of 152 caregivers over 18 years of age with a minimum of 4 months' experience as nonprofessional caregivers were recruited from primary health care institutions. Nonprofessional caregivers were randomized into two groups. In the intervention group, each caregiver installed a smartphone app and used it for 28 days. This app offered them daily activities that were based on 10 recommendations to promote positive mental health. The level of positive mental health, measured using the Positive Mental Health Questionnaire (PMHQ), and caregiver burden, measured using the 7-item short-form version of the Zarit Caregiver Burden Interview (ZBI-7), were the primary outcomes. Users' satisfaction was also measured.

Results: In all, 113 caregivers completed the study. After the first month of the intervention, only one factor of the PMHQ, F1–Personal satisfaction, showed a significant difference between the groups, but it was not clinically relevant (0.96; *P*=.03). However, the intervention group obtained a higher mean change for the overall PMHQ score (mean change between groups: 1.40; *P*=.24). The results after the third month of the intervention showed an increment of PMHQ scores. The mean difference

of change in the PMHQ score showed a significant difference between the groups (11.43; *P*<.001; *d*=0.82). Significant changes were reported in 5 of the 6 factors, especially F5–Problem solving and self-actualization (5.69; *P*<.001; *d*=0.71), F2–Prosocial attitude (2.47; *P*<.001; *d*=1.18), and F3–Self-control (0.76; *P*=.03; *d*=0.50). The results of the ZBI-7 showed a decrease in caregiver burden in the intervention group, although the results were inconclusive. Approximately 93.9% (46/49) of the app users indicated that they would recommend the app to other caregivers and 56.3% (27/49) agreed that an extension of the program's duration would be beneficial.

Conclusions: The app-based intervention program analyzed in this study was effective in promoting positive mental health and decreasing the burden of caregivers and achieved a high range of user satisfaction. This study provides evidence that mobile phone app–based intervention programs may be useful tools for increasing nonprofessional caregivers' well-being. The assessment of the effectiveness of intervention programs through clinical trials should be a focus to promote internet-based programs in health policies.

Trial Registration: ISRCTN Registry ISRCTN14818443; http://www.isrctn.com/ISRCTN14818443

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7264-5

(JMIR Mhealth Uhealth 2021;9(1):e21708) doi:10.2196/21708

KEYWORDS

clinical trial; caregiver; mobile phone app; intervention program; positive mental health; mobile health; health promotion; randomized controlled trial; nursing; caregiving; technology

Introduction

The increase in life expectancy has contributed to the progressive aging of the population, which has modified the epidemiological pattern, marked by an increase in chronic diseases. Chronic noncommunicable diseases are the main cause of mortality and morbidity in the world, accounting for the death of 38 million people worldwide each year [1]. In Spain, chronic diseases constitute one of the main public health problems, as they are the cause of the increase in morbidity, disability, and impairment, with a resulting impact on the lives of both patients and their caregivers [2].

In Spain, nonprofessional caregivers are on the front line of care for people with physical and/or psychological dependencies. The most common type of nonprofessional caregiver is a family member, althought there is an increase in the number of caregivers contracted by families or the person being cared for who are not trained health care professionals [3,4]. According to the national and international literature, there is a consensus that people who take on the role of nonprofessional caregivers—continuously and/or over a long period of time—must carry out multiple and/or complex tasks, which often leads to feelings of discouragement and stress [5,6]. Evidence suggests that nonprofessional caregivers may experience negative symptoms such as sleep disturbance, fatigue, depression, and anxiety [7,8]. Experiencing such difficulties involves a process that continuously tests their physical capacity and positive mental health, which often leads to overburdening and/or caregiver burnout [8-10].

Since the caregiver's role is an important element in ensuring the well-being of the person being cared for [11], intervention in this population group becomes a priority. There are intervention and/or support programs that have been positively evaluated [12-15], although most of them are based on in-person training. However, research has shown that many caregivers underutilize the supports available to them and instead try to handle everything by themselves [16,17]. The barriers that hinder access to in-person support programs include a lack of coordinated home care services [18], the multiple occupations and intensive dedication required by family care [19], geographical or transportation limitations, and the caregivers' own health problems [20,21]. These obstacles often make access to these resources difficult. The development and evaluation of mobile health (mHealth) tools, supervised by health professionals and adapted to the time constraints of caregivers, might be a useful strategy for an intervention program aimed at this population group.

We live in a digital age, in which digital literacy is a necessary competence for the practice of any profession and even more so in the field of health care. Nurses must develop new digital strategies to carry out their professional activities [22]. Traditional models of health care are changing with the development of information and communication technologies (ICTs) and the incorporation of mHealth solutions. Digital health intervention programs have had a significant impact on the care of chronic diseases, providing access to electronic health records, apps, and health portals [21,23-26].

Apps for smartphones and tablets have become indispensible and complementary tools to health care [27,28]. They provide an opportunity to lead ICT-based projects to empower people, teach them how to manage their health, improve their quality of life, and achieve well-being [21,22]. A recent review highlighted the importance of implementing new technologies in health care policies—specifically, in that case study, telemedicine [29]. It stressed that further studies should be carried out aimed at a consistent analysis and follow-up of patients after such intervention programs as a strategy to demonstrate that they are necessary [30].

There are currently more than 325,000 health apps available for health system users to download and interact with on their mobile devices. These tools can help improve the possibilities for continuity of care for the population, optimize existing resources, and increase the quality of care [28,31].

A systematic review by Lorca-Cabrera et al [21] concluded that there are very few studies on the evaluation of apps aimed at improving the health and welfare of caregivers. This reality, aggravated by the lack of regulation of effective apps for this purpose, can lead to health issues for patients and their caregivers due to potentially erroneous or inaccurate information provided by these apps [32,33]. It is therefore necessary to design and evaluate the apps in which intervention program for caregivers are conducted and measure their effectiveness. In this way, caregivers will benefit from the full potential of the apps to manage their self-care and improve their quality of life [34].

This study was based on a previous project in which a website [25,35] was developed for caregivers of patients with chronic diseases [25]. Given the evidence mentioned above, this study aimed to evaluate a digital intervention program of care for caregivers, using a mobile app that promotes positive mental health and/or reduces overburdening. The theoretical basis for the promotion of caregivers' health was Lluch-Canut's positive mental health assessment model, along with her decalogue of practical recommendations [36,37]. The protocol of this clinical trial study was previously published [38].

Methods

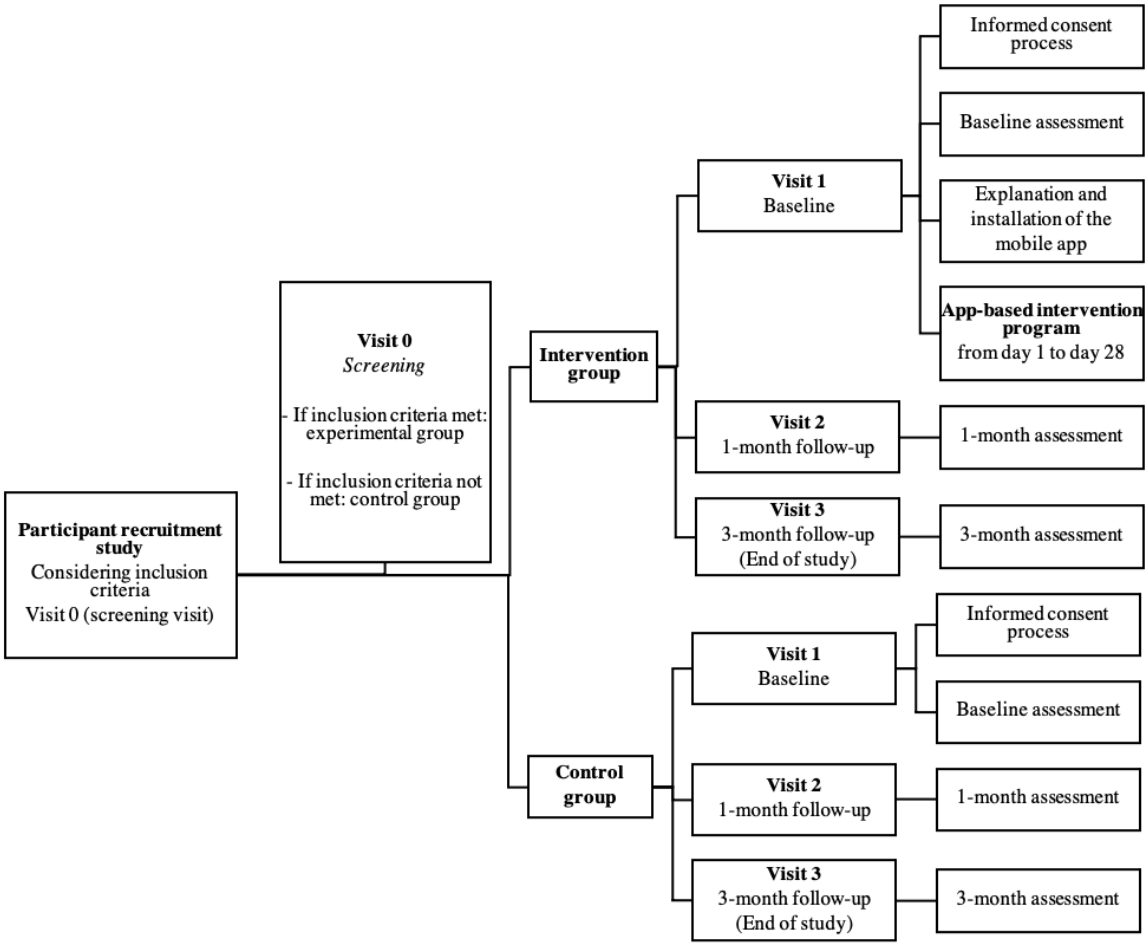
Aims

This study aimed to evaluate the effectiveness of a smartphone app–based intervention program compared with a standard intervention program for caregivers in primary health care institutions. The usability and satisfaction of the app were a focus of this study as well. The hypothesis was that an app-based intervention program would improve caregivers' mental health and decrease their burden.

Design

A randomized controlled trial (RCT) was conducted. Participants were randomly assigned to either the experimental group or the control group (Figure 1). The RCT was registered in the International Standard Randomized Controlled Trial Number registry (ISRCTN: 14818443; May 24, 2019).

Figure 1. Participant flowchart.



Sample Size Calculation

The study aimed to show differences between an app-based intervention and standard intervention with a standardized effect size (Cohen *d*) of 0.33 or larger. A standardized effect of 0.33 can be considered the lowest limit of a moderate clinical effect [39] and is based on a meta-analysis of well-being intervention research [40] and a recent RCT [41]. The total population of caregivers was uncertain, so the sample size was calculated considering an α risk of .05 and a β risk of .20. The minimum number of participants required to obtain evidential data of the results was 108 subjects, with 54 subjects each in the experimental and control groups. An estimated 30% loss to follow-up was taken into account for each group and a difference

greater than or equal to 10 was recognized as statistically significant.

Recruitment

Primary health care is the first point of access to the public health system in Catalonia, Spain. The rest of the services in the system can be accessed by referral from primary care, except for medical emergencies, which can be accessed directly in the event of urgent need. The primary care center is the place for on-site care where people have to go when they have a health problem or want to prevent an illness. It provides diagnosis and care for the main health problem; health and social care; health promotion services; preventive, curative, and rehabilitative care; home care service; urgent or continuous care; and sexual health care. Primary care services are part of the basic common core portfolio, which are services fully covered by public funding.

The eligible caregivers were recruited by nurses from 7 primary health care centers in central and southern Catalonia. Each nurse invited caregivers they knew who met the criteria to participate in the study. The inclusion criteria for the participants were as follows: (1) primary or secondary nonprofessional caregiver of someone with a chronic disease; (2) over 18 years of age; (3) minimum of 4 months of experience as a caregiver; (4) knowledge of Spanish or Catalan (the app was available in both languages and the user could choose his/her preference); (5) user of a mobile device and the WhatsApp mobile app; (6) access to a mobile device with an Android operating system and internet access; and (7) signature of informed consent.

Randomization and Blinding

The online randomization procedure was carried out on an individual basis. Caregivers who agreed to take part in the study—who had a phone that supported the app and met the inclusion criteria—were randomly assigned. The randomization was stratified by the primary health center attended (7 centers in total), gender, age, and perceived level of well-being. A computer program allocated the participants using a generated randomization list. Given the nature of the study, it was not possible to blind the caregivers and the professionals.

Details of the Intervention and Control Groups

Both study groups received the same standard intervention for caregivers by the nursing staff in the primary health care centers. Every year, the nurses assess the burden of the caregivers using the validated Zarit Burden Interview tool. If they identify a high level of burden, they refer the caregiver for psychological intervention. There is no specific protocol or intervention to promote positive attitudes toward caregiving.

Control Group

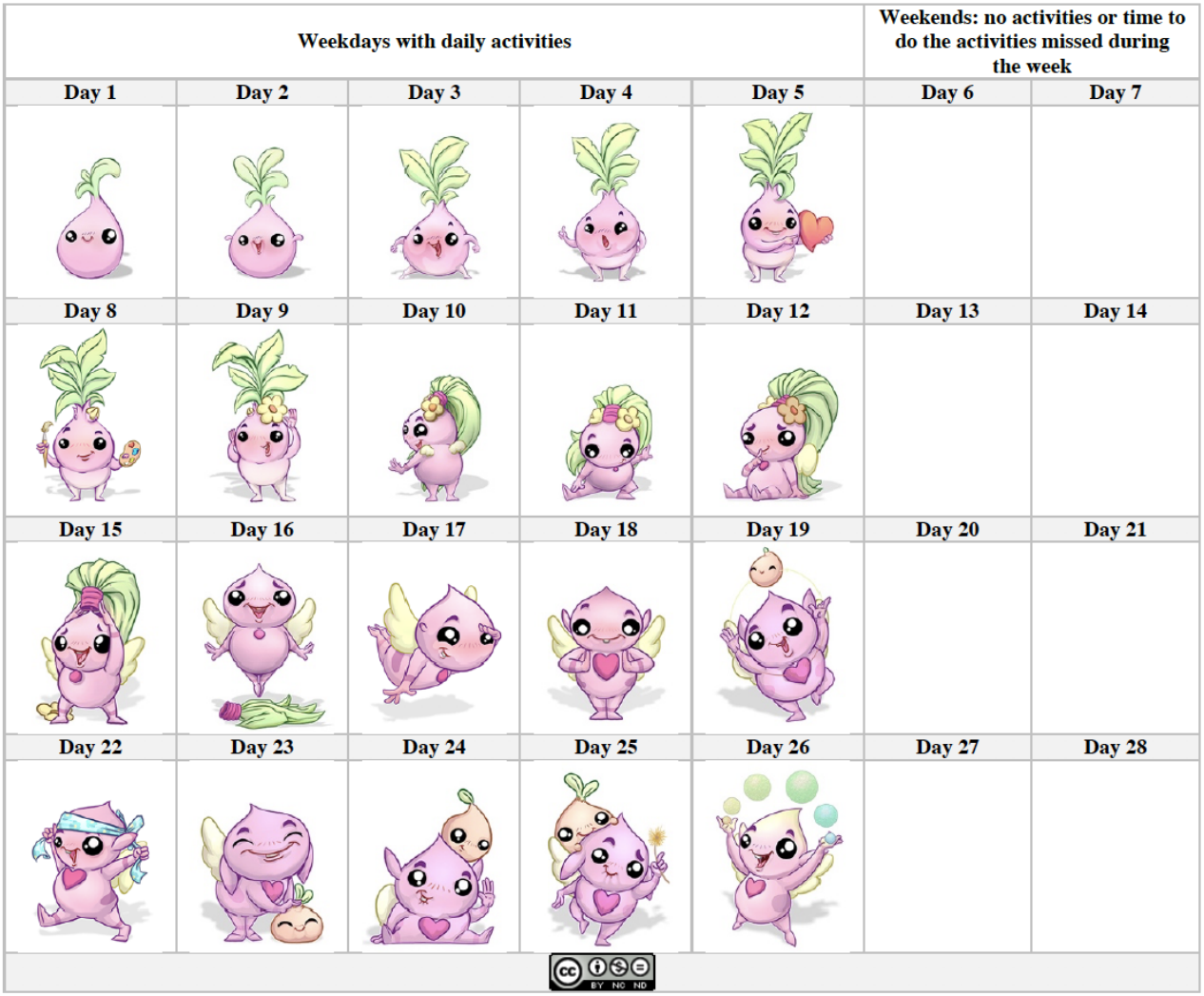
Participants in the control group received the standard intervention for caregivers by the nurses at their primary health care center of reference.

Intervention Group

Caregivers in the intervention group received the standard nursing care in addition to a free smartphone app (the TIVA app), which involved a 28-day intervention program. The researcher had to register the app and set a starting date for the intervention program, as agreed upon by the caregiver and the nurse in charge. During this period of time, the caregiver had the smartphone app active on his or her mobile device.

In the intervention program, the TIVA app offered participants an activity daily from Monday to Friday. These activities were related to the decalogue of positive mental health described by Lluch-Canut [36]. The decalogue includes 10 recommendations to promote positive mental health. For each recommendation, 2 activities were created by a group of experts who were part of the research group. After carrying out each activity, the caregiver expressed whether it was useful or not. The app provided a motivational quote every day and asked the caregiver, “Hello, how are you today?” Although there were no activities during the weekend, the app still recommended that caregivers visit the website developed in a previous related project [35]. The final activity offered caregivers an opportunity to register on the app’s website. This allowed them to be connected to other caregivers and to have access to any news posted there. The app includes gamification entailing an ad hoc character named TIVA—from the Spanish word “posiTIVA,” meaning “positive.” This character grows up and changes every time the caregiver performs the daily activity (Figure 2).

Figure 2. Evolution of the main character of the TIVA app.



Data Collection

The study was performed from September 2019 to November 2019. For the intervention group, primary outcome data were collected through the app, and secondary outcome data were collected by nurses through ad hoc questionnaires. There were three time points for data collection—at baseline (when the caregiver agreed to participate), and at 1 month and 3 months after baseline.

Measures and Outcomes

Primary Outcome

The primary outcome was related to an increase in the positive mental health score and a decrease in the caregiver burden score in the intervention group compared with the control group. This was measured using Lluch-Canut's validated positive mental health questionnaire (PMHQ [42]). The PMHQ consists of 39 items distributed among 6 factors that describe positive mental health: F1–Personal satisfaction (8 items), F2–Prosocial attitude (5 items), F3–Self-control (5 items), F4–Autonomy (5 items), F5–Problem-solving and self-actualization (9 items), and F6–Interpersonal relationship skills (7 items). The items are in the form of positive or negative statements that patients rate on a scale from 1 to 4 according to how frequently they occur: 1=always or almost always, 2=quite often, 3=sometimes, and 4=rarely or never. The PMHQ provides a global score for positive mental health (sum of the item scores) as well as specific scores for each factor. The global positive mental health value ranges from 39 points (low positive mental health) to 156 points (high positive mental health). The minimum and maximum values for each factor are as follows: F1 8-32, F2 5-20, F3 5-20, F4 5-20, F5 9-36, and F6 7-28.

Caregiver burden was assessed using the 7-item short-form version of the Zarit Caregiver Burden Interview (ZBI-7), which was validated by Martín Carrasco et al [43] and Regueiro Martínez et al [44]. It measures the caregiver's perceived burden in providing care. The 7 items are assessed on a 5-point Likert scale, ranging from 0=never to 4=almost always. Item scores are summed to obtain a total score ranging from 0 to 28, with higher scores indicating greater burden. The questions focus on major areas such as caregivers' health, psychological well-being, finances, social life, and the caregiver-patient relationship.

Secondary Outcome

The secondary outcome was related to the usability and satisfaction regarding the app by the intervention group. An ad

hoc questionnaire was administered by the nurse in charge. This questionnaire was created by the research group to evaluate the usability and satisfaction of the app-based intervention program. Qualitative data were also collected by the nurse through an open interview to obtain feedback on the user experience from the app-based intervention program users.

Data Analysis

Descriptive statistics, including means for continuous variables and proportions for categorical variables, were used to summarize the characteristics of the participants. Analyses were conducted using an intention-to-treat analysis. Bivariate analysis was performed to calculate the mean change between the baseline and follow-up (at 1 month and 3 months). Next, to estimate the difference between the two groups, we calculated the difference between the mean change in the intervention group and the mean change in the control group. Due to the nonnormal distribution of the sample, nonparametric tests were used (Kolmogorov-Smirnov test, $P<.001$). To compare differences between the groups, the Mann-Whitney U test and Wilcoxon test were used. A P value ≤0.05 was considered significant. Cohen d analysis was performed to measure the effect size. Dropouts were not analyzed. Statistical analysis was performed using SPSS Statistics software (version 25 for Mac; IBM Corp).

If more than two items were missing from either of the instruments, the total score was not calculated and the data were considered missing. Missing data from the PMHQ and ZBI-7 were handled through average imputation of answered items, as long as no more than 40% of the items were missing. If more than 40% of the items were missing, the overall score was not calculated and the data were considered missing.

Validity and Reliability/Rigor

CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) [45,46] was used to guide the design and implementation of the RCT. The researchers, who are experienced with clinical trials, monitored the study design, study protocols, patient recruitment, blinding, subject dropouts, and patient information confidentiality. The scrupulous study design ensured quality management and high external validity.

Ethical Considerations

The study was approved by the Ethics Committee of Institut d'Atenció Primaria Jordi Gol (reference number: PI18/207). Caregivers were informed about the content, purpose, and procedure of the study. Written informed consent was given. Patients were reassured that their withdrawal would not prevent them from receiving the care that they would normally receive. The study was conducted in accordance with the principles of the Declaration of Helsinki, revised and updated, and followed Spain's best practice guidelines (Buena Práctica Clínica). No negative impact was expected on the participants of the study. Data confidentiality was protected under the Spanish law governing the protection of personal data (Ley Orgánica 3/2018 de Protección de Datos de Carácter Personal). The participants were identified by research codes, and research information remained confidential.

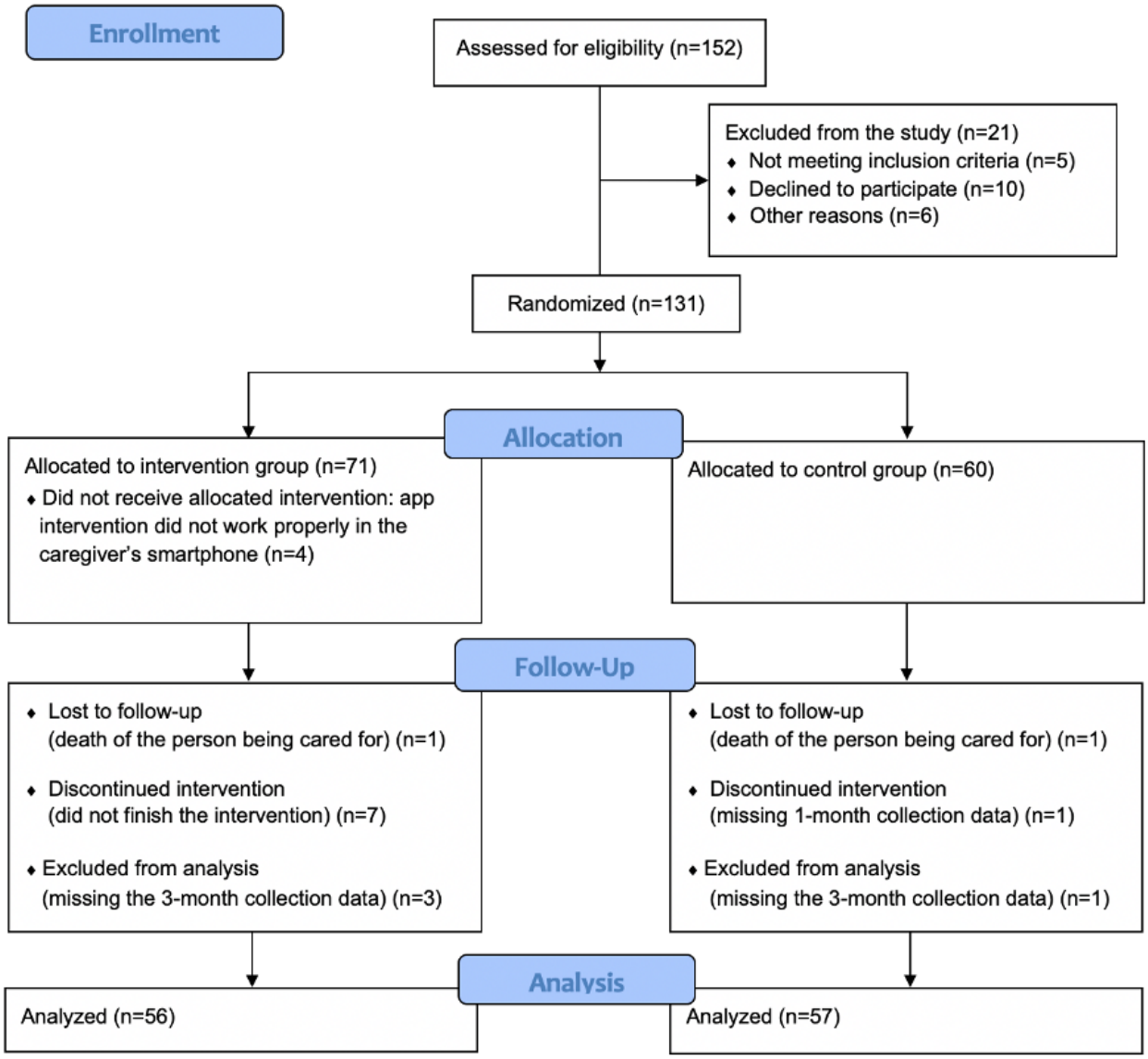
Results

Recruitment in the RCT

A total of 152 caregivers were assessed for eligibility by the nurses in the primary health centers, and 131 agreed to participate and met the inclusion criteria. During the allocation into the intervention group, 4 caregivers were excluded because the app did not work properly on their smartphones. An additional 2 caregivers dropped out of the study because of the death of the person being cared for. Finally, 13 participants were lost to follow-up, including 7 participants in the intervention group who did not complete the app activities).

In all, 113 participants completed the trial: 56 in the intervention group and 57 in the control group. Thus, 79% (56/71) of the sample in the intervention group finished the intervention program. The loss to follow-up rate for the study was 13.7% (18/131) (Figure 3).

Figure 3. Participant flow in the study.



Characteristics of the Participants

Caregivers

The baseline sociodemographic characteristics of the intervention and control groups are presented in Table 1. Overall, there was a large difference in the numbers of male and female caregivers in the study (8.0% versus 92.0%), which reflects the reality of the gender of caregivers. In terms of nationality, only 4.4% of caregivers in the sample were born outside of Spain. Marital status was similar between the groups. The relationship with the person being cared for and his/her situation of dependency was similar between the groups as well.

Table 1. Baseline sociodemographic characteristics of the intervention and control groups.

Sociodemographic characteristics	Control group (n=57)	Intervention group (n=56)	Total (N=113)
Age (years)			
Mean (SD)	64.35 (13.49)	56.89 (13.49)	60.65 (12.37)
Min-max	31-94	28-75	28-94
Gender, n (%)			
Men	3 (5.3)	6 (10.7)	9 (8.0)
Women	54 (94.7)	50 (89.3)	104 (92.0)
Nationality, n (%)			
Spanish	55 (96.5)	53 (94.6)	108 (95.6)
Other	2 (3.5)	3 (5.4)	5 (4.4)
Level of studies, n (%)			
No studies	7 (12.3)	3 (1.8)	10 (7.1)
Middle school	30 (52.6)	14 (25.0)	44 (38.9)
High school	4 (7.0)	9 (16.1)	13 (11.5)
Professional studies	5 (8.8)	12 (21.4)	17 (15.0)
College	11 (19.3)	18 (32.1)	29 (25.7)
No answer	0 (0.0)	2 (3.6)	2 (1.8)
Marital status, n (%)			
Single	10 (17.5)	10 (17.9)	20 (17.7)
Married	37 (64.9)	36 (64.3)	73 (64.6)
Divorced	4 (7.0)	8 (14.3)	12 (10.6)
Widowed	6 (10.5)	2 (3.6)	8 (7.1)
Occupation, n (%)			
Unpaid job	14 (24.6)	6 (10.7)	20 (17.7)
Paid job	22 (38.6)	28 (50.0)	50 (44.2)
Retired	18 (31.6)	18 (32.1)	36 (31.9)
Unemployed	3 (5.3)	4 (7.1)	7 (6.2)
Do you have free time? , n (%)			
Yes	53 (93.0)	55 (98.2)	108 (95.6)
No	4 (7.0)	1 (1.8)	5 (4.4)
Type of caregiver, n (%)			
Primary	54 (94.7)	43 (76.8)	97 (85.8)
Secondary	3 (5.3)	13 (23.2)	16 (14.2)
Relationship with the person being cared for, n (%)			
Parent	33 (57.9)	39 (69.6)	72 (63.7)
Partner	15 (26.3)	10 (17.9)	25 (22.1)
Other but family-related (eg, grandparent, father/mother-in-law)	7 (12.3)	3 (5.4)	10 (8.8)
Other with no family relation (eg, friend, neighbor)	0 (0.0)	2 (3.6)	2 (1.8)
Nonprofessional caregiver contracted by the family or the person cared for	2 (3.5)	2 (3.6)	4 (3.5)

Some differences in variables between the groups had to be taken into consideration. Although age was taken into consideration during the randomization process to avoid an age bias, there was a considerable difference in the mean age of the groups (64.35 years versus 56.89 years in the control and intervention groups, respectively). This result was related to the

4 caregivers who did not receive their allocated intervention. Those 4 participants were over 65 years of age, and although they had smartphones, the phones did not have enough memory to support the app. Researchers considered that the 8-year age difference between the groups was reasonable considering the intended sample and its lower percentage among those who used smartphones in the control group. The participants' level of studies also needs to be taken into consideration, as the intervention group had a greater percentage of individuals with a higher level of education. This difference could be related to the age difference. Regarding occupations, 50.0% (28/56) of participants in the intervention group had a paid job compared with 38.6% (22/57) in the control group. Another difference to be highlighted is that 23.2% (13/56) of participants in the intervention group were secondary caregivers compared with 5.3% (3/57) in the control group.

Caregivers were asked about their perceptions of their personal well-being and tasks. No major differences between the groups were found (Table 2).

Table 2. Characteristics of caregivers' perceptions of their own well-being.

	Mean score (SD)[a]		
Characteristics	Control group (n=57)	Intervention group (n=56)	Total (N=113)
Perception of well-being	7.18 (1.48)	7.08 (1.48)	7.13 (1.39)
Level of burden	6.77 (2.08)	6.34 (2.08)	6.56 (2.32)
Level of satisfaction with caregiving tasks	7.61 (1.82)	7.98 (1.82)	7.80 (1.69)
Level of difficulty of caregiving duties	5.72 (2.66)	5.70 (2.66)	5.71 (2.50)
Level of the demands of the person being cared for	5.28 (3.18)	5.50 (3.18)	5.39 (3.06)

[a]Score from 0 to 10 (0=lowest score, 10=highest score).

Individuals Being Cared For

Caregivers were asked questions about the person in their care. The average age (control group: 84.88 years; intervention group: 81.63 years) and gender (control group: 61.4% women; intervention group: 67.9% women) of the individuals being cared for were similar between the groups (Table 3).

Table 3. Characteristics of individuals being cared for.

Characteristics	Control group (n=57)	Intervention group (n=56)	Total (N=113)
Age (years)			
Mean (SD)	84.88 (8.95)	81.63 (10.16)	83.27 (9.66)
Min-max	58-102	53-100	53-102
Sex, n (%)			
Men	22 (38.6)	18 (32.1)	40 (35.4)
Women	35 (61.4)	38 (67.9)	73 (64.6)
Situation of dependence of the person being cared for, n (%)			
Multiple chronic conditions	25 (43.9)	35 (62.5)	60 (53.1)
Alzheimer	14 (24.6)	10 (17.9)	24 (21.2)
Fragility	12 (21.1)	4 (7.1)	16 (6.2)
Stroke	5 (8.8)	2 (3.6)	7 (14.2)
Tetraplegic	0 (0.0)	2 (3.6)	2 (1.8)
Neoplasia	0 (0.0)	2 (3.6)	2 (1.8)
Parkinson	0 (0.0)	1 (1.8)	1 (0.9)
Schizophrenia	1 (1.8)	0 (0.0)	1 (0.9)

Primary Outcome

The primary outcome was assessed using the PMHQ and ZBI-7. One of the main differences between the groups was related to the higher baseline PMHQ scores in the control group than in the intervention group. In fact, F2–Prosocial attitude, F3–Self-control, and F5–Problem solving and self-actualization are the factors behind this difference. The scores between the groups on the other factors were similar (Table 4).

Table 4. Descriptive results of the Positive Mental Health Questionnaire (PMHQ) (total scores and factor scores) and the 7-item short-form version of the Zarit Caregiver Burden Interview (ZBI-7).

Items	Control group (n=57)		Intervention group (n=56)	
	Mean (SD)	Range	Mean (SD)	Range
PMHQ total score				
Baseline	120.10 (20.32)	84.00-152.00	98.60 (10.96)	51.00-127.00
1 month	118.94 (20.05)	80.00-150.00	101.55 (14.70)	39.00-147.00
3 months	121.68 (19.52)	85.00-156.00	114.41 (20.30)	85.00-152.00
F1: Personal satisfaction				
Baseline	25.52 (4.38)	11.00-32.00	24.87 (4.47)	8.00-32.00
1 month	24.60 (4.30)	11.00-31.00	25.59 (3.95)	9.00-30.00
3 months	23.44 (5.37)	9.00-32.00	24.23 (6.17)	11.00-32.00
F2: Prosocial attitude				
Baseline	16.07 (3.41)	9.00-20.00	11.21 (1.99)	6.00-20.00
1 month	16.32 (3.72)	6.00-20.00	11.33 (2.51)	5.00-20.00
3 months	16.84 (2.96)	10.00-20.00	14.98 (3.42)	9.00-20.00
F3: Self-control				
Baseline	14.40 (3.81)	6.00-19.00	11.32 (2.35)	7.00-16.00
1 month	14.75 (3.83)	7.00-20.00	11.24 (2.90)	5.00-20.00
3 months	15.28 (3.51)	5.00-20.00	13.73 (3.25)	7.00-20.00
F4: Autonomy				
Baseline	16.19 (3.07)	5.00-20.00	16.23 (3.08)	5.00-20.00
1 month	15.54 (3.64)	5.00-20.00	16.40 (3.34)	5.00-20.00
3 months	15.37 (4.21)	5.00-20.00	15.21 (4.52)	5.00-20.00
F5: Problem solving and self-actualization				
Baseline	26.05 (8.49)	10.00-36.00	15.30 (5.26)	9.00-33.00
1 month	26.46 (8.33)	9.00-36.00	15.11 (5.87)	9.00-36.00
3 months	29.21 (6.77)	10.00-36.00	25.39 (8.76)	9.00-36.00
F6: Interpersonal relationship skills				
Baseline	21.86 (3.43)	12.00-28.00	19.63 (2.93)	7.00-25.00
1 month	21.29 (3.75)	12.00-28.00	19.75 (3.27)	7.00-25.00
3 months	21.54 (4.29)	13.00-28.00	20.85 (4.12)	13.00-28.00
ZBI-7[a] score				
Baseline	19.77 (5.38)	9.00-29.00	18.80 (5.64)	7.00-32.00
1 month	20.56 (5.24)	9.00-31.00	18.29 (5.34)	7.00-32.00
3 months	20.70 (5.44)	8.00-32.00	17.69 (5.52)	7.00-29.00

[a]The ZBI-7 was completed by 92 caregivers (43 from the control group, 49 from the intervention group). The data were missing at random, with more than 40% of answer sheets being incomplete.

Assessments were conducted at baseline, and at 1 month and 3 months following the app intervention with the intervention group. Immediately after the intervention (Table 5), there were no statistically significant differences in changes in the PMHQ (P=.24) or ZBI-7 (P=.24) scores between the groups. There were statistically significant—but not clinically relevant—differences in the mean change in F1–Personal satisfaction (0.96; P=.03; d=–0.00) between the groups. Each group obtained statistically significant differences in ZBI-7 scores. While ZBI-7 scores decreased in the intervention group, they increased in the control group. However, there were no clinically relevant differences in the ZBI-7 scores between the groups.

Table 5. Comparison of mean changes at the 1-month follow-up between and within the intervention and control groups.[a]

Mean change at 1-month follow-up	Intervention group–control group		Intervention group		Control group	
	Mean difference of change (95% CI)	*P* value	Mean change (95% CI)	*P* value	Mean change (95% CI)	*P* value
PMHQ[b] total score	1.40 (–3.98 to 5.72)	.24	0.37 (–4.09 to 4.83)	.21	–1.16 (–3.72 to 1.41)	.69
F1–Personal satisfaction	0.96 (–0.62 to 2.47)	.03[a]	0.22 (–1.17 to 1.61)	.07	–0.93 (–1.89 to 0.32)	.20
F2–Prosocial attitude	–0.07 (–1.08 to 0.68)	.18	0.11 (–0.46 to 0.68)	.92	0.25 (–0.24 to 0.74)	.17
F3–Self-control	0.51 (–1.11 to 0.87)	.86	–0.1 (–0.89 to 0.85)	.98	0.35 (–0.43 to 1.13)	.31
F4–Autonomy	0.79 (–0.38 to 1.79)	.14	0.22 (–0.82 to 1.27)	.17	–0.65 (–1.47 to 0.17)	.38
F5–Problem solving and self-actualization	0.05 (–2.65 to 1.13)	.45	–0.26 (–1.45 to 0.93)	.35	0.40 (–0.93 to 1.73)	.95
F6–Interpersonal relationship skills	0.77 (–0.71 to 1.86)	.18	0.09 (–0.99 to 1.18)	.51	–0.59 (–0.58 to 0.08)	.11
ZBI-7[c] score	–0.55 (3.98 to 5.72)	.24	–0.51 (–1.51 to 0.49)	.05	0.79 (0.16 to 1.42)	.01

[a]Analyzed using Mann-Whitney U test and Wilcoxon signed rank test, with *P* values ≤.05 considered statistically significant.

[b]PMHQ: Positive Mental Health Questionnaire.

[c]ZBI-7: 7-item short-form version of the Zarit Caregiver Burden Interview.

The comparison of the mean changes at the 3-month follow-up (Table 6) revealed statistically significant differences in PMHQ scores between the intervention and control groups (11.43; *P*<.001; *d*=0.82). F5–Problem solving and self-actualization stood out for its clinically relevant mean difference of change (5.69; *P*<.001; *d*=0.71). However, F2–Prosocial attitude was the factor with the largest effect size (mean difference of change of 2.47; *P*<.001; *d*=1.18). F1–Personal satisfaction and F6–Interpersonal relationship skills had smaller effect sizes, but there were statistically significant differences between the groups (F1: 1.36; *P*=.05; *d*=0.36; and F6: 1.39; *P*=.04; *d*=0.25). F3–Self-control had a statistically significant difference and a moderate effect size (0.76; *P*=.03; *d*=0.50). F4–Autonomy did not show clinically relevant differences between the groups (–0.25; *P*=.99). ZBI-7 scores showed statistically significant differences with moderate clinically relevant results (–2.03; *P*<.001; *d*=–0.68).

Table 6. Comparison of mean changes at 3-month follow-up between and within the intervention and control groups.[a]

Mean change at 3-month follow-up	Intervention group–control group			Intervention group		Control group	
	Mean difference of change (95% CI)	*P* value	Cohen *d* (95% CI)[b]	Mean change (95% CI)	*P* value	Mean change (95% CI)	*P* value
PMHQ[c] total score	11.43 (8.92 to 22.33)	<.001	0.82 (0.43 to 1.21)	14.94 (9.14 to 20.72)	<.001	3.51 (–1.54 to 8.56)	.78
F1–Personal satisfaction	1.36 (–0.62 to 2.47)	.05	0.36 (–0.01 to 0.75)	–0.96 (–2.73 to 0.82)	.89	–2.32 (–4.41 to –0.24)	.01
F2–Prosocial attitude	2.47 (–0.03 to 3.68)	<.001	1.18 (0.78 to 1.58)	3.73 (2.79 to 4.67)	<.001	1.26 (0.51 to 2.00)	.02
F3–Self-control	0.76 (–1.61 to 1.87)	.03	0.50 (0.12 to 0.88)	2.31 (1.13 to 3.50)	<.001	155 (0.18 to 2.93)	.22
F4–Autonomy	–0.25 (–0.78 to 1.79)	.99	–	–1.44 (2.78 to –0.10)	.26	–1.19 (–2.76 to 0.38)	.29
F5–Problem solving and self-actualization	5.69 (–2.65 to 8.13)	<.001	0.71 (0.32 to 1.09)	10.29 (7.41 to 13.17)	<.001	4.60 (1.96 to 7.25)	.02
F6–Interpersonal relationship skills	1.39 (–0.71 to 2.85)	.04	0.25 (–0.12 to 0.62)	1.00 (–0.30 to 2.30)	.03	–0.39 (–1.73 to 0.95)	.44
ZBI-7[d] score	–2.03 (–5.07 to 1.36)	<.001	0.68 (–1.08 to –0.27)	–1.10 (–2.24 to 0.04)	<.001	0.93 (–0.08 to 1.93)	.04

[a]Analyzed using Mann-Whitney U test and Wilcoxon signed rank test, with *P* values ≤.05 considered statistically significant.

[b]Cohen *d* was only reported when the *P* value of the mean difference was statistically significant.

[c]PMHQ: Positive Mental Health Questionnaire.

[d]ZBI-7: 7-item short-form version of the Zarit Caregiver Burden Interview.

Secondary Outcome

The level of satisfaction with the app was high. Questions 1 to 3 on the ad hoc questionnaire related to the operating system and had an average score of 93.9%. Regarding the organization of the app (questions 4 and 5), users did not report many difficulties in using the app. Users liked the daily phrases, claiming that they helped to improve their mental health (Q6: 43/49, 87.8%). Also, 100% (49/49) of users rated the activities as being easy (Q7), and 91.8% (45/49) of users found the TIVA character to encourage them to continue using the app (Q8). The majority of caregivers (46/49, 93.9%) responded that they would recommend the app and 27 of 49 (56.3%) users indicated that they felt the intervention should last longer (Table 7).

Table 7. Satisfaction with the app by the intervention group (n=49).[a]

Questions	Value, n (%)
Q1: Was the installation of the app easy?	
Yes	41 (83.6)
No	8 (16.4)
Q2: Was the operation of the app well adapted to your mobile device?	
Yes	48 (98.0)
No	1 (2.0)
Q3: Was the response of the mobile app fast?	
Yes	49 (100)
No	0 (0)
Q4: Did the app provide information on the steps to follow?	
Always	42 (85.7)
Frequently	4 (8.2)
Sometimes	3 (6.1)
Q5: Was there ever a time when you did not know what to do?	
Sometimes	13 (26.5)
Never	36 (73.5)
Q6: Do you feel that the daily phrases have helped to improve your mental well-being?	
Yes	43 (87.8)
No	6 (12.2)
Q7: Did you find the activities easy to do?	
Yes	49 (100)
No	0 (0)
Q8: Did the evolution of TIVA encourage you to continue using the mobile app?	
Yes	45 (91.8)
No	4 (8.2)
Q9: Would you recommend the app to other caregivers?	
Yes	46 (93.9)
No	3 (5.4)
Q10: Would you extend the intervention time?	
Yes	27 (56.3)
No	21 (43.8)

[a]Of the 56 participants in the intervention group, app satisfaction data from 7 participants were lost.

Discussion

Principal Findings

This study aimed to asses the effectiveness of a digital intervention program to promote positive mental health among nonprofessional caregivers and to evaluate the usability and satisfaction of the app-based intervention program. The results demonstrated that the implementation of an app-based intervention program for caregivers significantly contributed to enhancing their positive mental health. In fact, results showed that the program produced a larger effect on the caregivers' lives after 3 months of the intervention, which suggests that it is an effective long-term program. In addition, the participants in the study reported a high satisfaction rate with the app.

The intervention program seemed to be effective in relation to the differences in F1–Personal satisfaction scores between the groups at 1-month follow-up. Although the other factors did not reflect any significant differences between the groups, F4–Autonomy and F6–Interpersonal relationship skills obtained a positive mean difference of change, which reflects greater improvement of the scores in the intervention group. In addition to these program results, the level of burden of the control group seemed to increase after 1 month; in contrast, the ZBI-7 scores decreased in the intervention group, although the mean difference of change was not statistically significant.

The results seemed to be more satisfactory after 3 months of the intervention. There were significant differences between the groups on all factors, with the exception of F4–Autonomy, which was the only factor with scores that decreased from 1-month to 3-month follow-up.

The factor with the biggest increase after 3 months of intervention was F5–Problem solving and self-actualization. The mean change in the intervention group increased. These improvements could be related to the increase in confidence and caregiving information promoted by the program. A recent review about caregiver programs highlighted the importance of interventions that include program resolution strategies [47]. These strategies include the need for interventions to be easy and provide a connection with health care providers. Both elements are included in these app-based programs.

The app-based intervention program was also shown to be effective by comparing the scores of the intervention group on F2–Prosocial attitude, F3–Self-control, and F6–Interpersonal relationship skills with those of the control group. Programs involving group social support have already been tested and demonstrate their efficacy in caregivers. However, a few studies demonstrate that online strategies could promote those elements. In fact, a meta-analysis performed in 2017 highlighted the need to promote this type of program, as there are many isolated caregivers who cannot access on-site caregiver support programs [48]. The results of our study provide the first step in promoting this kind of program to foster prosocial attitudes through ICT strategies and with the support of primary health care nurses.

Another outcome of the study relates to F1–Personal satisfaction. In fact, this factor slightly increased in the intervention group after the first month, and decreased after 3 months. Therefore, interventions related to this factor are effective to prevent a further decrease over time. The research team believes that this result may be related to the messages of coaching and motivation that the app included, which were designed by the research team, including psychologists and nurses. Those motivational phrases were created with the aim of recognizing the caregiver's task and its value.

In this section, it is important to highlight that after the app-based intervention program ended, users had the option of keeping the app on their smartphones, which would continue to display the question "Hello, how are you today?" every day. Furthermore, if they used the app, they could see the final evolution of TIVA and access the website [35]. Although not formally registered, this element seemed to satisfy the caregivers, who had made their request to the nurses responsible for data collection. A previous study shows that caregiver support is strongly related to their need for more care-related information and more information on available resources [49]. These needs could be related to our outcomes and to increased effectiveness in the intervention group after 3 months.

Results related to the ZBI-7 showed that after 3 months, the intervention group had a statistically significant decrease in caregiver burden, with a moderate size effect between the groups. However, researchers consider that these results are inconclusive considering the lack of information recorded in this study about the evolution of the person being cared for, which could have varied considerably after 3 months of intervention. Evidence shows how caregiver burden is directly associated with the caregiver's duties, and further studies should evaluate the evolution of the characteristics of the caregiving situation [50,51].

The secondary outcome related to the app-based intervention program was satisfactory in terms of user acceptance. Users found the app easy to use, which was one of the goals of the app's development. The literature highlights the need for these types of online programs to have an easy operating system to prevent dropouts [52,53]. According to informal feedback from caregivers and nurses, the character of the program, TIVA, was one of the reasons for the high degree of loyalty of the intervention program users to the app. In fact, many caregivers expressed several positive emotions after witnessing TIVA's evolution every day.

Similar to applications designed for other types of users, new technologies designed to care for caregivers can increase satisfaction and motivation by empowering them, which promotes healthy changes in their behaviors and improves their quality of life [53]. Currently available applications have worked on aspects related more to the care of the chronic patient than to the well-being of the caregiver. With this application, and as recommended by other authors, behavioral interventions have been performed using "persuasive health technologies," which are becoming much more promising approaches that encourage healthy behaviors [54]. Interventions that improve positive mental health should be an integral part of emotional support for caregivers and used in conjunction with advice from health professionals, improving their relationship without overloading face-to-face care. Most of these programs were conducted in a

traditional face-to-face format. Nursing has the responsibility of initiating changes with the use of new technology and, as shown with the TIVA program, mHealth must be used as an indispensable working tool to reach all users, including those who cannot be reached in person [55].

Several studies emphasize the need to implement online intervention programs with gamification [21,52,53]. During the TIVA app's creation, several options were considered to implement gamification in order to ensure the caregivers' involvement during the 28 days of the intervention program. A game-based option was not considered, as it would have required more of the caregiver's time and thus posed a risk of increasing dropouts.

On the other hand, the program allowed the self-management of time when planning the daily operation of the program. In this way, adherence was facilitated and the risk of dropouts was also reduced, as was also shown in another study [56].

Although the increase in the use of new technologies opens up a new space in the promotion of health for patients and caregivers, we identified in consonance with other authors that most of the existing digital support programs were more commercial than functional and it is necessary to promote studies that provide scientific evidence on the impact of these applications in the health field [57].

Therefore, the TIVA app seems to have been a good choice for this app-based intervention program. However, we can only express qualitative data on this outcome, so this should be assessed in further studies.

Limitations and Future Work

A number of important issues remain to be addressed by future research. First, the basic characteristics of the participants showed some differences. There was an 8-year difference in the mean ages of the groups, which was related to individuals' knowledge of how to use a smartphone. Further studies may need to compare the same age ranges with a minimum sample size for each group. This was not possible in our case, as the sample was not large enough to do this comparison. A similar problem occurred in relation to the level of education: the intervention group appeared to have a higher level of education than the control group. This could also have led to a bias in the results given that previous studies have linked the level of education to better coping strategies and level of resilience [58-60]. Despite these differences in sociodemographic characteristics, both groups obtained similar scores on characteristics of well-being and the person being cared for, which strengthens the validity of the results. Nevertheless, further studies with larger sample sizes should be undertaken to avoid possible biases.

Second, the level of positive mental health measured using the PMHQ appeared to be different at baseline for the intervention and control groups. The intervention group had a significantly lower level of positive mental health. This could have resulted in a bias in our study, as it might have been easier to show an improvement in the intervention group than in the control group. However, user satisfaction seems to be consistent in the results. Nevertheless, we believe that further studies need to be conducted to consider the baseline level of positive mental health, as well as the type of caregiver, as criteria to be included during the randomization process to verify the results.

Third, another limitation of the study is related to the type of caregiver. There was a higher percentage of secondary caregivers in the intervention group than in the control group. However, as the analysis was conducted with a consideration of pre- and postintervention scores, significant differences are valid.

Fourth, at the baseline of the study, caregiver burden was assessed using the ZBI-7. At the same time, data related to the characteristics of the person being cared for and information about the caregiving tasks (including the time per week dedicated to caregiving and caregiving duties) were recorded. However, while the ZBI-7 was also assessed after 1 month and after 3 months, the rest of the information was not. We believe that results related to caregiver burden should be carefully considered, and further studies should include an assessment of changes of caregivers' duties and characteristics of the person cared for throughout the study.

Fifth, the app-based intervention program includes the features described in the "Methods," which includes a total of 20 activities. Those activities, developed by the research team, were created with consideration of the 10 recommendations from Lluch-Canut's model for promoting positive mental health [36,37]. Each activity cannot be linked directly to a factor from the model, as each activity promotes more than one factor. This is the reason why a redesign of the activities/features cannot be linked directly to a factor. Based on the results of this study, we consider, as a future line of research, a qualitative evaluation of the actual design of the app-based intervention program with the research team, stakeholders, and caregivers who participated in this study in the intervention group.

Conclusions

The app-based intervention program analyzed in this study can be considered effective, with a high degree of user satisfaction. The development of a 28-day program with interventions based on Lluch-Canut's positive mental health framework was successful in terms of user satisfaction and usability. The effectiveness of the program as it relates to positive mental health was demonstrated with positive results. However, we conclude that the program's effectiveness should be verified by further studies where the baseline level of positive mental health and type of caregiver are included as randomization criteria.

This app-based intervention program showed increased positive mental health levels after 1 month and 3 months. The factors with the greatest long-term effect were F2–Prosocial attitude, F3–Self-control, and F5–Problem solving and self-actualization. However, activities related to F4–Autonomy should be revised, as the effectiveness was only maintained after 1 month and not after 3 months.

The app's user satisfaction was encouraging. Users considered the app's operational system and activities to be appropriate. The character created especially for the app, TIVA, with a programmed evolution after each activity, was considered to

be a fundamental element to encourage users to continue with the app-based intervention program.

The results of the study encourage the promotion of app-based intervention programs for caregivers and endorse their effectiveness and user satisfaction. We believe that a further study involving stakeholders and participants should be considered to evaluate the adequacy of the activities in the actual app-based program in order to address the need to redesign activities/features of the app-based intervention program to increase the program's effectiveness.

Acknowledgments

The authors gratefully acknowledge the support of study participants and partner organizations. The study group members of project PI16/01576 who were involved in the implementation of the clinical trial are Anna Falcó-Pegueroles, Montserrat Boqué-Cavallé, Amalia Morro-Grau, Belen Vargas-Motera, Gerard Mora-López, Monica Mulet, Aurelia Sánchez, Rut Martí-Arqués, José Ramón Martínez, and Rosa Raventós-Torner.

This study was funded by Instituto de Salud Carlos III (Spain) through the project reference number PI16/01576 (cofounded by FEDER founds/European Regional Development Found). It has a financial role: EUR € 9689 (US $11,772.62) the first year, EUR € 9075 (US $11,026.58) the second year, and EUR € 9680 (US $11,761.68) the third year. The funds have been used to cover the design and creation of the app. These funders were not involved in the design of the Spanish study, data collection, study management, data analysis, data interpretation, writing of the report, or the decision to submit the report for publication.

Authors' Contributions

All authors have approved the final version of the manuscript and meet at least one of the following criteria: substantial contributions to conception and design, data acquisition, or data analysis and interpretation; and drafting the article or critically revising it for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 1975 KB - mhealth_v9i1e21708_app1.pdf]

References

1. WHO. Primary Health Care – Now More Than Ever. The World Health Report 2008. World Health Organization: World Health Organization; 2008. URL: https://www.who.int/whr/2008/en/ [accessed 2019-04-20]
2. Generalitat de C. Departament de Salut. Pla de Salut. Pla Salut Catalunya. Barcelona: Departament de Salut; 2016:2016.
3. Ruiz Valencia LF, Gordillo Sierra AM, Galvis López CR. Factores condicionantes básicos en cuidadores informales de pacientes crónicos en el domicilio. Rev Cuid 2019 Sep 13;10(3). [doi: 10.15649/cuidarte.v10i3.608]
4. Rosas C, Neri AL. Quality of life, burden, family emotional support: a model for older adults who are caregivers. Rev. Bras. Enferm 2019;72(suppl 2):169-176. [doi: 10.1590/0034-7167-2018-0439]
5. Ngangana PC, Davis BL, Burns DP, McGee ZT, Montgomery AJ. Intra-family stressors among adult siblings sharing caregiving for parents. J Adv Nurs 2016 Aug 01;72(12):3169-3181. [doi: 10.1111/jan.13065]
6. Hamama L, Sharon M. Posttraumatic Growth and Subjective Well-Being among Caregivers of Chronic Patients: A Preliminary Study. J Happiness Stud 2012 Dec 20;14(6):1717-1737. [doi: 10.1007/s10902-012-9405-8]
7. Rodrigues AM, Ferré-Grau C, Ferreira PL. Being an Informal Caregiver of a Person with a Pressure Ulcer in the Azores Islands. Advances in Skin & Wound Care 2015;28(10):452-459. [doi: 10.1097/01.asw.0000471191.11548.dd]
8. Deek H, Hamilton S, Brown N, Inglis SC, Digiacomo M, Newton PJ, et al. Family-centred approaches to healthcare interventions in chronic diseases in adults: a quantitative systematic review. J Adv Nurs 2016 Jan 10;72(5):968-979. [doi: 10.1111/jan.12885]
9. Pucciarelli G, Ausili D, Galbussera AA, Rebora P, Savini S, Simeone S, et al. Quality of life, anxiety, depression and burden among stroke caregivers: A longitudinal, observational multicentre study. J Adv Nurs 2018 May 31;74(8):1875-1887. [doi: 10.1111/jan.13695]
10. Machado BM, Dahdah DF, Kebbe LM. Cuidadores de familiares com doenças crônicas: estratégias de enfrentamento utilizadas no cotidiano. Cad. Bras. Ter. Ocup 2018;26(2):299-313. [doi: 10.4322/2526-8910.ctoAO1188]
11. Zabalegui Yárnoz A, Navarro Díez M, Cabrera Torres E, Gallart Fernández-Puebla A, Bardallo Porras D, Rodríguez Higueras E, et al. Eficacia de las intervenciones dirigidas a cuidadores principales de personas dependientes mayores de 65 años. Una revisión sistemática. Revista Española de Geriatría y Gerontología 2008 May;43(3):157-166. [doi: 10.1016/s0211-139x(08)71176-4]

12. Becqué YN, Rietjens JAC, Heide A, Witkamp E. The effectiveness of a nurse‐led intervention to support family caregivers in end‐of‐life care: Study protocol for a cluster randomized controlled trial. J Adv Nurs 2020 Mar 03;76(5):1266-1272. [doi: 10.1111/jan.14326]
13. Kuo L, Huang H, Liang J, Kwok Y, Hsu W, Su P, et al. A randomized controlled trial of a home-based training programme to decrease depression in family caregivers of persons with dementia. J Adv Nurs 2016 Oct 17;73(3):585-598. [doi: 10.1111/jan.13157]
14. Sousa L, Sequeira C, Ferré-Grau C. 'Living together with dementia' – Conceptual validation of training programme for family caregivers: Innovative practice. Dementia 2018 Mar 07;19(4):1333-1342. [doi: 10.1177/1471301218762565]
15. Teixeira SMA, Coelho JCF, Sequeira CADC, Lluch i Canut MT, Ferré-Grau C. The effectiveness of positive mental health programs in adults: A systematic review. Health Soc Care Community 2019 May 29. [doi: 10.1111/hsc.12776]
16. Rempel GR, Neufeld A, Kushner KE. Interactive Use of Genograms and Ecomaps in Family Caregiving Research. J Fam Nurs 2016 Jul 24;13(4):403-419. [doi: 10.1177/1074840707307917]
17. Boele FW, van Uden-Kraan CF, Hilverda K, Weimer J, Donovan HS, Drappatz J, et al. Neuro-oncology family caregivers' view on keeping track of care issues using eHealth systems: it's a question of time. J Neurooncol 2017 May 26;134(1):157-167. [doi: 10.1007/s11060-017-2504-y]
18. Naganathan G, Kuluski K, Gill A, Jaakkimainen L, Upshur R, Wodchis WP. Perceived value of support for older adults coping with multi-morbidity: patient, informal care-giver and family physician perspectives. Ageing and Society 2015 Aug 03;36(9):1891-1914. [doi: 10.1017/s0144686x15000768]
19. Duran-Parra M, Torres CC, Arboleda LB, Rivera-Carvajal R, Franco S, Santos J. Effectiveness of an Educational Nursing Intervention on Caring Ability and Burden in Family Caregivers of Patients with Chronic Non-Communicable Diseases. A Preventive Randomized Controlled Clinical Trial. Invest Educ Enferm 2019 Feb;37(1):e4. [doi: 10.17533/udea.iee.v37n1e04]
20. Klemm PR, Hayes ER, Diefenbeck CA, Milcarek B. Online support for employed informal caregivers: psychosocial outcomes. Comput Inform Nurs 2014 Jan;32(1):10-20. [doi: 10.1097/CIN.0000000000000009] [Medline: 24284908]
21. Lorca-Cabrera J, Ferré-Grau C, Martí-Arques R, Raigal-Aran L, Falcó-Pegueroles A, Albacar-Riobóo N. Effectiveness of health web-based and mobile app-based interventions designed to improve informal caregiver's well-being and quality of life: A systematic review. International Journal of Medical Informatics 2020 Feb;134:104003. [doi: 10.1016/j.ijmedinf.2019.104003]
22. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
23. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
24. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A Web-Based Psychoeducational Program for Informal Caregivers of Patients With Alzheimer's Disease: A Pilot Randomized Controlled Trial. J Med Internet Res 2015 May 12;17(5):e117. [doi: 10.2196/jmir.3717]
25. Lleixá-Fortuño M, García-Martínez M, Lluch-Canut MT, Roldán-Merino J, Espuny-Vidal C, Ferré-Grau C, et al. Diseño de una web 2.0 para cuidadoras de personas con problemas crónicos de salud. Pixel-Bit 2015(47):149-161. [doi: 10.12795/pixelbit.2015.i47.10]
26. Lorca-Cabrera J, Martí-Arques R, Albacar-Rioboo N, García-García M, Mora-López G, Lleixà-Fortuño M. Technological support for non-professional caregivers for patients with chronic conditions. Metas Enfermeria 2016;19:21 [FREE Full text]
27. Green Paper on mobile Health (mHealth). European Commission.: European Commission; 2014. URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2019-03-01]
28. Current Status and Future Trends in Mobile Health. Researc2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-03-03]
29. Avanesova AA, Shamliyan TA. Worldwide implementation of telemedicine programs in association with research performance and health policy. Health Policy and Technology 2019 Jun;8(2):179-191. [doi: 10.1016/j.hlpt.2019.04.001]
30. Papa R, Efthymiou A, Lamura G, Piccinini F, Onorati G, Papastavrou E, et al. Review and Selection of Online Resources for Carers of Frail Adults or Older People in Five European Countries: Mixed-Methods Study. JMIR Mhealth Uhealth 2020 Jun 17;8(6):e14618. [doi: 10.2196/14618]
31. Grossman MR, Zak DK, Zelinski EM. Mobile Apps for Caregivers of Older Adults: Quantitative Content Analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162. [doi: 10.2196/mhealth.9345]
32. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2016 Oct 18. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
33. Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]

34. Adler NJ, Aycan Z. Cross-Cultural Interaction: What We Know and What We Need to Know. Annu. Rev. Organ. Psychol. Organ. Behav 2018 Jan 21;5(1):307-333. [doi: 10.1146/annurev-orgpsych-032117-104528]
35. Cuidadoras cronicos. cuidadorascronicos. Tarragona: Universitat Rovira i Virgili; 2014. URL: www.cuidadorascronicos.com [accessed 2019-03-12]
36. Lluch-Canut T. Decalogue of positive mental health. In: Universitat de Barcelona. Barcelona: Repositori Universitat de Barcelona; 2011.
37. Lluch-Canut M. Construcción De Una Escala Para Evaluar La Salud Mental Positiva. Universitat de Barcelona 1999:319 [FREE Full text]
38. Ferré-Grau C, Raigal-Aran L, Lorca-Cabrera J, Ferré-Bergadá M, Lleixà-Fortuño M, Lluch-Canut MT, et al. A multi-centre, randomized, 3-month study to evaluate the efficacy of a smartphone app to increase caregiver's positive mental health. BMC Public Health 2019 Jul 5;19(1). [doi: 10.1186/s12889-019-7264-5]
39. Garrett AD, Lipsey MW. Design Sensitivity: Statistical Power for Experimental Research. The Statistician 1993;42(2):195. [doi: 10.2307/2348989]
40. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
41. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Computers in Human Behavior 2009 May;25(3):749-760. [doi: 10.1016/j.chb.2009.02.003]
42. Lluch-Canut T, Puig-Llobet M, Sánchez-Ortega A, Roldán-Merino J, Ferré-Grau C. Assessing positive mental health in people with chronic physical health problems: correlations with socio-demographic variables and physical health status. BMC Public Health 2013 Oct 5;13(1). [doi: 10.1186/1471-2458-13-928]
43. Martín Carrasco M, Salvadó I, Nadal Álava S, Laurindo Carlos M, Rico JM, Lanz P, et al. Adaptación para nuestro medio de la escala de sobrecarga del cuidador (Caregiver Burden Interview) de Zari. Revista de gerontología 1996;6:338-345 [FREE Full text]
44. Regueiro Martínez A, Pérez-Vázquez A, Gómara Villabona SM, Carmen Ferreiro Cruz M. Escala de Zarit reducida para la sobrecarga del cuidador en atención primaria. Atención Primaria 2007 Apr;39(4):185-188. [doi: 10.1157/13100841]
45. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
46. 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]
47. Lorio CM, Romano M, Woods JJ, Brown J. A Review of Problem Solving and Reflection as Caregiver Coaching Strategies in Early Intervention. Infants & Young Children 2020;33(1):35-70. [doi: 10.1097/iyc.0000000000000156]
48. Parker Oliver D, Patil S, Benson JJ, Gage A, Washington K, Kruse RL, et al. The Effect of Internet Group Support for Caregivers on Social Support, Self-Efficacy, and Caregiver Burden: A Meta-Analysis. Telemed J E Health 2017 Aug;23(8):621-629. [doi: 10.1089/tmj.2016.0183] [Medline: 28328392]
49. Bruening R, Sperber N, Miller K, Andrews S, Steinhauser K, Wieland GD, et al. Connecting Caregivers to Support: Lessons Learned From the VA Caregiver Support Program. J Appl Gerontol 2019 Jan 18;39(4):368-376. [doi: 10.1177/0733464818825050]
50. Sutcliffe C, Giebel C, Bleijlevens M, Lethin C, Stolt M, Saks K, et al. Caring for a Person With Dementia on the Margins of Long-Term Care: A Perspective on Burden From 8 European Countries. Journal of the American Medical Directors Association 2017 Nov;18(11):967-973.e1. [doi: 10.1016/j.jamda.2017.06.004]
51. Lethin C, Renom-Guiteras A, Zwakhalen S, Soto-Martin M, Saks K, Zabalegui A, et al. Psychological well-being over time among informal caregivers caring for persons with dementia living at home. Aging & Mental Health 2016 Jul 27;21(11):1138-1146. [doi: 10.1080/13607863.2016.1211621]
52. Phongtankuel V, Shalev A, Adelman RD, Dewald R, Dignam R, Baughn R, et al. Mobile Health Technology Is Here—But Are Hospice Informal Caregivers Receptive? Am J Hosp Palliat Care 2018 Jun 03;35(12):1547-1552. [doi: 10.1177/1049909118779018]
53. Davoudi M, Najafi Ghezeljeh T, Vakilian Aghouee F. Effect of a Smartphone-Based App on the Quality of Life of Patients With Heart Failure: Randomized Controlled Trial. JMIR Nursing 2020 Nov 2;3(1):e20747. [doi: 10.2196/20747]
54. McLean A. mHealth Apps as Effective Persuasive Health Technology: Contextualizing the "Necessary" Functionalities. JMIR Nursing 2020 Jul 15;3(1):e19302. [doi: 10.2196/19302]
55. Wilson R, Small J. Care Staff Perspectives on Using Mobile Technology to Support Communication in Long-Term Care: Mixed Methods Study. JMIR Nursing 2020 Sep 29;3(1):e21881. [doi: 10.2196/21881]
56. Potts C, Bond R, Ryan A, Mulvenna M, McCauley C, Laird E, et al. Ecological Momentary Assessment Within a Digital Health Intervention for Reminiscence in Persons With Dementia and Caregivers: User Engagement Study. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17120 [FREE Full text] [doi: 10.2196/17120] [Medline: 32420890]
57. Siqueira do Prado L, Carpentier C, Preau M, Schott A, Dima AL. Behavior Change Content, Understandability, and Actionability of Chronic Condition Self-Management Apps Available in France: Systematic Search and Evaluation. JMIR Mhealth Uhealth 2019 Aug 26;7(8):e13494. [doi: 10.2196/13494]

58. Lai R, Tensil M, Kurz A, Lautenschlager NT, Diehl-Schmid J. Perceived Need and Acceptability of an App to Support Activities of Daily Living in People With Cognitive Impairment and Their Carers: Pilot Survey Study. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16928. [doi: 10.2196/16928]
59. Jannah R, Haryanto J, Kartini Y, Yusuf A. Factors associated with perceived health among caregivers who are taking care of elderly with schizophrenia. International Journal of Nursing and Health Services 2019;2(4):213-218. [doi: 10.35654/ijnhs.v2i4.153]
60. Meyer OL, Liu X, Tancredi D, Ramirez AS, Schulz R, Hinton L. Acculturation level and caregiver outcomes from a randomized intervention trial to enhance caregivers' health: evidence from REACH II. Aging & Mental Health 2017 Apr 24;22(6):730-737. [doi: 10.1080/13607863.2017.1317330]

Abbreviations

ICTs: information and communication technologies
mHealth: mobile health
PMHQ: Positive Mental Health Questionnaire
RCT: randomized controlled trial
ZBI-7: 7-item short-form version of the Zarit Caregiver Burden Interview

Edited by T Rashid Soron; submitted 23.06.20; peer-reviewed by W Zhang, A Hall, R Miranda, S Patil, C Cheng; comments to author 03.08.20; revised version received 24.09.20; accepted 22.12.20; published 22.01.21.

Please cite as:
Ferré-Grau C, Raigal-Aran L, Lorca-Cabrera J, Lluch-Canut T, Ferré-Bergadà M, Lleixá-Fortuño M, Puig-Llobet M, Miguel-Ruiz MD, Albacar-Riobóo N
A Mobile App–Based Intervention Program for Nonprofessional Caregivers to Promote Positive Mental Health: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(1):e21708
URL: http://mhealth.jmir.org/2021/1/e21708/
doi:10.2196/21708
PMID:33480852

Original Paper

Young People's Experiences Using an On-Demand Mobile Health Sexual and Reproductive Health Text Message Intervention in Kenya: Qualitative Study

Jefferson Mwaisaka[1,2], MA; Lianne Gonsalves[3,4,5], PhD; Mary Thiongo[1], MSc; Michael Waithaka[1], MSc; Hellen Sidha[6], MPH; Otieno Alfred[7], PhD; Carol Mukiira[8], MA; Peter Gichangi[1,9,10], PhD

[1]International Centre for Reproductive Health, Kenya, Mombasa, Kenya

[2]Department of Population Family and Reproductive Health, College of Health Sciences, University of Ghana, Accra, Ghana

[3]UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Sexual and Reproductive Health and Research, World Health Organization, Geneva, Switzerland

[4]Swiss Tropical and Public Health Institute, Basel, Switzerland

[5]University of Basel, Basel, Switzerland

[6]National Council for Populations and Development, Kenya, Nairobi, Kenya

[7]Population Studies and Research Institute, University of Nairobi, Nairobi, Kenya

[8]African Institute for Development Policy, Nairobi, Kenya

[9]Ghent University, Ghent, Belgium

[10]Technical University of Mombasa, Mombasa, Kenya

Corresponding Author:
Jefferson Mwaisaka, MA
International Centre for Reproductive Health, Kenya
Tom Mboya Street
Mombasa
Kenya
Phone: 254 723 206 132
Email: jeffmwaisaka@gmail.com

Abstract

Background: Digital health usability assessments can help explain how well mobile health (mHealth) apps targeting young people with sexual and reproductive health (SRH) information performed and whether the intended purpose was achieved. However, few digital health assessments have been conducted to evaluate young people's perceptions regarding mHealth system interactions and content relevance on a wide range of SRH topics. In addition, the majority of randomized controlled trials (RCTs) have focused on push messaging platforms; therefore, the mHealth field lacks sufficient RCTs investigating on-demand mHealth SRH platforms.

Objective: The objective of this study was to explore young people's experiences using an on-demand SRH mHealth platform in Kenya.

Methods: We used qualitative data related to the usability of an mHealth platform, Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome (ARMADILLO), collected at the end of the intervention period. A total of 30 in-depth interviews (IDIs) were held with the intervention participants (15 women and 15 men) to elicit their experiences, opinions, and perspectives on the design and content of the ARMADILLO platform. The study participants were randomly selected from a list of intervention arm participants to participate in the IDIs. The interviews were later transcribed verbatim, translated into English, and coded and analyzed thematically using NVivo version 12 software (QSR International).

Results: Respondents reported varied user experiences and levels of satisfaction, ranging from ease of use by the majority of the respondents to systematic frustrations that prevented some participants from progressing to other stages. Interesting features of the mHealth platform included the immediate response participants received when requesting messages, weekly remunerated quizzes, and perceived *ability* of educative and informative content and messages to change behaviors. Proposed enhancements to the platform included revising some concepts and words for easy understanding and increasing the interactivity of the platform,

whereby young people could seek clarity when they came across difficult terms or had additional questions about the information they received.

Conclusions: The importance of understanding the range of health literacy and technological variations when dealing with young people cannot be overemphasized. Young people, as mHealth end users, must be considered throughout intervention development to achieve optimum functionality. In addition, young people targeted with mHealth SRH interventions must be sensitized to the interactions on mHealth platforms or any other digital health apps if implemented in a nonresearch setting for optimal use by the targeted audience.

(JMIR Mhealth Uhealth 2021;9(1):e19109) doi:10.2196/19109

KEYWORDS

mHealth; mobile phones; sexual and reproductive health; Kenya

Introduction

Background

Globally, sexual and reproductive health (SRH) programs use mobile phones as communication platforms to reach out to young people aged between 10 and 24 years with information and services on a wide range of SRH topics [1]. The proliferation of mobile phones and the significant advancements in wireless technologies provide innovative ways to deliver health information to young people [2]. In addition, the flexibility, accessibility, confidentiality, and convenient nature of mobile phones make them appealing to young people seeking sensitive SRH information.

Assessing the usability and acceptability of various mobile health (mHealth) interventions designed to provide factual SRH information to adolescents and young people is an essential step in designing and implementing sustainable, appealing mHealth programs [3]. The World Health Organization (WHO) indicates that *usability* determines whether the mHealth intervention can be used as intended by the users by focusing on the quality of interaction between the user and the technology [4]. Usability assessments can be used to identify needs, design, develop, implement, and evaluate appropriate and effective mHealth interventions. Usability assessments involving young people as the targeted users can help explain how well the mHealth app functioned and whether the platform achieved its intended purpose [5].

Several studies have explored the potential of pushed text messaging to improve knowledge and affect SRH behavior change among young people and hard-to-reach populations [1,6-9]. However, there has been less research on interactive on-demand systems where health information is triggered by users' requests, traditionally via SMS menus [10]. One enabler of mHealth adoption in developing countries such as Kenya is the interactive interface feature of health apps [11]. Interactive mHealth programs have been found to be feasible, acceptable, and potentially effective in supporting behavior change such as smoking cessation [12] and depressive symptoms among individuals with spinal cord injury [13]. In Kenya, interactive on-demand 2-way mHealth interventions have been shown to be superior to 1-way text message interventions, particularly for improving medication adherence among HIV-positive individuals; however, a significant knowledge gap still exists as to why on-demand 2-way mHealth interventions are deemed superior [14,15]. This could be because of inadequate formal usability and acceptability assessments to compare the 2 sets of mHealth delivery models. This lack of insight may contribute to future challenges such as integrating mHealth into health services and the adoption of mHealth interventions by policy makers for scale-up purposes [14]. In addition, existing on-demand mHealth apps in Kenya such as *Daktari popote*, meaning *Doctor anywhere*, *Hello Doctor*, and MedAfrica, which allow users to search and filter health information and locate doctors and hospitals, are largely internet based and would require users to have smartphones to access health services [16].

Specifically, for mHealth text message interventions targeting adolescents and young people, there are very few, if any, formal usability and acceptability assessments of interactive and on-demand mHealth SRH interventions, particularly in low-income countries [17]. The Mobile for Reproductive Health (m4RH) program is a unique program designed to expand family planning information, albeit for the general public, which is known to have been evaluated. The evaluation of the m4RH presented user satisfaction information with respect to the mHealth platform's contents [18,19]; however, this evaluation did not capture the complete usability of the mHealth system, including user interactions and how effective the interactive mHealth platform was perceived by the users. For this reason, there is still a need for evaluations, particularly for those mHealth interventions designed to improve SRH knowledge and service uptake among young people, to focus on understanding targeted real users' interactions. Such assessments can significantly benefit mHealth developers, users, and policy makers. The expected functionality of mHealth apps and recommendations gathered from such assessments will provide helpful guides for improving the usability of mHealth apps targeting young people.

Study Objectives

This study sought to evaluate the usability of an interactive on-demand mHealth platform aimed at providing SRH information to young people in Kenya. Specifically, the assessment determined whether and to what extent young people found the digital intervention user friendly and useful by examining their experiences and perceptions regarding the system's design, ease of interaction, and platform contents.

Methods

Overview

This usability assessment was nested within the broader Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome (ARMADILLO) study. ARMADILLO was a 3-arm (*intervention*, *control*, and *contact*) individual randomized controlled trial (RCT; International Standard Randomized Controlled Trial Number [ISRCTN]: 85156148) aimed at developing and evaluating an on-demand system for young people to access and receive SRH information through SMS. The trial has been described in detail elsewhere [20]. In brief, the larger mHealth (ARMADILLO) study was conducted in Kwale County of Kenya, between January and July 2018, and with young male and female study participants aged 18 to 24 years. Compared with other counties in the coastal region of Kenya, Kwale County is among the top 2 counties with the highest prevalence of teenage pregnancies (24%), which is higher than both regional and national averages reported at 20.8% and 18%, respectively [21]. In addition, according to the Kenya Demographic and Health Survey of 2014, Kwale County had the lowest median age at first sexual intercourse among women, 16.6 years in the entire coastal region of Kenya, and the fifth lowest nationally [21]. Moreover, the county had a lower contraceptive prevalence rate (CPR) of 38.2% compared to the national CPR of 53.2% [21].

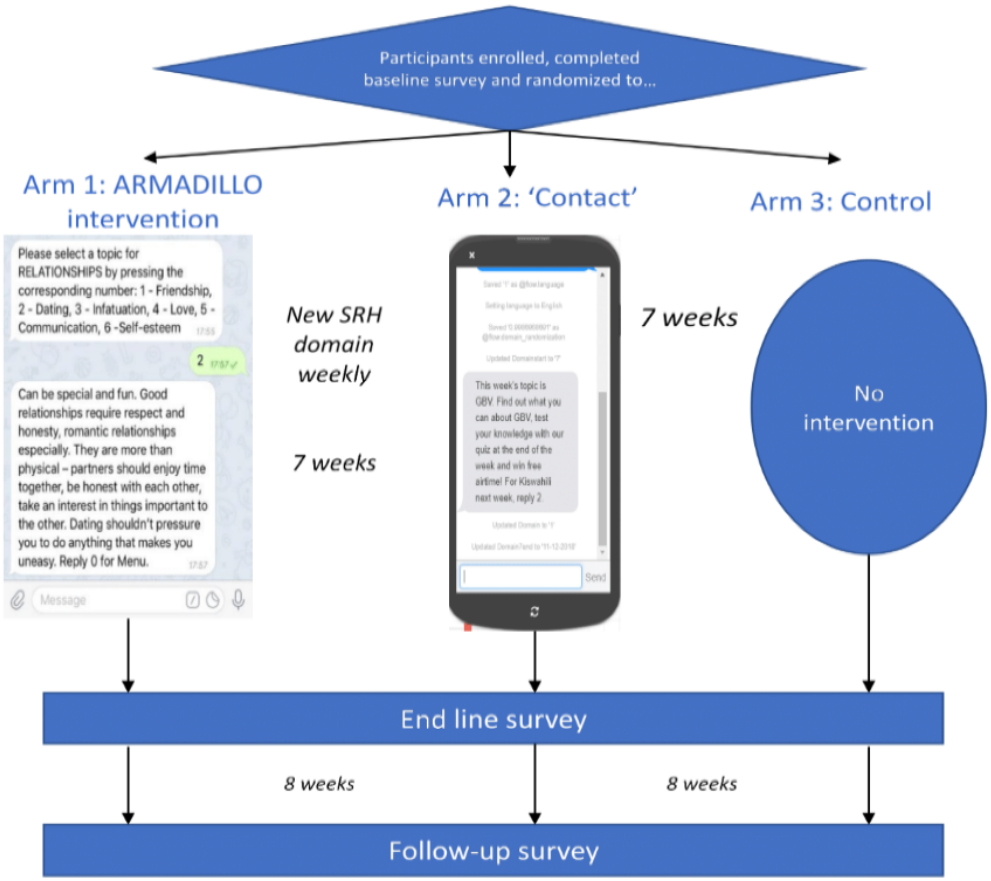
Participants' Recruitment

The ARMADILLO study used household-based surveys and multistage random sampling to recruit participants for the main study. We used census data obtained from the Kenya National Bureau of Statistics to identify and enumerate all households in the study zone. This was followed by a census of households with eligible study participants. A list of potential participants to be sampled was then randomly generated using a computer-based random number generator. The research team then made a second visit to the selected households to inform eligible participants about the study and seek consent from them. A total of 740 young men and women aged 18 to 24 years were selected for the study. This sample provided 80% power to detect a 10% change in mean number of myths believed by young people from baseline to end line, assuming that the baseline level of belief was 0.55, with an assumed SD of 0.30, accounting for a dropout rate of 20% [20]. To minimize contamination, only 1 eligible youth was selected from each household. This was followed by randomly allocating participants into intervention, contact, or control groups using a computer-based algorithm. The allocation followed a 1:1:1 ratio; this allocation was overseen by a member of the research team who did not interact with the study participants. After the arm allocation, the participants' 7-week interaction with the relevant arm began the following day [20].

Intervention Procedure

The intervention period was from January 20, 2018, to March 10, 2018, lasting 7 weeks with outcome assessments conducted at baseline and end line. The trial's primary outcome was to assess the ability of on-demand SRH information delivered via SMS to dispel myths and misconceptions around contraception. Each week during the intervention period (Figure 1), participants randomized to the intervention arm received a given SRH domain pushed to their phones and subsequently requested information (subdomains) related to that particular topic. The contact arm participants received weekly topics only and were instructed to learn on their own. The control arm received no messages. This paper highlights the experiences of only intervention arm participants.

Figure 1. Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome trial diagram. ARMADILLO: Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome; SRH: sexual and reproductive health.



Intervention Arm Design

Following enrollment into the system, the intervention arm participants received an introductory SMS followed by a second SMS asking them to specify whether they would prefer messages in English or Swahili (language preference could be changed throughout the intervention). They then received their first, randomly selected weekly domain, marking the start of their 7-week intervention period. The platform presented the intervention arm participants with an SMS menu of various subtopics from the weekly domain (Figure 2), and the intervention participants were asked to select a number to learn more about the selected subdomain.

Interaction with the platform incurred no charges from the participants, and all costs were billed to the study. Intervention arms participants were prompted to engage with weekly quizzes, after which free airtime of US $ 0.50 was credited to their phones by the study irrespective of whether their responses were correct or not. Per ethics requirements, participants were periodically reminded that they could opt out of the study by sending the word *STOP* to the short code, given as 21438. All interactions with the system were free for participants, with any SMS costs reverse billed to the study.

Participants could request a given subdomain using a number-based menu (Figure 3)—each subdomain request resulted in 1 to 3 SMS worth of content pushed to young people's mobile phones. The number-based menu was selected, as it is commonly used by mobile phone users in Kenya to purchase mobile airtime and/or internet bundles directly from their service provider and for remote mobile money transfers. Finally, weekly domains were sent with an opt-out option where participants could unsubscribe from the platform at will.

Figure 2. Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome intervention arm architecture showing domains and subdomains. ARMADILLO: Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome. Pregnancy Prevention. GBV: gender-based violence; IUD: intrauterine device; E-pills: emergency contraceptive pills; OC: oral contraceptives; PMTCT: prevention of mother to child transmission of HIV.

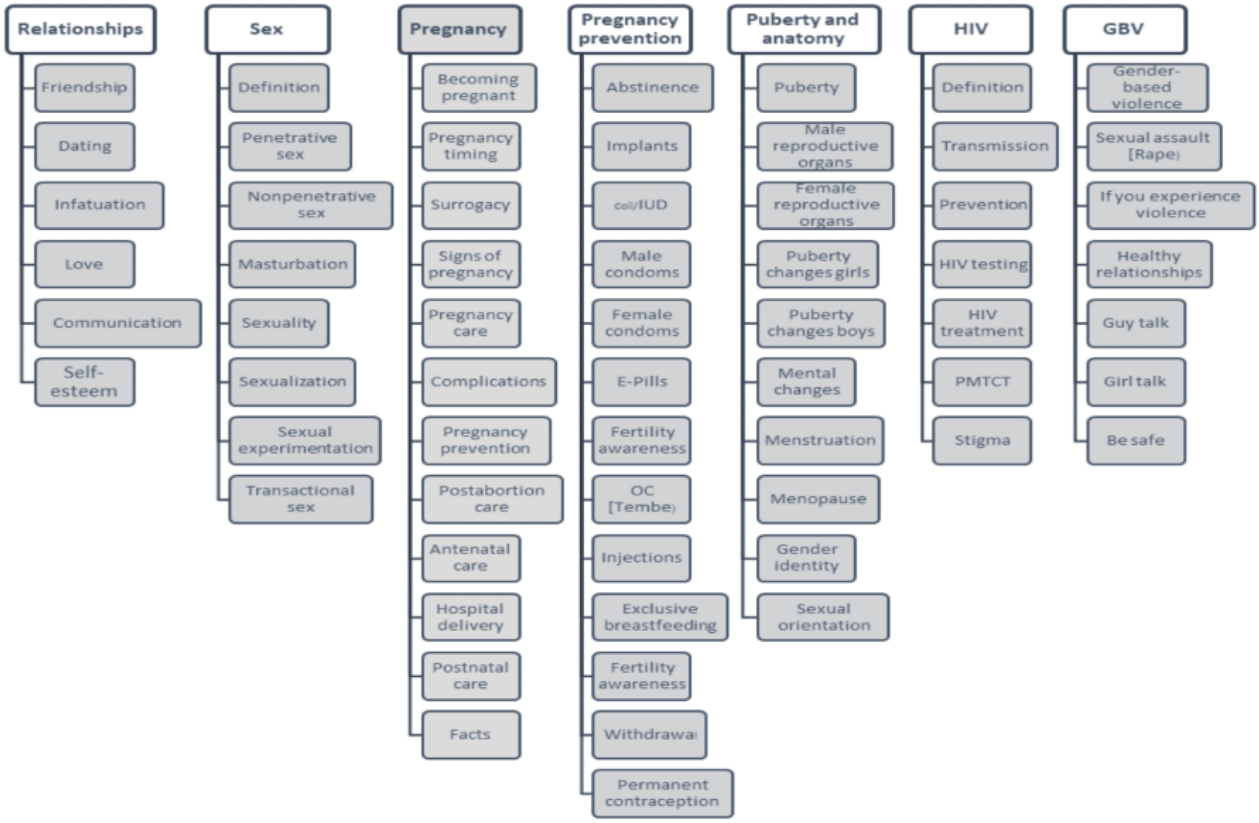
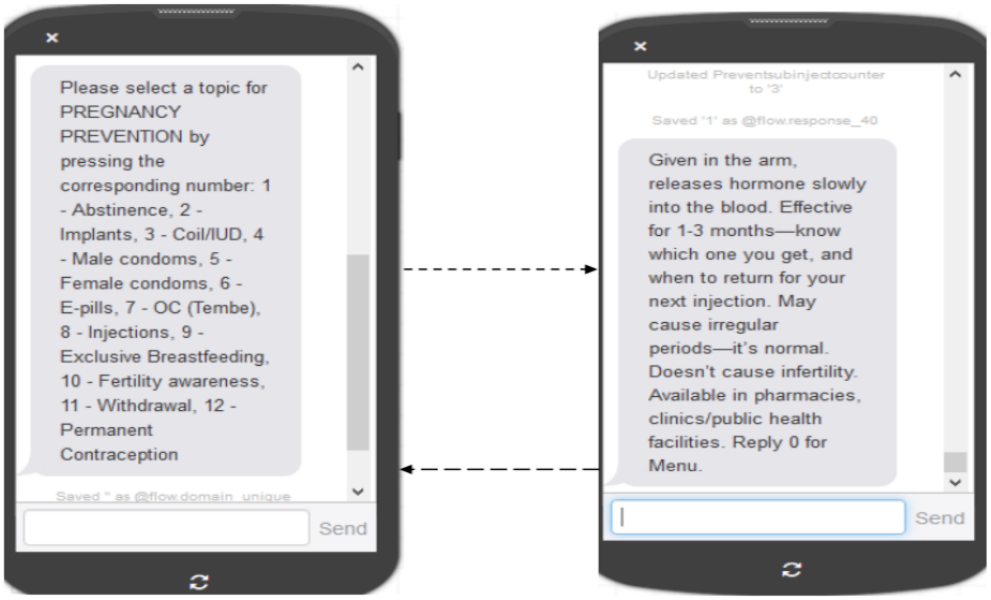


Figure 3. Example Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome subdomain menu with sample 'injection (#8) message requested' (English version).



Data Collection

Qualitative data related to the ARMADILLO platform were collected at the end of the 7-week intervention period of the study. The in-depth interviews (IDIs) were held between April 2018 and June 2018. Out of 206 intervention arm participants who did not drop out of the study and completed both baseline and end line assessments, 30 intervention arm participants (15 females and 15 males) were randomly selected from a list of intervention arm participants to participate in an IDI. The concept of saturation in qualitative methods limiting the number of IDIs to between 25 and 30 participants required to reach

thematic saturation [22] was employed. The IDIs sought to elicit participants' experiences, opinions, and perspectives on the design and content of the ARMADILLO platform. Following randomization, the study team made phone calls to the selected intervention arm participants to check on their availability. If one was not available, another participant was randomly selected until a predetermined sample size of 30 intervention respondents was achieved.

A total of 4 interviewers (2 men and 2 women) were trained to conduct the interviews and were provided with a semistructured interview guide. All interviewers had university-level education, 2 were public health experts, whereas the other 2 were social scientists. The interview guides were developed by the research team; the interview questions determined by the research team were reviewed by experts from the International Centre for Reproductive Health-Kenya. After the review of the interview questions, the guides were piloted with young people aged 18 to 24 years from a neighboring county. Suggested changes arising from piloting were reported to the research team and appropriate modifications were made. The guide content included questions on experiences while interacting with the platform; the typical question was "Tell me about your experience interacting with the platform, how was it? Easy? Difficult?" Other questions included what participants liked and/or disliked about the ARMADILLO platform and participants' views about the ARMADILLO domains, including their relevance and comprehensibility as well as suggestions for improvements.

All interviews were conducted in person at a private room in a local drop-in center (a nongovernmental health facility where young people and key populations receive comprehensive package of SRH services). Each data collector interviewed participants of the same sex. All interviews were conducted in Swahili, and following the consent of the respondents, all discussions were audio recorded.

The interviews were later transcribed verbatim, translated into English, and coded and analyzed thematically using NVivo (QSR International) version 12 software. Two researchers read through all the 30 transcripts independently; repeated readings of all the transcripts to search for meanings and patterns were done to further familiarize ourselves with the data. This was followed by code development from the data where the researchers identified important statements from the transcripts and attached a label to them to guide in developing themes. Four research meetings were held throughout the coding process to discuss emerging ideas arising from the coded data sets. These meetings helped in theme development. Our code development and thematic analysis was data driven, and we used inductive coding to create codes based on the responses of the study participants. Our analysis was shaped by the frequency of similar codes and recurrent ideas based on similarities in the data. Following theme development, the researchers reviewed the coded data extracts from each theme for coherence and consistency until all the researchers were satisfied that the data have been logically presented in a useful manner. A framework thematic analysis approach [23] was used to develop the themes. The definition of *usability of digital health interventions* by the WHO informed the coding process [4]. Our indicators of analysis, therefore, sought to determine whether the mHealth intervention worked, in that, it was easy to interact with (user friendly), and whether young people found the platform useful. Our analysis and findings are presented in accordance with the Consolidated Criteria for Reporting Qualitative Research [24].

Ethical clearance was obtained from the WHO Institutional Review Board (Protocol WHO A65892 core) and the Kenyatta National Hospital-University of Nairobi Ethics and Review Committee (KNH-UoN-ERC P550/09/2014).

Results

Demographic Characteristics and Phone Ownership of Study Participants

An equal number of men and women (15 each) were interviewed in this study. The mean age for all participants was 21.13 years, with an SD of 1.96. A greater proportion of women 33% (5/15) compared with men 27% (4/15) had postsecondary education (Table 1). More men 40% (6/15) had secondary school education as their highest education level compared with their women counterparts (4/15, 27%). More women 87% (13/15) than men (8/15, 53%) owned smartphones. This study also established that more than half of the study participants had their current phones for more than a year men, 53% (8/15) and women, 80% (12/15), as shown in Table 1.

Qualitative findings from this study are categorized into following 3 themes:

1. High and low points of an on-demand system
2. Navigating the system's supportive features: what worked and what didn't?
3. SRH content relevance and comprehension: Was it useful?

Table 1. Sociodemographic characteristics of study participants by sex (N=30).

Characteristic	Youth, men			Youth, women		
	Age 18-19 years (n=3, 20%)	Age 20-24 years (n=12, 80%)	Total (n=15, 100%)	Age 18-19 years (n=4, 27%)	Age 20-24 years (n=11, 73%)	Total (n=15, 100%)
Highest school level, n (%)						
Never attended school	N/A[a]	1 (8.3)	1 (6.6)	1 (25.0)	N/A	1 (6.7)
Primary school	2 (66.7)	2 (16.7)	4 (26.7)	2 (50.0)	3 (27.3)	5 (33.3)
Secondary school	N/A	6 (50.0)	6 (40.0)	1 (25.0)	3 (27.3)	4 (26.7)
Postsecondary school	1 (33.3)	3 (25.0)	4 (26.7)	N/A	5 (45.4)	5 (33.3)
Relationship status, n (%)						
Single	2 (66.7)	6 (50.0)	8 (53.3)	2 (50.0)	4 (36.4)	6 (40.0)
Dating or friends with benefits	1 (33.3)	5 (41.7)	6 (40.0)	N/A	7 (63.6)	7 (46.7)
Married or engaged	N/A	1 (8.3)	1 (6.7)	2 (50.0)	N/A	2 (13.3)
Have a child, n (%)						
Yes	N/A	2 (16.7)	2 (13.3)	N/A	N/A	N/A
No	3 (100)	10 (83.3)	13 (86.7)	4 (100)	11 (100)	15 (100)
Phone type, n (%)						
Analog phone (only for calls and text messages)	N/A	3 (25.0)	3 (20.0)	N/A	1 (9.1)	1 (6.7)
Multimedia phone (has a camera and MP3[b] player but does not have the internet or apps)	1 (33.3)	3 (25.0)	4 (26.7)	1 (25.0)	N/A	1 (6.7)
Smartphone (has internet and apps)	2 (66.7)	6 (50.0)	8 (53.3)	3 (75.0)	10 (90.9)	13 (86.6)
Duration of phone ownership, n (%)						
Less than 1 month	N/A	1 (8.3)	1 (6.7)	1 (25.0)	N/A	1 (6.7)
1-3 months	1 (33.3)	N/A	1 (6.7)	N/A	N/A	N/A
3-6 months	N/A	4 (33.3)	4 (26.7)	N/A	N/A	N/A
6-12 months	1 (33.3)	N/A	1 (6.7)	1 (25.0)	1 (9.1)	2 (13.3)
More than one year	1 (33.3)	7 (58.4)	8 (53.3)	2 (50.0)	10 (90.9)	12 (80.0)

[a]N/A: not applicable.

[b]MP3: MPEG audio layer-3.

High and Low Points of an On-Demand System

On-demand systems are common in Kenya. When asked about their experiences interacting with the ARMADILLO platform, a majority of the IDI participants reported that the platform's procedures were easy to follow and that they were able to navigate through the platform on their own. Both female and male participants reported positive user experiences and were generally able to move through the steps without any difficulties. The similarity of the mHealth platform with other on-demand mobile services offered by the majority of the mobile service providers in Kenya to buy airtime, pay bills, and order for other services was largely associated with the ease of platform use:

> *I can say it was easy. ... I reached the messages by following the instructions that were given, they were simple instructions to follow then get to a conclusion* [23-year-old male youth]

> *When I received the messages, I used to follow the directions like they were sending let's say from number 1-12, so I went through them one by one ... I used to send one after the other, when I send the first one, they bring me the second one until I finished. And they used to respond to what I was asking them* [21-year-old female youth]

However, the ease of use was not entirely uniform. Reported restrictions included repeat instructions to participants, preventing them from progressing to other stages. A few users reported difficulties completing some functions, such as getting back to the menu by selecting *0*. This was because some participants reported that they were used to the feature (reply *0* for menu) being called *home* in other apps as opposed to menu, whereas others preferred it to be called the *go back* feature. In addition, a few participants mistakenly put a number not available on the menu and kept getting an *I don't understand*

your selection response. When designing this mHealth platform, such challenges did not seem straightforward, probably based on the assumption that the navigation process was clear and that young people were adept at using mobile phones:

> *I experienced difficulties two times since every time I responded I could see those are the same questions being asked and when I reply am told they do not understand my response, until the third time that is when I entered one option and I realized that they were topics, I did not know and understand the first time.* [19-year-old female youth]

That said, generally, when participants were unclear, they sought assistance from their peers in the study. Despite being informed that they could reach out to the study team in case of any technical challenges, young people in this study tended to be far more comfortable seeking help from their peers than from the research team. This speaks to the component of peer-to-peer learning among young people, typical of how they rely on their peers for advice. Although no female participants reported reaching out to their peers for technical support, male participants in this study reported some challenges in progressing through the system and preferred to reach out to their friends:

> *First, I did not know how to go about it, but I got advice from someone else who instructed me and told me how it was and how to use it.* [22-year-old male youth]

> *I inquired from a friend, I did not know how to go about it and there was someone called XXX or I am not supposed to mention him ... I followed him and he explained to me.* [20-year-old male youth]

A feature of on-demand systems is the ability to receive targeted information on a topic based on a user's interest *in the moment.* Respondents appreciated the *immediate response* they received from the system when requesting a message from a subdomain. Participants did not indicate any long breaks or delays that would have discouraged them from interacting with the platform. Almost all participants interviewed provided this positive comment regarding the ARMADILLO platform. The capacity for real-time feedback resulted in some motivation by most users, which made them continue to interact with the system. Perhaps this is an interesting observation. The impatient nature of young people was clearly displayed; mHealth platforms, therefore, need to provide real-time support to increase their usability when targeting young people:

> *... when you request for something there was immediate feedback ... and it's like there is somebody attending to you, there were no delays.* [18-year-old male youth]

> *What I liked most about ARMADILLO was the information they shared, it was so helpful and there were no delays. Anyone who wanted to get the whole information would press a number and one would get it instantly.* [23-year-old male youth]

A limitation of on-demand systems is that, similar to push systems, there is no ability to engage further for questions and clarifications beyond what the developers have predefined. This *static* nature caused some frustration among participants, who wanted additional clarity when they came across difficult terms or had additional questions about the information received. This concern came from both male and female participants; young people wanted to learn more about SRH issues, particularly on pregnancy prevention methods. This finding suggests that a platform allowing young people to ask SRH questions via text messages and being responded to by trained SRH experts would appeal more to young people and contribute to improved mHealth interactions:

> *I do not know about the others if they received, or maybe they had people to ask who understood but as for me I had no one to ask, I would give an answer to get the credit but ... I just want to request if one was wrong then they should be able to direct them into getting the right answer, because one does not know what the right or wrong answer is since whether right or wrong one still received credit* [21-year-old female youth]

> *There was a time I received some information about something, but I was not contented because I did not understand ... there was an issue about family planning, using of pills and injection ... There are some that I understood, and there are others that I needed more explanation, not everything that one reads they get to understand.* [22-year-old male youth]

Navigating the System's Supportive Features: What Worked and What Didn't?

The weekly, remunerated quizzes (initially developed to encourage participants to engage with the system and motivate them to participate) were well received by study participants. Some participants were explicit about being motivated by the free airtime to interact with the system. However, others felt that the weekly quizzes were a good feature that positively challenged their knowledge. The majority of participants expressed their excitement with the weekly quizzes terming them as learning motivators that interested and made them want to engage more with the platform:

> *I liked the questions part because they promote you with credit at the end of answering and also, they give you a duration if you are free, they ask you questions then after answering they send you credit* [23-year-old female youth]

> *What I liked was when I was asked the weekly questions, some questions I did not know about them, while others I did; so, I used to think if I was able to answer them, then am good ... That challenge was what I liked.* [21-year-old female youth]

However, 2 additional design features were a source of confusion for the participants. First, a number of participants reported that they ignored the first messages, as they did not recognize the ARMADILLO short code (21438) and thought that the messages were spam, sent every so often by local service providers. Although short codes are the preferred channels for mHealth bulk SMS text messaging, the skepticism tendency

displayed by young people implies that the proliferation of SMS services and spam messages could threaten the mHealth platform's acceptability and usability, particularly by young people who appeared cautious when interacting with unverified sources. For this study, had the messages been sent with an *ARMADILLO* name, chances are that participants would not have ignored their first messages:

> *When I received the message I did not know, any time I would receive the messages and think they are from ... service provider ...and ignored them only to realize they were from ARMADILLO.* [24-year-old male youth]

> *I was shocked ... I have never gotten such questions or messages, ... I was like who are these people? Of all the people why did they choose me... as in to ask me because ... I did not know they [people sending the messages] were the ones like the same with ARMADILLO.* [18-year-old female youth]

Second, participants were periodically reminded that they could opt out of the system (and study) by sending *STOP* back to the short code. Unfortunately, some participants thought that this opt-out option was similar to *periodic* unsubscribing from mobile provider services, which allow users to opt out (eg, when not interested in something) but also opt *in* later on. However, to comply with research ethics requirements, opting out meant dropping out of the study altogether:

> *For example, if I wanted to dismiss myself from the group it was a bit of a challenge and I would keep asking myself... for example if I dismiss myself from the system, I did not know whether I would continue getting the messages later on or not, so I did not know how to go about that.* [22-year-old female youth]

SRH Content Relevance and Comprehension: Was It Useful?

Participants generally felt that the ARMADILLO contents were relevant to them and people of their age. Messages were described as being educative and informative. Not only did the participants receive new information but they were also conveyed the information with clarity, devoid of the shame mostly associated with SRH talk. SRH knowledge, candidly described, resonated with the young participants of ARMADILLO. Topics touching on sex, HIV, and pregnancy were the most mentioned domains from the platform by both male and female participants:

> *...they have helped me to overcome some myths, be open minded, you know...as in .... There are myths like, for instance that young girls cannot get pregnant or get AIDs ... There are also those that say when one engages in sex for the first time, she cannot get pregnant. Something of that sort makes someone wise; you can understand that it is important to use protection when engaging in sex* [21-year-old male youth]

> *What I liked most is that most issues discussed were issues that were not openly discussed as they are considered shameful, in fact in this generation it has been difficult for us to sit with our parents and be explained to that 'nowadays you are supposed to be this and that way'... when I came across ARMADILLO, I got the opportunity that I have been missing ... I have even known how to avoid certain things.* [23-year-old male youth]

Anecdotally, some messages were purported to have inspired intent to change behavior. Participants reported going for HIV testing and deciding to use condoms as a result of the messages:

> *The other ones like HIV testing...as in there is a time I had stayed long without testing, When I read those messages, I got the courage to go and get tested because it had been long.* [21-year-old female youth]

> *... to prevent myself, as in health wise, to prevent myself by using a condom that is what made me happy, I read about how I can help myself when I meet a woman or how to ensure she does not get pregnant.* [20-year-old male youth]

Although a majority of the participants reported satisfaction with the platform's use of language (in both English and Swahili), a few indicated that certain concepts and words used were difficult to understand. The language challenge was largely reported by the female participants. Unlike their male counterparts who reported reaching out to their friends when they experienced challenges navigating the platform, the female participants chose to use the web in search of explanations, whereas others chose to keep their uncertainties to themselves:

> *At times the English used was hard and it became complicated, but I would use google for the meaning if I did not understand, because their English was deeper, so I just 'googled' the meaning and I go through it.* [18-year-old female youth]

> *Difficult was the family planning [apart from condoms and emergency pills], the subdomains about family planning [pregnancy prevention]; even though I requested, [when] they gave me the messages I was like ... I am not understanding anything.* [21-year-old female youth]

Discussion

Principal Findings

In this qualitative study, we describe young peoples' experiences using an on-demand mHealth intervention and thoughts about the content. Young participants in the ARMADILLO trial had a variety of experiences when interacting with the mHealth platform. Once a weekly domain was opened, the majority of participants reported enjoying the freedom to select the subdomain they wished to read at their convenience. However, although designed for intuitive engagement, the ease of use as envisioned by the study was not uniform. A few users reported getting confused by the aspects of the system. Others expressed frustration when they reached the end of a message and were left with additional questions. Weekly quizzes aimed at motivating the study participants to interact with the system were appreciated. The guaranteed mobile credit was probably responsible for a fair amount of participants' enthusiasm.

Misunderstandings about the system's *stop* feature and short code caused additional confusion. Finally, the content itself was widely appreciated and understood by most participants; a few, however, cited comprehension difficulty.

mHealth Situation in Kenya

Most of the myriad mHealth projects in Kenya (both on-demand and pushed services) are rarely evaluated to assess their usability by the targeted audience [17]. One exception is an evaluation of the m4RH program, which was designed to expand the access to family planning information for the general public with an on-demand menu. The m4RH usability evaluation assessed acceptability, feasibility, and potential behavioral impact [18]. The success of the project was evaluated based on the number of users interacting with the m4RH system, how users learned about the platform, and users reporting satisfaction with the program and its contents [18,19]. Similar to this ARMADILLO assessment, the m4RH assessment demonstrated the potential of reaching young people with factual and timely SRH information, including contraceptive facts. In addition, the (m4RH) assessment established that adolescents and young adults were the most frequent users of the system [18], with users appreciating the simplicity of the language used and the privacy that characterized the delivery of SRH information [19]. More assessments are needed to evaluate young people's perceptions of the ease of use and acceptability of mHealth systems and their contents.

mHealth Future Considerations

When a majority of participants successfully engaged with the ARMADILLO intervention, why choose to focus this paper on the usability pain points? The success and adoption of mHealth interventions largely depends on the targeted end users' interaction with the technology and their belief that using the platform will benefit their health [4]. With regard to content development, the ARMADILLO study had a robust formative phase [25] that involved vetting SRH content with young people for comprehension, relevance, and tone. That said, during its implementation, a small minority of participants indicated that they struggled to understand the content. This speaks to the importance of understanding the full spectrum of health literacy among the targeted population, particularly when dealing with a population as heterogeneous as young people.

Issues related to the construction of the system itself (the *stop* feature, any confusion about the short code and progressing through the system) might possibly have been identified by conducting additional pre-RCT *real-world* pretest of the system. Others, including sustained frustration at not being able to obtain additional information, likely did not affect the objectives of the broader RCT but had implications for implementing similar services. One preferred feature of the ARMADILLO intervention was the instant gratification of immediately receiving a targeted message when requested. This is the natural strength of an on-demand system where the desired information comes when queried.

Sentiments about the desire for real-time feedback were observed with another digital study in Kenya [26]. This can happen in a few ways. One option is to have persons (peers or health experts) providing real-time responses to questions, as has been done in Mozambique's Geracao Biz program [27]—this can, however, be resource intensive. Rapid improvements in artificial intelligence and its application to chatbots mean that users may soon be able to receive more *human* responses to real-time queries. Springster's Big Sis is an interesting early example [25], although evaluation is needed. A less expensive option is to link users of on-demand or push systems to platforms with more flexibility and space for content (eg, websites) and hotlines. Integrating with already-existing systems avoids needless duplication. Creating robust and interoperable platforms ensures the success of mHealth initiatives in readiness for scale-up [28].

Limitations

This study was not without limitations, one being that intervention participants not interviewed in this study might have different perspectives on the ARMADILLO platform compared with those reported in this paper. However, the random selection of intervention arm study participants and participants' diversities in terms of age and gender strengthen our recruitment criteria by avoiding a biased selection. Second, respondents in this study might have given socially desirable responses by either overreporting the positives or underreporting of undesirable platform features. The consistency of our findings with other mHealth assessment studies supports the validity of this study.

Conclusions

Findings from this study add evidence to the underexplored area of mHealth users' experiences targeting young people. The need to consider end users throughout the mHealth development for optimum functionality to be achieved is key. This will not only create a sense of ownership but also give the mHealth initiative a practical approach while promoting its adoption through various media appealing to young people. With continued attention to digital health broadly and an entire new generation of delivery channels (from influencers to chatbots), program developers need to pay careful attention to understanding how users interact with a system, rather than make assumptions about *what works* when engaging with young people.

Acknowledgments

The authors acknowledge Winnie Wangari for her work as a research assistant during the study period. The authors appreciate Nelly Marete for her inputs during the writing of the manuscript. This work was supported by the UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), a cosponsored programme executed by the World Health Organization (WHO). The named authors alone are responsible for the views expressed in this publication and do not necessarily represent the decisions or the policies of the

UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP) or the World Health Organization (WHO).

Authors' Contributions

MJ, GL, and GP conceptualized the study. MJ, GL, GP, and TM oversaw data collection. MJ participated in data analysis, and MJ, GL, and GP drafted the manuscript. TM, MC, WM, SH, OA, GP, MJ, and GL provided critical input in the development and revision of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Ippoliti NB, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. Reprod Health 2017 Jan 17;14(1):11 [FREE Full text] [doi: 10.1186/s12978-016-0276-z] [Medline: 28095855]
2. Zhang D, Lai J. Can convenience and effectiveness converge in mobile web? A critique of the state-of-the-art adaptation techniques for web navigation on mobile handheld devices. Int J Hum Comput Interact 2011:- [FREE Full text] [doi: 10.1080/10447318.2011.559876]
3. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, et al. The usability and acceptability of an adolescent mhealth HIV/STI and drug abuse preventive intervention in primary care. Behav Med 2016 May 25:1-12. [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]
4. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: World Health Organization; 2016.
5. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190. [doi: 10.2196/mhealth.8758]
6. Oerther SE, Manjrekar P, Oerther DB. Utilizing mobile health technology at the bottom of the pyramid. Procedia Engineering 2014;78:143-148. [doi: 10.1016/j.proeng.2014.07.050]
7. Wesolowski A, Eagle N, Noor A, Snow R, Buckee C. Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS One 2012:- [FREE Full text] [doi: 10.1371/journal.pone.0035319]
8. Varshney U. Mobile health: four emerging themes of research. Decis Support Syst 2014 Oct;66:20-35. [doi: 10.1016/j.dss.2014.06.001]
9. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
10. Classification of Digital Health Interventions. WHO. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf;jsessionid=1FCDEAAD8CE7584B36A24E45F3616928?sequence=1 [accessed 2020-12-17]
11. Aamir J, Ali SM, Kamel Boulos MN, Anjum N, Ishaq M. Enablers and inhibitors: a review of the situation regarding mHealth adoption in low- and middle-income countries. Health Policy Technol 2018 Mar;7(1):88-97. [doi: 10.1016/j.hlpt.2017.11.005]
12. McClure J, Anderson M, Bradley K, An L, Catz S. Evaluating an adaptive and interactive mhealth smoking cessation and medication adherence program: a randomized pilot feasibility study. JMIR Mhealth Uhealth 2016 Aug 3;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]
13. Kryger M, Crytzer T, Fairman A, Quinby E, Karavolis M, Pramana G, et al. The effect of the interactive mobile health and rehabilitation system on health and psychosocial outcomes in spinal cord injury: randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e14305 [FREE Full text] [doi: 10.2196/14305] [Medline: 31464189]
14. Chiang N, Guo M, Amico K, Atkins L, Lester R. Interactive two-way mhealth interventions for improving medication adherence: an evaluation using the behaviour change wheel framework. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e87 [FREE Full text] [doi: 10.2196/mhealth.9187] [Medline: 29650504]
15. Lester R, Ritvo P, Mills E, Kariri A, Karanja S, Chung M, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov;376(9755):1838-1845 [FREE Full text] [doi: 10.1016/s0140-6736(10)61997-6]
16. Kosgei B. The Role of Mobile Technology in Kenya Healthcare. i3 Consult. URL: https://www.i3consult.com/the-role-of-mobile-technology-in-kenya-healthcare/ [accessed 2020-12-19]
17. Njoroge M, Zurovac D, Ogara E, Chuma J, Kirigia D. Assessing the feasibility of eHealth and mHealth: a systematic review and analysis of initiatives implemented in Kenya. BMC Res Notes 2017 Feb 10;10(1):90 [FREE Full text] [doi: 10.1186/s13104-017-2416-0] [Medline: 28183341]

18. Mobile for Reproductive Health Project: Kenya. 2013. URL: https://www.fhi360.org/sites/default/files/media/documents/m4rh-kenya-brief.pdf [accessed 2020-12-17]
19. Vahdat H, L'Engle KL, Plourde K, Magaria L, Olawo A. There are some questions you may not ask in a clinic: providing contraception information to young people in Kenya using SMS. Int J Gynaecol Obstet 2013 Nov;123(Suppl 1):e2-e6 [FREE Full text] [doi: 10.1016/j.ijgo.2013.07.009] [Medline: 24012514]
20. 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]
21. Kenya Demographic and Health Survey 2014. Kenya National Bureau of Statistics. 2015. URL: https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf [accessed 2020-12-17]
22. Dworkin S. Sample size policy for qualitative studies using in-depth interviews. Arch Sex Behav 2012 Dec;41(6):1319-1320 [FREE Full text] [doi: 10.1007/s10508-012-0016-6] [Medline: 22968493]
23. Nowell L, Norris J, White D, Moules N. Thematic analysis. Int J Qual Method 2017 Oct 2;16(1):160940691773384 [FREE Full text] [doi: 10.1177/1609406917733847]
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Gonsalves L, L'Engle KL, Tamrat T, Plourde K, Mangone E, 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 7;12:67 [FREE Full text] [doi: 10.1186/s12978-015-0059-y] [Medline: 26248769]
26. Ivanova O, Wambua S, Mwaisaka J, Bossier T, Thiongo M, Michielsen K, et al. Evaluation of the elimika pilot project: improving art adherence among HIV positive youth using an ehealth intervention in Mombasa, Kenya. Afr J Reprod Health 2019 Mar;23(1):100-110. [doi: 10.29063/ajrh2019/v23i1.10] [Medline: 31034176]
27. Chandra-Mouli V, Gibbs S, Badiani R, Quinhas F, Svanemyr J. Programa Geração Biz, Mozambique: how did this adolescent health initiative grow from a pilot to a national programme, and what did it achieve? Reprod Health 2015 Feb 17;12:12 [FREE Full text] [doi: 10.1186/1742-4755-12-12] [Medline: 25971669]
28. Tomlinson M, Rotheram-Borus M, Swartz L, Tsai A. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]

Abbreviations

ARMADILLO: Adolescent/Youth Reproductive Mobile Access and Delivery Initiatives for Love and Life Outcome
CPR: contraceptive prevalence rate
IDI: in-depth interview
mHealth: mobile health
m4RH: Mobile for Reproductive Health
RCT: randomized controlled trial
SRH: sexual and reproductive health
WHO: World Health Organization

Edited by G Eysenbach; submitted 03.04.20; peer-reviewed by M Bangha, C Mudogo, S Hewage, E Burner; comments to author 12.06.20; revised version received 20.08.20; accepted 14.09.20; published 15.01.21.

Please cite as:
Mwaisaka J, Gonsalves L, Thiongo M, Waithaka M, Sidha H, Alfred O, Mukiira C, Gichangi P
Young People's Experiences Using an On-Demand Mobile Health Sexual and Reproductive Health Text Message Intervention in Kenya: Qualitative Study
JMIR Mhealth Uhealth 2021;9(1):e19109
URL: http://mhealth.jmir.org/2021/1/e19109/
doi:10.2196/19109
PMID:33448930

Original Paper

Mobile App Support for Cardiopulmonary Resuscitation: Development and Usability Study

Sune Dueholm Müller[1], PhD; Kasper Glerup Lauridsen[2,3,4], MD, MSc; Amra Hadrovic Palic[1], MSc; Lotte Nygaard Frederiksen[1], MSc; Morten Mathiasen[5], MSc; Bo Løfgren[2,3,6], MD, PhD, FESC, FAHA

[1]Department of Management, Aarhus University, Aarhus, Denmark

[2]Department of Medicine, Randers Regional Hospital, Randers, Denmark

[3]Research Center for Emergency Medicine, Aarhus University Hospital, Aarhus, Denmark

[4]Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[5]Research and innovation, Business Academy Aarhus, Viby, Denmark

[6]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Corresponding Author:
Sune Dueholm Müller, PhD
Department of Management
Aarhus University
Fuglesangs Allé 4
Aarhus, 8210
Denmark
Phone: 45 87165023
Email: sdm@mgmt.au.dk

Abstract

Background: The user requirements for in-hospital cardiopulmonary resuscitation (CPR) support apps are understudied. To study usability, functionality, and design based on user requirements, we applied a mixed methods research design using interviews, observations, and a Kano questionnaire to survey perspectives of both physicians and nurses.

Objective: This study aims to identify what an in-hospital CPR support app should include to meet the requirements and expectations of health care professionals by evaluating the *CprPrototype* app.

Methods: We used a mixed methods research design. The qualitative methods consisted of semistructured interviews and observations from an advanced life support (ALS) course; both provided input to the subsequent questionnaire development. The quantitative method is a questionnaire based on the Kano model classifying user requirements as *must-be*, *one-dimensional* (attributes causing satisfaction when present and dissatisfaction when absent), *attractive*, *indifferent*, and *reverse* (attributes causing dissatisfaction when present and satisfaction when absent). The questionnaire was supplemented with comment fields. All respondents were physicians and nurses providing ALS at hospitals in the Central Denmark Region.

Results: A total of 83 physicians and nurses responded to the questionnaire, 15 physicians and nurses were observed during ALS training, and 5 physicians were interviewed. On the basis of the Kano questionnaire, 53% (9/17) of requirements were classified as *indifferent*, 29% (5/17) as *attractive*, and 18% (3/17) as *one-dimensional*. The comments revealed 7 different categories of user requirements with noticeable differences between those of physicians and nurses: *technological challenges*, *keep track of time*, *documentation and history*, *disturbing element*, *improvement areas: functions*, *improvement areas: design*, and *better guidance*.

Conclusions: The study provides recommendations to developers on the user requirements that need to be addressed when developing CPR support apps. Three features (*one-dimensional* attributes) must be incorporated in an in-hospital CPR support app: *reminder of rhythm check*, *reminder of resuscitation drugs*, and *differentiate between adults and children*. In addition, 5 features (*attractive* attributes) would result in higher user satisfaction: *all functions on one side*, *access to the patient journal in the app*, *automatic time recording when cardiac arrest is called*, *sound to guide the chest compression rate* (metronome), and *send CPR history to the DANARREST(Danish in-hospital cardiac arrest registry) database*.

(JMIR Mhealth Uhealth 2021;9(1):e16114) doi:10.2196/16114

KEYWORDS

the Kano model; cardiopulmonary resuscitation; healthcare; smartphone apps; public health; ALS CPR algorithm; app evaluation; mobile phone

Introduction

New digital technologies are developing rapidly, and health care organizations are increasingly adopting and adapting to these technologies to serve clinical needs [1,2]. To mention a few benefits, such technologies support better clinical decision making, facilitate better communication, and potentially improve patient outcomes [1,3]. Among the technologies that have become commonplace within health care are mobile devices, which have led to rapid growth in the development of medical software apps [1,4-6]. These include apps used for cardiopulmonary resuscitation (CPR) guidance and support [7,8]. Most available apps used to support CPR are targeting laypersons performing basic life support [7,9].

In contrast, little is known about apps for advanced life support (ALS) during in-hospital cardiac arrests. The challenges of health care professionals with in-hospital resuscitations are dealing with high cognitive load, as they have to coordinate tasks in a team and plan timely rhythm analysis and drug administrations while considering the reversible causes of cardiac arrest [10]. Failure to adhere to guidelines may adversely impact survival [11], and consequently, apps have been suggested as potential cognitive aids to improve ALS guideline adherence [12].

However, it is important to study user requirements to improve clinical usability during resuscitation [13]. Shah and Robinson [14] argue that understanding users' needs during development determines the success or failure of technology development. Martin et al [15] support this assertion by stating that investments in research on user requirements benefit not only the developer but also the user and the entire health care sector. The proper elicitation of requirements is more likely to aid in the development of technologies that will support and be used in clinical work. Therefore, research and development of an in-hospital CPR support app based on user requirements of health care professionals is a timely and relevant subject. The development of an efficient CPR support app will contribute to the improvement of the manner in which CPR will be performed in the future. To that end, this study aims to identify what an in-hospital CPR support app should include to meet the requirements and expectations of health care professionals by evaluating the *CprPrototype* app. This translates into the following research question: What are the user requirements for an app for in-hospital CPR support?

Methods

The CprPrototype App

This study seeks to elicit user requirements for an app for in-hospital CPR support by evaluating the *CprPrototype* app, developed by physicians from Aarhus University Hospital, researchers from Aarhus University, and developers from Aarhus Business Academy. The app is based on the European Resuscitation Council guideline for ALS (adapted from Soar et al [10]; Figure 1).

When performing CPR, the user of the CprPrototype app can choose the algorithm for shockable rhythms or nonshockable rhythms (screenshot 1 in Figure 2). The app then starts a 2-min cycle with a countdown for the next rhythm check. Depending on the rhythm, the app instructs the user when to prepare specific resuscitation drugs, including the dose (screenshot 2 in Figure 2). An available feature is the ability to see a list of possible reversible causes of cardiac arrest (screenshot 3 in Figure 2). The app continuously keeps track of time, and every action performed by the user in the app is tracked and stored in the app's log (*History*) feature (screenshot 4 in Figure 2).

Figure 1. Cardiopulmonary resuscitation (CPR) algorithm.

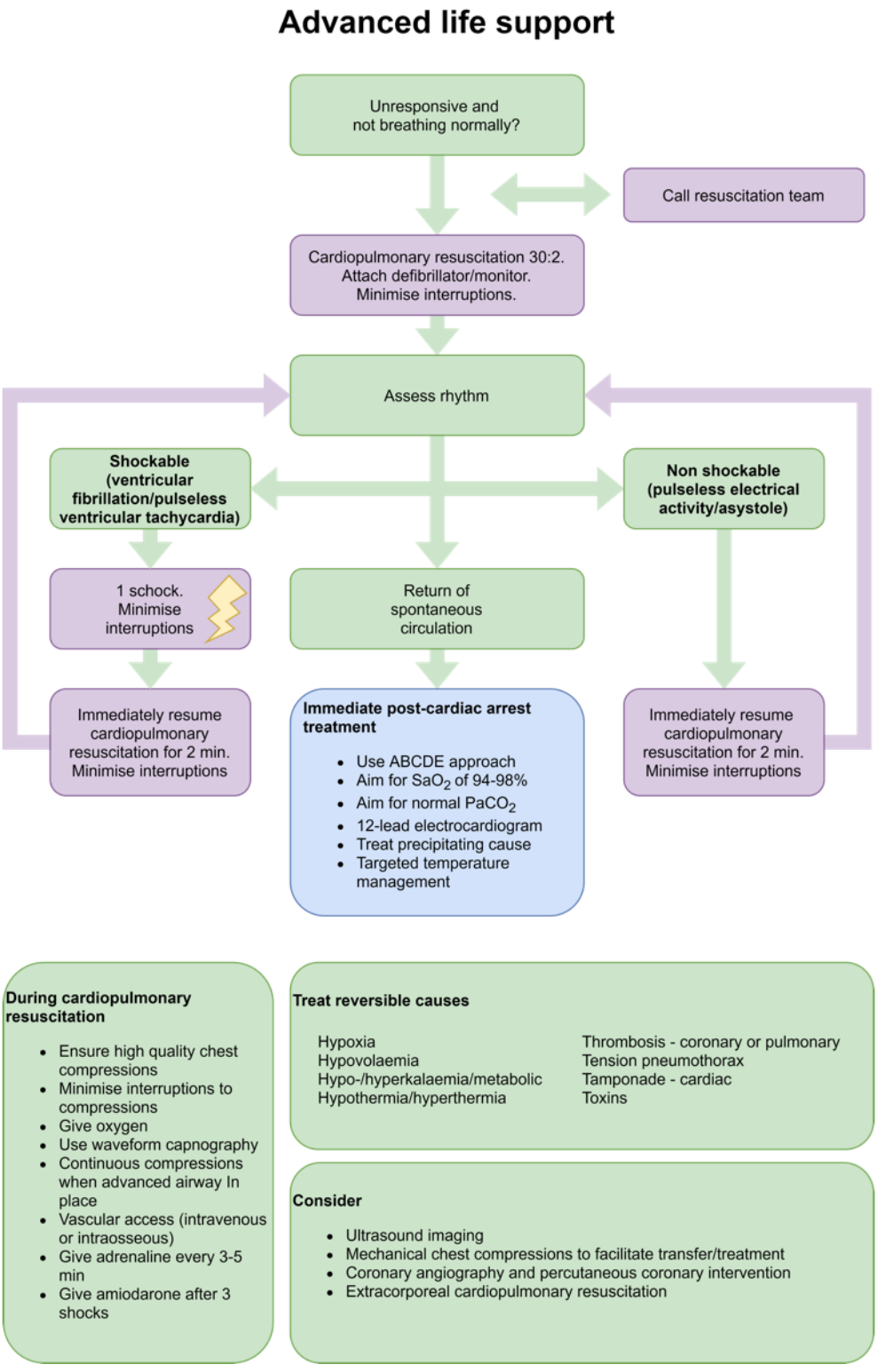
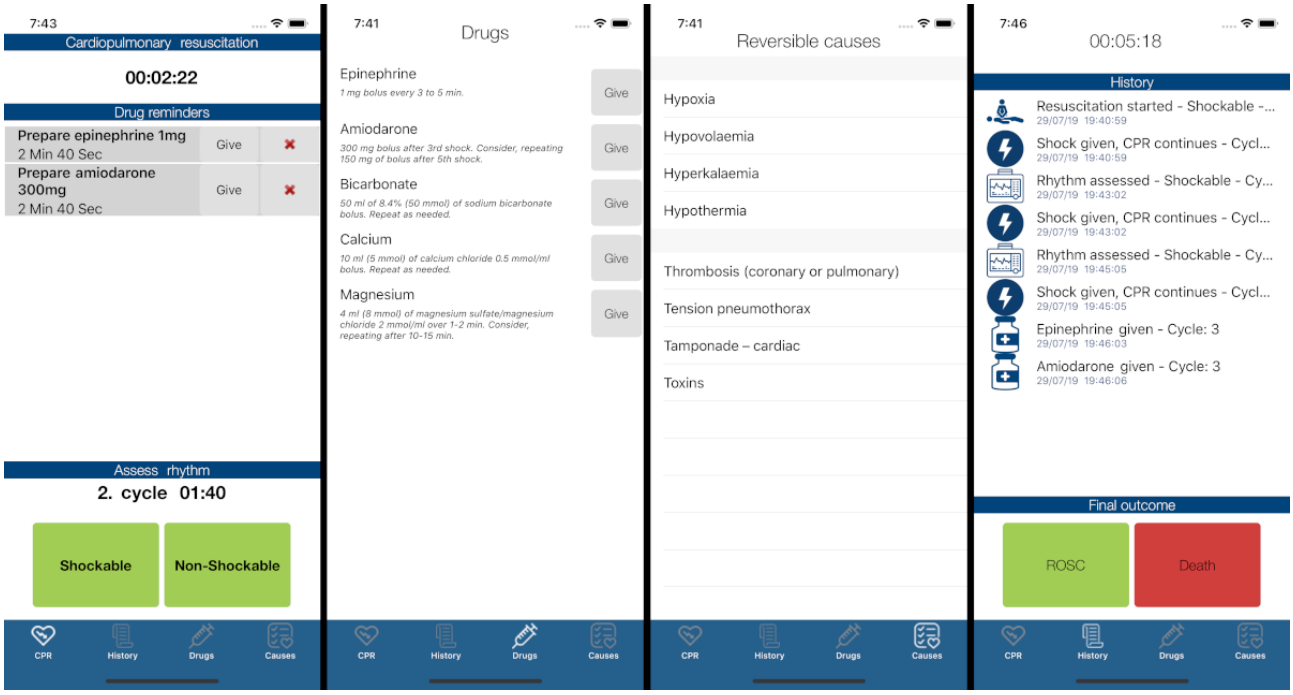


Figure 2. Screenshot of CprPrototype app. ROSC: Return of Spontaneous Circulation.



Design

As the app serves to support in-hospital CPR, it is important from a development perspective to understand the needs of nurses and physicians during resuscitation. Therefore, we applied a mixed methods research design to study user requirements.

First, we observed an ALS course to understand the unspoken needs of nurses and physicians in providing ALS. The ALS course included simulations of cardiac arrests, and the participants practiced each step of ALS. During the simulations, the nurses and physicians were divided into teams. We followed each team through participant observation in an effort to uncover the needs that were not verbally articulated by the course participants. We took extensive field notes that were subsequently coded and analyzed to identify common themes [16]. The themes (resuscitation challenges, smooth collaboration, information overload, efficient communication, and need for structure and overview) provided insight into the work processes and communication of resuscitation team members, their unspoken needs for support and guidance, and helped prepare the subsequent interviews.

Second, we conducted interviews with physicians who served as resuscitation team leaders. The interviews were semistructured, based on insights from our observations. An interview guide was used to ensure the structure and comparable answers. The interviews were transcribed and analyzed through systematic text condensation to uncover their meaning [17]. The analysis facilitated our understanding of how they practice ALS, medical terminology and technology used during resuscitation, and use case scenarios for the app.

Third, we used the interview results together with the observations to develop questions for the Kano-type questionnaire. The questionnaire was sent to the respondents with an attached video of how the app functions. Furthermore, the questionnaire data were collected and managed using the REDCap (REDCap consortium; research electronic data capture) software platform [18,19].

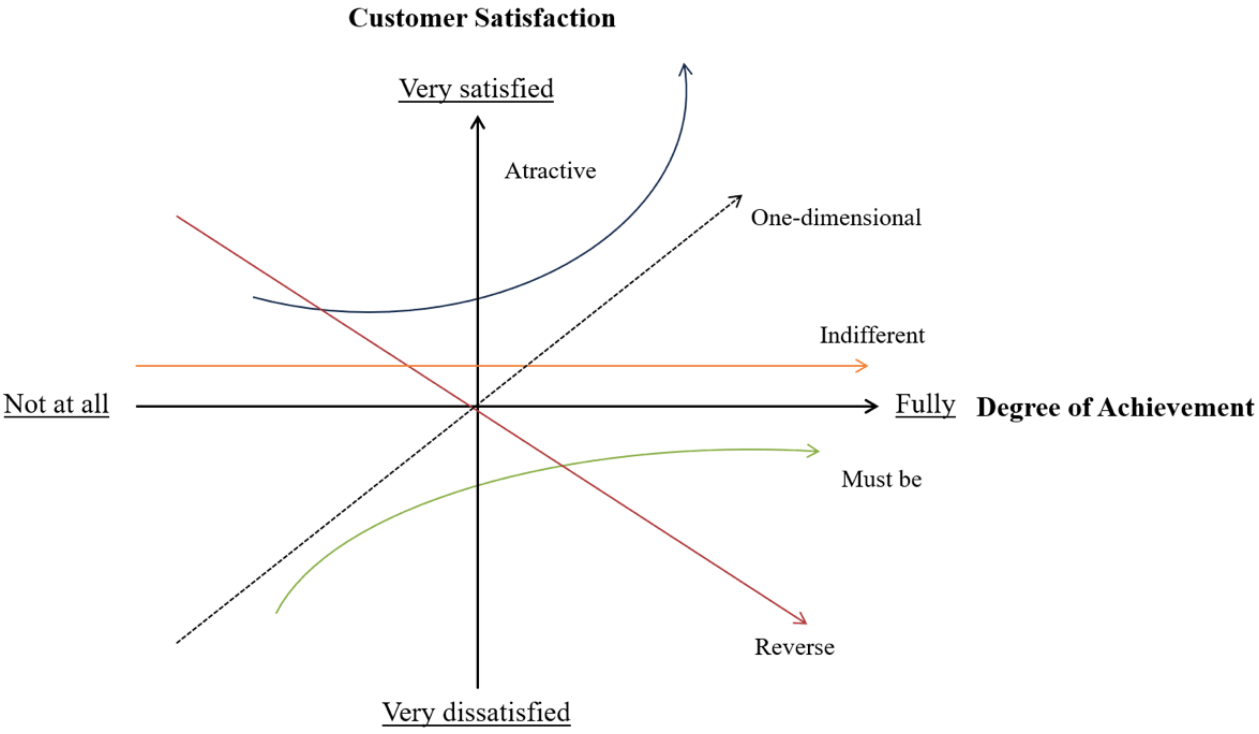
The study complies with the ethical principles for medical research involving human subjects. According to the Danish National Committee on Biomedical Research Ethics, the study does not require approval from an ethical review committee. The study took place from February to May 2019.

The Kano Model

To elicit user requirements and improve user satisfaction in developing health care products and services, the Kano model has previously been used [13,20,21]. The model provides insight into user requirements and whether different features (quality attributes) of a service or product contribute to greater or lower customer satisfaction [21]. Thus, it guides in prioritizing between user requirements and identifying opportunities when designing or improving products and services based on customer needs [13].

Kano et al [22] proposed a 2-dimensional quality model to classify and categorize an element of a service or product. The model is based on the motivator-hygiene theory by Herzberg et al [23], positing that the factors causing satisfaction are different from those causing dissatisfaction. The model helps visualize the relationship between the product's functionality and customer satisfaction (Figure 3—adapted from Witell and Löfgren [24]). The model serves to explain the role of various quality attributes in determining customer satisfaction as a basis for developing a product or service.

Figure 3. The Kano model. The horizontal axis shows the level of fulfillment of a requirement, and the vertical axis displays the level satisfaction with a requirement.



As illustrated, the model classifies the quality attributes into 5 categories: *must-be*, *one-dimensional*, *attractive*, *indifferent*, and *reverse quality* (listed based on importance). The *must-be* attributes are expected by the customer and do not result in increased customer satisfaction, but if these attributes are not present, customers are dissatisfied. The *one-dimensional* attributes cause satisfaction when present and cause dissatisfaction when absent. The *attractive* attributes are unexpected and delight the customer, which increases customer satisfaction, but they do not cause dissatisfaction when absent, because they are not expected. Therefore, these attributes often reflect unspoken needs. Finally, the *indifferent* attributes neither cause satisfaction nor dissatisfaction, and *reverse* attributes result in dissatisfaction when present and cause satisfaction when absent [24].

We used the Kano 5-level questionnaire to classify the features into 5 categories. The features are first evaluated according to a functional question (how respondents feel if a particular feature is present) and subsequently a dysfunctional question (how respondents feel if the feature is not present). The respondents had to choose between 5 possible responses (Textbox 1).

Textbox 1. Example questions and possible responses from the Kano questionnaire.

- If the app is able to differentiate between adults and children, how would you feel?
  - I like it that way
  - It must be that way
  - I am neutral
  - I can accept it to be that way
  - I dislike it that way
- If the app is not able to differentiate between adults and children, how would you feel?
  - I like it that way
  - It must be that way
  - I am neutral
  - I can accept it to be that way
  - I dislike it that way

When the survey was completed, all answers were evaluated and placed in an evaluation table [21] (Table 1) and categorized into M (must-be), O (one-dimensional), A (attractive), I (indifferent), R (reverse), and Q (questionable) quality attributes. If an answer is categorized as questionable, it indicates a conflicting answer because the respondent has answered *like* to both the functional and dysfunctional question, making the response invalid [21]. All questions from the survey are listed in the evaluation table, where an attribute is assigned according to the most frequently used response category (Tables 1, 2, and 3).

The coefficient of customer satisfaction shows how strongly a feature influences satisfaction or dissatisfaction, which helps developers prioritize user requirements. The coefficient consists of positive (satisfaction) and negative (dissatisfaction) values, and the coefficient shows the satisfaction or dissatisfaction with the presence or absence of a feature. The positive value in the formula below shows the satisfaction when a requirement is met, and the negative value shows dissatisfaction when a requirement is not met. The coefficient is calculated as follows (adapted from the study by Berger et al [25]):

A, O, M, and I indicate the frequency of each category shown in the evaluation table. The negative sign in front of the dissatisfaction formula emphasizes the negative influence on customer satisfaction when the requirement is not met or if the feature is not part of the product [20]. Features of the evaluated product or service that yield high positive and negative values should be prioritized and addressed [21].

Table 1. Kano evaluation table.

Characteristics	Dysfunctional				
	Like it	Must-be	Neutral	Accept it	Dislike
Functional					
Like it	Q[a]	A[b]	A	A	O[c]
Must-be	R[d]	I[e]	I	I	M[f]
Neutral	R	I	I	I	M
Accept it	R	I	I	I	M
Dislike	R	R	R	R	Q

[a]Q: questionable.

[b]A: attractive.

[c]O: one-dimensional.

[d]R: reverse.

[e]I: indifferent.

[f]M: must-be.

Table 2. Kano evaluation table—all results (n=83).

Questions	A[a], n (%)	M[b], n (%)	O[c], n (%)	R[d], n (%)	Q[e], n (%)	I[f], n (%)	Category
1. Have a phone in your hand during resuscitation	24 (29)	4 (5)	1 (1)	7 (8)	1 (1)	46 (55)	I
2. Use the app during resuscitation	33 (40)	N/A[g]	6 (7)	5 (6)	1 (1)	38 (46)	I
3. All functions on one side	28 (34)	5 (6)	19 (23)	3 (4)	1 (1)	27 (33)	A
4. Bigger text in the app	13 (16)	5 (6)	9 (11)	5 (6)	1 (1)	50 (60)	I
5. Bigger icons in the app	10 (12)	5 (6)	7 (8)	7 (8)	1 (1)	53 (64)	I
6. Color on the alarm	12 (14)	1 (1)	4 (5)	7 (8)	1 (1)	58 (70)	I
7. Reminder of rhythm check	17 (20)	16 (19)	33 (40)	4 (5)	1 (1)	12 (14)	O
8. Reminder of resuscitation drugs	15 (18)	13 (16)	35 (42)	2 (2)	2 (2)	16 (19)	O
9. Differentiate between adults and children	9 (11)	14 (17)	43 (52)	2 (2)	1 (1)	14 (17)	O
10. Sound on the alarm by the end of a 2-minute cycle	12 (14)	7 (8)	25 (30)	10 (12)	2 (2)	27 (33)	I
11. Turn of the alarm by the end of a 2-minute cycle	16 (19)	13 (16)	18 (22)	4 (5)	1 (1)	31 (37)	I
12. Vibration instead of alarm by the end of a 2-minute cycle	12 (14)	4 (5)	12 (14)	18 (22)	4 (5)	33 (40)	I
13. Access to the patient journal in the app	26 (31)	1 (1)	7 (8)	22 (27)	3 (4)	24 (29)	A
14. Automatic time recording when cardiac arrest is called	39 (47)	5 (6)	18 (22)	2 (2)	1 (1)	18 (22)	A
15. More information about the reversible causes for cardiac arrest	23 (28)	7 (8)	20 (24)	4 (5)	1 (1)	28 (34)	I
16. Sound to guide the chest compression rate (metronome)	29 (35)	4 (5)	8 (10)	19 (23)	1 (1)	22 (27)	A
17. Send CPR[h] history to the DANARREST[i] database	36 (43)	3 (4)	22 (27)	1 (1)	1 (1)	20 (24)	A

[a]A: attractive.

[b]M: must-be.

[c]O: one-dimensional.

[d]R: reverse.

[e]Q: questionable.

[f]I: indifferent.

[g]N/A: not applicable.

[h]CPR: cardiopulmonary resuscitation.

[i]DANARREST: Danish in-hospital cardiac arrest registry.

Table 3. Kano evaluation table—physicians (n=53).

Questions	A[a], n (%)	M[b], n (%)	O[c], n (%)	R[d], n (%)	Q[e], n (%)	I[f], n (%)	Category
1. Have a phone in your hand during resuscitation	13 (25)	1 (2)	N/A[g]	6 (11)	1 (2)	32 (60)	I
2. Use the app during resuscitation	17 (32)	N/A	5 (9)	5 (9)	1 (2)	25 (47)	I
3. All functions on one side	20 (38)	3 (6)	14 (26)	2 (4)	1 (2)	13 (25)	A
4. Bigger text in the app	7 (13)	2 (4)	6 (11)	4 (8)	1 (2)	33 (62)	I
5. Bigger icons in the app	7 (13)	3 (6)	4 (8)	6 (11)	1 (2)	32 (60)	I
6. Color on the alarm	7 (13)	N/A	2 (4)	3 (6)	1 (2)	40 (75)	I
7. Reminder of rhythm check	10 (19)	10 (19)	21 (40)	3 (6)	1 (2)	8 (15)	O
8. Reminder of resuscitation drugs	9 (17)	8 (15)	22 (42)	1 (2)	2 (4)	11 (21)	O
9. Differentiate between adults and children	5 (9)	7 (13)	28 (53)	1 (2)	1 (2)	11 (21)	O
10. Sound on the alarm by the end of a 2-minute cycle	8 (15)	4 (8)	14 (26)	6 (11)	2 (4)	19 (36)	I
11. Turn of the alarm by the end of a 2-minute cycle	11 (21)	12 (23)	11 (21)	2 (4)	1 (2)	16 (30)	I
12. Vibration instead of alarm by the end of a 2-minute cycle	8 (15)	4 (8)	7 (13)	14 (26)	3 (6)	17 (32)	I
13. Access to the patient journal in the app	19 (36)	N/A	5 (9)	17 (32)	2 (4)	10 (19)	A
14. Automatic time recording when cardiac arrest is called	26 (49)	2 (4)	11 (21)	2 (4)	1 (2)	11 (21)	A
15. More information about the reversible causes for cardiac arrest	15 (28)	4 (8)	13 (25)	2 (4)	1 (2)	18 (34)	I
16. Sound to guide the chest compression rate (metronome)	20 (38)	1 (2)	3 (6)	14 (26)	1 (2)	14 (26)	A
17. Send CPR[h] history to the DANARREST[i] database	27 (51)	1 (2)	12 (23)	1 (2)	1 (2)	11 (21)	A

[a]A: attractive.

[b]M: must-be.

[c]O: one-dimensional.

[d]R: reverse.

[e]Q: questionable.

[f]I: indifferent.

[g]N/A: not applicable.

[h]CPR: cardiopulmonary resuscitation.

[i]DANARREST: Danish in-hospital cardiac arrest registry.

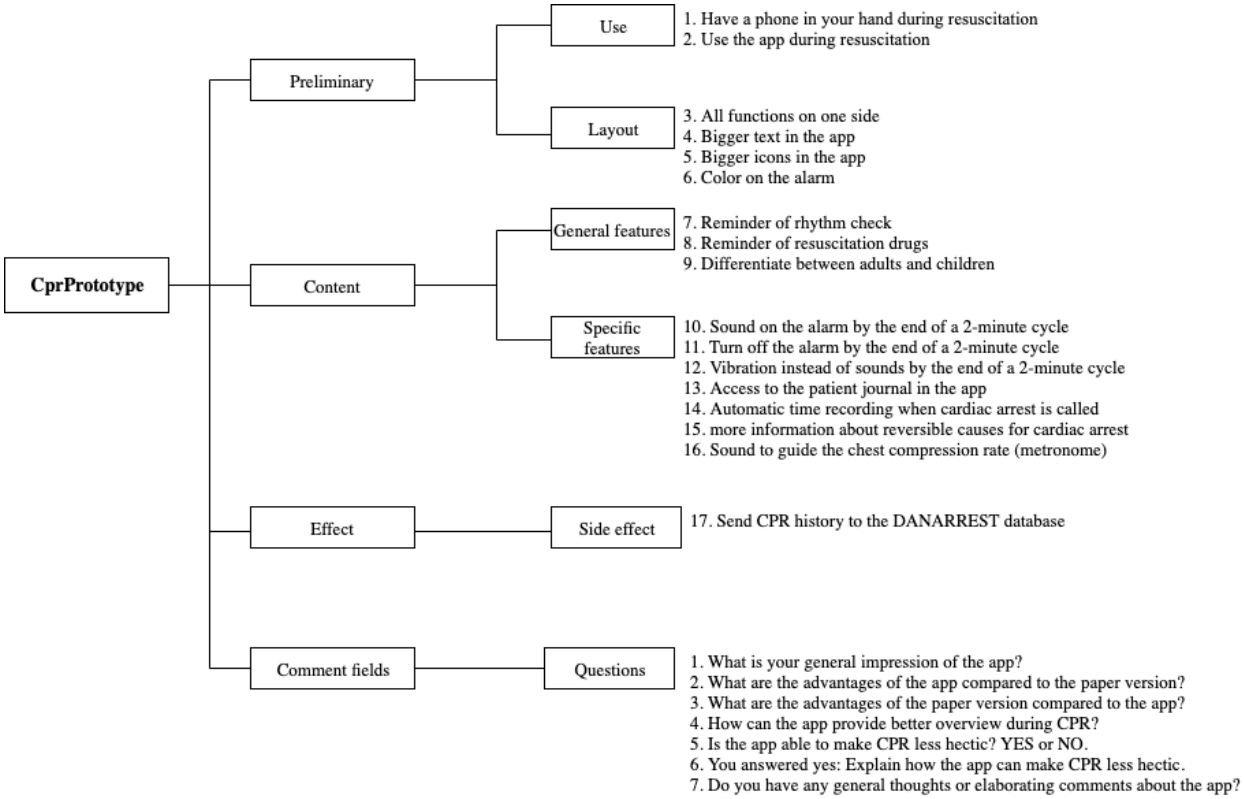
Questionnaire Design

Before distributing the Kano questionnaire to respondents, a pilot test was performed with 5 physicians to evaluate their understanding of the questions. Their feedback helped us modify questions to facilitate understanding and ensure reliability and validity by adapting the wording to the terminology used by both nurses and doctors. The final questionnaire design consists of 3 main categories of questions and 5 subcategories (Figure 4).

The Kano model is used to elicit user requirements and provides insight into the needs and priorities of users, but it does not provide a more detailed explanation of those requirements. To this end, supplementary methods are needed. Therefore, we added additional questions with comment fields that allow respondents to elaborate on answers. As the study participants had not used the CprPrototype app beforehand, we created a video in Danish that explains the app and its purpose [26].

An analysis of the comments was performed to identify themes and interpret the answers to the functional and dysfunctional questions.

Figure 4. The questionnaire framework. CPR: cardiopulmonary resuscitation; DANARREST: Danish in-hospital cardiac arrest registry.



Study Participants

Physicians and nurses are considered the users in this study, as they respond to in-hospital resuscitations [27].

We observed 15 physicians and nurses at an ALS course at a university hospital. In the semistructured interviews, a total of 5 physicians from different hospitals in the Central Denmark Region were interviewed. All 5 interviewees acted as team leaders during in-hospital resuscitation and provided feedback on the CprPrototype app. Finally, the questionnaire was sent to a total of 469 physicians and nurses from different hospitals in the Central Denmark Region who had previously participated in an ALS course. Questionnaires were distributed by course coordinators via email to safeguard their anonymity and ensure compliance with the General Data Protection Regulation (GDPR) [28]. The exact number of nurses compared with the number of physicians is unknown. Respondents were given the option of participating in a draw to win movie tickets.

Results

In total, 17.7% (83/469) of physicians and nurses responded. Overall, 64% (53/83) of respondents were physicians and 36% (30/83) were nurses.

A total of 17 app features were rated and classified using the Kano evaluation table. The Kano evaluation table shows the classification of each question from the questionnaire based on user requirements (Tables 2, 3, and 4). The tables below show the difference between the individual results obtained from physicians and nurses and the aggregated results. The overall Kano analysis (Table 2) indicates that more than half of the features are classified as *indifferent* (9/17, 53%), with only 5 being *attractive* (5/17, 29%) and 3 being *one-dimensional* (3/17, 18%). None of the features are classified as *must-be*. The frequent *indifferent* classification suggests that neither the presence nor absence of most of the evaluated features makes a difference to the user. *One-dimensional* means that the absence of the feature will cause dissatisfaction. The features classified as *one-dimensional* are questions 7, 8, and 9 (Table 2). Questions 3, 13, 14, 16, and 17 (Table 2) point to *attractive* features that delight the users because they are not expected but their absence does not cause dissatisfaction.

Both physicians and nurses were *indifferent* to most of the evaluated features (9/17, 53%). There is, however, a difference in user requirements between physicians and nurses regarding the *one-dimensional* and *attractive* categories. Overall, 29% (5/17) of the features were classified as *one-dimensional* by the nurses compared with 18% (3/17) by physicians. This difference indicates that nurses to a higher degree believe that one particular feature, *sending CPR history to the DANARREST database* (the national in-hospital cardiac arrest quality registry in Denmark), needs to be incorporated in the app and will cause dissatisfaction if it is not. In turn, based on the answers by the physicians, this feature is classified as *attractive*.

Table 4. Kano evaluation table—nurses (n=30).

Questions	A[a], n (%)	M[b], n (%)	O[c], n (%)	R[d], n (%)	Q[e], n (%)	I[f], n (%)	Category
1. Have a phone in your hand during resuscitation	11 (37)	3 (10)	1 (3)	1 (3)	N/A[g]	14 (47)	I
2. Use the app during resuscitation	16 (53)	N/A	1 (3)	N/A	N/A	13 (43)	A
3. All functions on one side	8 (27)	2 (7)	5 (17)	1 (3)	N/A	14 (47)	I
4. Bigger text in the app	6 (20)	3 (10)	3 (10)	1 (3)	N/A	17 (57)	I
5. Bigger icons in the app	3 (10)	2 (7)	3 (10)	1 (3)	N/A	21 (70)	I
6. Color on the alarm	5 (17)	1 (3)	2 (7)	4 (13)	N/A	18 (60)	I
7. Reminder of rhythm check	7 (23)	6 (20)	12 (40)	1 (3)	N/A	4 (13)	O
8. Reminder of resuscitation drugs	6 (20)	5 (17)	13 (43)	1 (3)	N/A	5 (17)	O
9. Differentiate between adults and children	4 (13)	7 (23)	15 (50)	1 (3)	N/A	3 (10)	O
10. Sound on the alarm by the end of a 2-minute cycle	4 (13)	3 (10)	11 (37)	4 (13)	N/A	8 (27)	O
11. Turn of the alarm by the end of a 2-minute cycle	5 (17)	1 (3)	7 (23)	2 (7)	N/A	15 (50)	I
12. Vibration instead of alarm by the end of a 2-minute cycle	4 (13)	N/A	5 (17)	4 (13)	1 (3)	16 (53)	I
13. Access to the patient journal in the app	7 (23)	1 (3)	2 (7)	5 (17)	1 (3)	14 (47)	I
14. Automatic time recording when cardiac arrest is called	13 (43)	3 (10)	7 (23)	N/A	N/A	7 (23)	A
15. More information about the reversible causes for cardiac arrest	8 (27)	3 (10)	7 (23)	2 (7)	N/A	10 (33)	I
16. Sound to guide the chest compression rate (metronome)	9 (30)	3 (10)	5 (17)	5 (17)	N/A	8 (27)	A
17. Send CPR[h] history to the DANARREST[i] database	9 (30)	2 (7)	10 (33)	N/A	N/A	9 (30)	O

[a]A: attractive.

[b]M: must-be.

[c]O: one-dimensional.

[d]R: reverse.

[e]Q: questionable.

[f]I: indifferent.

[g]N/A: not applicable.

[h]CPR: cardiopulmonary resuscitation.

[i]DANARREST: Danish in-hospital cardiac arrest registry.

The comments of the respondents provide additional insight into the questionnaire responses. A total of 68 respondents answered questions with comments. Of the 68 respondents, 25 were nurses and 43 were physicians. From these comments, we identified 7 main themes in an attempt to better understand the differences in user requirements between physicians and nurses (Table 5). The column *frequency* shows how many times both physicians and nurses made a comment, mentioning one of the themes. In the columns *Physicians* and *Nurses*, we show the frequency of themes among physicians and nurses, respectively.

We quantified the comments to use the resulting values as a means to identify patterns and thereby convert them into central themes. This allows us to discern additional differences between physicians and nurses regarding user requirements.

Table 5. Themes of comment fields.

Theme	Frequency	Physicians (n=43), n (%)	Nurses (n=25), n (%)
Technological challenges	24	23 (53)	1 (4)
Keep track of time	33	24 (56)	9 (36)
Documentation and history	24	14 (33)	10 (40)
Disturbing element	14	12 (28)	2 (8)
Areas of improvement			
Functions	19	15 (35)	4 (16)
Design	26	15 (35)	11 (44)
Better guidance	39	22 (51)	17 (68)

The themes in Table 5 are the most dominant among all the comments, revealing that the same themes are salient across the comments of both physicians and nurses, despite differences in frequency.

A clear difference in requirements is the preference of nurses for more information and guidance in the app compared with the physicians. As shown in Table 5, 68% (17/25) of nurses preferred better guidance during CPR in comparison with 51% (22/43) of physicians. One of the main concerns expressed by physicians is technological challenges, such as the risk of the smartphone running out of battery or crashing during CPR. In total, 53% (23/43) of physicians expressed this concern, whereas only 4% (1/25) of nurses commented on this concern. The physicians also fear that the app might become a disturbance during CPR (12/43, 28%). In contrast, only a few of the nurses mentioned the same theme (2/25, 8%).

Table 6 provides an overview of the degree of satisfaction of health care professionals when a requirement is met and dissatisfaction when a requirement is not met. Furthermore, the coefficient indicates a clear priority among the app features evaluated in the study [29]. The values of the coefficients range from 0 to 1. The closer the values are to 1, the more satisfied or dissatisfied a user is with a feature [21]. If a value is greater than 0.5, a feature is assumed to be important [29].

Table 6. Coefficient of user satisfaction.

Questions	Positive value	Negative value
1. Have a phone in your hand during resuscitation	0.33	−0.06
2. Use the app during resuscitation	0.50	−0.07
3. All functions on one side	0.59	−0.30
4. Bigger text in the app	0.28	−0.18
5. Bigger icons in the app	0.22	−0.16
6. Color on the alarm	0.21	−0.06
7. Reminder of rhythm check	0.64	−0.62
8. Reminder of resuscitation drugs	0.63	−0.60
9. Differentiate between adults and children	0.65	−0.71
10. Sound on the alarm by the end of a 2-minute cycle	0.52	−0.45
11. Turn of the alarm by the end of a 2-minute cycle	0.43	−0.39
12. Vibration instead of alarm by the end of a 2-minute cycle	0.39	−0.26
13. Access to the patient journal in the app	0.56	−0.13
14. Automatic time recording when cardiac arrest is called	0.71	−0.28
15. More information about the reversible causes for cardiac arrest	0.55	−0.34
16. Sound to guide the chest compression rate (metronome)	0.58	−0.19
17. Send CPR[a] history to the DANARREST[b] database	0.71	−0.30

[a]CPR: cardiopulmonary resuscitation.

[b]DANARREST: Danish in-hospital cardiac arrest registry.

As shown in Table 6, 3 of the features display positive values of 0.63 and 0.65, which indicate a high degree of satisfaction if the requirements are met. The same features display negative values between 0.60 and 0.71, which indicates high degrees of dissatisfaction if the requirements are not met. These features are classified as *one-dimensional* in Table 2. The coefficients, therefore, show developers how to prioritize among *one-dimensional* requirements when developing the app. The coefficients also show that questions 14 and 17, which are classified as *attractive*, also lead to high degrees of satisfaction if the corresponding features are implemented.

Discussion

Principal Findings

The 3 attributes classified as *one-dimensional* are app features that the user explicitly wants. The 3 features are *reminder of rhythm check*, *reminder of resuscitation drugs*, and *differentiate between adults and children*. App developers should focus on these but may also benefit from implementing the *attractive* features. According to Witell and Löfgren [24], *attractive* attributes are essential when striving for quality in products and services because of the likelihood of generating user satisfaction. This is supported by the coefficient values in Table 6, which display a user satisfaction of above 0.70 for all *attractive* features. Besides, implementing *attractive* features is a low-risk strategy, as low performance with regard to such features will not increase user dissatisfaction because they are unexpected. For example, one feature that could be added to a CPR support app to increase user satisfaction is *automatic time recording when cardiac arrest is called* (0.71). Thus, app developers may use the coefficients (Table 6) to prioritize the features to be implemented when developing an in-hospital CPR support app.

An analysis of the questionnaire comments by the respondents revealed 7 central themes. Most of the comments supported the results of the Kano questionnaire. When asked if they would use the app during resuscitation, the nurses saw it as an *attractive* feature, whereas the responses of the physicians were classified as *indifferent*. More than one theme supported these categorizations. The themes *technological challenges* and *disturbing element* were commented on frequently among physicians compared with nurses, which helped explain why the physicians were *indifferent* to *using an app during resuscitation*. In contrast, nurses frequently mentioned the value of receiving better guidance provided by the app.

Furthermore, physicians frequently made suggestions regarding functionality, whereas nurses made design suggestions. A possible explanation for the differences in user requirements between physicians and nurses is the difference in experience with ALS and the different roles they have during CPR, with physicians most frequently being team leaders. Some physicians have more experience with the ALS algorithm, which could account for their suggestions regarding functionality, whereas it is difficult for nurses to comment on functionality as they are less experienced with the algorithm. The study suggests that it might be beneficial in the future to ensure configurability of the CPR support app, depending on who the user is, given the different requirements of physicians and nurses. However, further research is necessary to investigate the differences between physicians and nurses and how to accommodate their different needs.

Comparison With Previous Work

Several studies that have applied the Kano model in a health care context have recently been published [13,21,29-34]. None of these studies use the Kano model to elicit user requirements in the development of an in-hospital CPR support app. Using the Kano model, we are able to support the claims of both Sulisworo and Maniquiz [32] and Gustavsson et al [13] that the Kano model is a practical tool to elicit different user requirements in a health care context and help prioritize between them. However, we recommend combining the Kano model with qualitative methods. We used observations and interviews to develop the Kano questionnaire, and we supplemented the Kano questionnaire with more open questions and comment fields, encouraging respondents to elaborate on their answers. This is a contribution to the existing Kano model methodology that allowed us to gain an in-depth understanding of user requirements and their priorities based on different roles (ie, physicians and nurses). The study by Gustavsson et al [13] shows the importance of incorporating the perspectives of individuals in multiple roles, which should be taken into consideration when using the Kano model. In doing so, practitioners and researchers can capture a wide range of different user needs. Our study corroborates the findings of Gustavsson et al [13] in the sense that our results show different user requirements based on the different roles of nurses and physicians.

Kalz et al [6] reported about an evaluation study of usability and quality criteria in developing an app for basic life support. In comparison, our study elicits user requirements for an ALS support app. Not only does our study offer recommendations for app developers to create value through implementation of specific *attractive* and *one-dimensional* features, but it also shows the importance of eliciting user requirements. Our results compare with those of Liao et al [34], who focus on exercise apps. Their study also recommends that app developers focus on value creation and invoke positive emotions through *attractive* features. Although there are similarities between the studies, our study extends the findings of Liao et al [34] by recommending that developers also focus on *one-dimensional* features and not only *attractive* features. They focus on *attractive* features because it enhances the strategic advantage in a competitive market with thousands of other apps similar to the one they are developing.

One study suggests that CPR support apps help improve the performance of ALS [12], although the literature on the subject is scarce. Low et al [12] found that test groups using an app during a simulation test improved the quality of ALS compared with the control group not using the app. Our study is the first to focus on app support for in-hospital CPR. Although our study elicits user requirements for such an app, we cannot provide additional support to the existing evidence that CPR support apps help improve CPR performance [12]. However, our study documents the user requirements for apps that are intended to precisely accomplish that goal.

Limitations

A limitation of this study is the low response rate. In total, 469 potential respondents were contacted of which 83 responded, which translates into a response rate of 18%. Compliance with the European GDPR regulation necessitated that the questionnaires were distributed by health care coordinators who were in possession of the names and email addresses of the respondents. However, indirect contact with respondents limited our ability to encourage participation through personal contact and reminders. Furthermore, the study only included physicians and nurses from the Central Denmark Region. However, the Danish health care system is homogeneous. Therefore, the results are, in all probability, comparable across Danish regions.

Furthermore, this study shows challenges in eliciting user requirements for a CPR support app, as most of the features are classified as *indifferent*. One reason for the classification can be attributed to the respondents not having used the CprPrototype app in real life but only having seen it presented in a video. Consequently, they may have difficulty articulating their requirements. In the comment fields, most respondents mentioned that they would like to try the app. If the respondents try the app in a simulation, they may have more specific comments regarding the features they need. Thus, a limitation is that this study does not include a simulated or a clinical resuscitation attempt. Future research should address this limitation. Thus, clinical investigations need to be done once the user requirements have been incorporated.

Conclusions

When developing a product or service such as an app for clinical use, focusing on user requirements is essential. Therefore, we address a knowledge gap by using the Kano model to elicit the

user requirements for an in-hospital CPR support app. In total, 3 requirements classified as *one-dimensional* should be prioritized and incorporated in the app: *reminder of rhythm check*, *reminder of resuscitation drugs*, and *differentiate between adults and children*. This study also revealed 5 *attractive* requirements that should be prioritized in developing CPR apps to increase user satisfaction: *all functions on one side*, *access to the patient journal in the app*, *automatic time recording when cardiac arrest is called*, *sound to guide the chest compression rate(metronome)*, and *send CPR history to the DANARREST database*.

Although this study shows an increasing use of mobile apps during CPR and highlights the importance of eliciting user requirements, our study is uniquely able to provide recommendations to developers on the specific user requirements that should be addressed when developing CPR support apps.

Looking toward the future, it will be important to ensure that the CprPrototype app complies with the European Union Medical Device Regulation of 2017 [35] and relevant national legislation before it can be used in clinical medical practice without fear of personal liability [36]. The next step is to integrate the app with widely used defibrillators into a coherent CPR decision support system for monitoring physiological processes and guiding CPR based on dynamic algorithms. This integration will pave the way for systems interoperability, so the information can be used in, for example, digital hospital command centers [37].

Acknowledgments

The authors would like to thank all the participants and respondents who participated in the study. The authors would also like to thank all the coordinators who helped distribute the questionnaire.

Conflicts of Interest

None declared.

References

1. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
2. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
3. Prgomet M, Georgiou A, Westbrook JI. The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review. J Am Med Inform Assoc 2009;16(6):792-801 [FREE Full text] [doi: 10.1197/jamia.M3215] [Medline: 19717793]
4. Nasser F BinDhim, Trevena L. There's an app for that: a guide for healthcare practitioners and researchers on smartphone technology. Online J Public Health Inform 2015;7(2):e218 [FREE Full text] [doi: 10.5210/ojphi.v7i2.5522] [Medline: 26392848]
5. Gan SK, Koshy C, Nguyen P, Haw Y. An overview of clinically and healthcare related apps in Google and Apple app stores: connecting patients, drugs, and clinicians. Sci Phone Appl Mob Devices 2016 Jul 19;2(1). [doi: 10.1186/s41070-016-0012-7]
6. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014 Mar 19;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]
7. Metelmann B, Metelmann C, Schuffert L, Hahnenkamp K, Brinkrolf P. Medical correctness and user friendliness of available apps for cardiopulmonary resuscitation: systematic search combined with guideline adherence and usability evaluation. JMIR Mhealth Uhealth 2018 Nov 6;6(11):e190 [FREE Full text] [doi: 10.2196/mhealth.9651] [Medline: 30401673]
8. Gruber J, Stumpf D, Zapletal B, Neuhold S, Fischer H. Real-time feedback systems in CPR. Trends Anaesthesia Critical Care 2012 Dec;2(6):287-294 [FREE Full text] [doi: 10.1016/j.tacc.2012.09.004]
9. Ahn C, Cho Y, Oh J, Song Y, Lim TH, Kang H, et al. Evaluation of smartphone applications for cardiopulmonary resuscitation training in South Korea. Biomed Res Int 2016;2016:6418710 [FREE Full text] [doi: 10.1155/2016/6418710] [Medline: 27668257]
10. Soar J, Nolan JP, Böttiger BW, Perkins GD, Lott C, Carli P, Adult advanced life support section Collaborators. European resuscitation council guidelines for resuscitation 2015: section 3. Adult advanced life support. Resuscitation 2015 Oct;95:100-147. [doi: 10.1016/j.resuscitation.2015.07.016] [Medline: 26477701]
11. Ornato JP, Peberdy MA, Reid RD, Feeser VR, Dhindsa HS, NRCPR Investigators. Impact of resuscitation system errors on survival from in-hospital cardiac arrest. Resuscitation 2012 Jan;83(1):63-69. [doi: 10.1016/j.resuscitation.2011.09.009] [Medline: 21963583]
12. Low D, Clark N, Soar J, Padkin A, Stoneham A, Perkins GD, et al. A randomised control trial to determine if use of the iResus application on a smart phone improves the performance of an advanced life support provider in a simulated medical

emergency. Anaesthesia 2011 Apr;66(4):255-262 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.06649.x] [Medline: 21401537]
13. Gustavsson S, Gremyr I, Kenne Sarenmalm E. Using an adapted approach to the Kano model to identify patient needs from various patient roles. TQM J 2016 Jan 11;28(1):151-162 [FREE Full text] [doi: 10.1108/tqm-04-2013-0050]
14. Shah SG, Robinson I. User involvement in healthcare technology development and assessment: structured literature review. Int J Health Care Qual Assur Inc Leadersh Health Serv 2006;19(6-7):500-515. [doi: 10.1108/09526860610687619] [Medline: 17100220]
15. Martin JL, Murphy E, Crowe JA, Norris BJ. Capturing user requirements in medical device development: the role of ergonomics. Physiol Meas 2006 Aug;27(8):R49-R62. [doi: 10.1088/0967-3334/27/8/R01] [Medline: 16772664]
16. Ghauri P, Grønhaug K, Strange R. Research Methods in Business Studies. Cambridge, United Kingdom: Cambridge University Press; 2020.
17. Brinkmann S, Kvale S. InterViews: Learning the Craft of Qualitative Research Interviewing. Thousand Oaks, CA: SAGE Publications; 2015.
18. 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]
19. 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 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
20. Hussain A, Mkpojiogu E, Kamal F. Eliciting User Satisfying Requirements for an E-health Awareness System Using Kano Model. In: Proceedings of the 14th International Conference On Applied Computer And Applied Computational Science. 2015 Presented at: ACACOS'15; April 23-25, 2015; Kuala Lumpur, Malaysia.
21. Chang W, Chang Y. Patient satisfaction analysis: identifying key drivers and enhancing service quality of dental care. J Dental Sci 2013 Sep;8(3):239-247 [FREE Full text] [doi: 10.1016/j.jds.2012.10.006]
22. Kano N, Seraku N, Takahashi F, Tsuji S. Attractive quality and must-be quality. J Japanese Soc Qual Control 1984;14(2):147-156 [FREE Full text] [doi: 10.20684/quality.14.2_147]
23. Herzberg F, Mausner B, Snyderman B. The Motivation to Work. New York, USA: Wiley; 1959.
24. Witell L, Löfgren M. Classification of quality attributes. Manag Serv Qual 2007 Jan 30;17(1):54-73. [doi: 10.1108/09604520710720674]
25. Berger C, Blauth R, Boger D. Kano's methods for understanding customer-defined quality. Center Qual Manag J 1993;2(4):2-28.
26. Mathiasen M. Mobil App Til Avanceret Genoplivning. YouTube. 2019 Mar 11. URL: https://www.youtube.com/watch?v=Qihi3MhXdAE [accessed 2020-12-10]
27. Lauridsen KG, Schmidt AS, Adelborg K, Løfgren BO. Organisation of in-hospital cardiac arrest teams - a nationwide study. Resuscitation 2015 Apr;89:123-128. [doi: 10.1016/j.resuscitation.2015.01.014] [Medline: 25617486]
28. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2020-12-10]
29. Rasche P, Mertens A, Brandl C, Liu S, Buecking B, Bliemel C, et al. Satisfying product features of a fall prevention smartphone app and potential users' willingness to pay: web-based survey among older adults. JMIR Mhealth Uhealth 2018 Mar 27;6(3):e75 [FREE Full text] [doi: 10.2196/mhealth.9467] [Medline: 29588268]
30. Keller AC, Bergman MM, Heinzmann C, Todorov A, Weber H, Heberer M. The relationship between hospital patients' ratings of quality of care and communication. Int J Qual Health Care 2014 Feb;26(1):26-33. [doi: 10.1093/intqhc/mzt083] [Medline: 24257162]
31. Mitrabasu N. Integrating Kano model and Herzberg two factor theory to unveil the third quality factor of patient satisfaction in a multispecialty outdoor medical centre. Int J Emerg Sci Eng 2013;1(7):1-5.
32. Paraschivescu A, Cotîrlet A. Kano model. Eco Transdiscip Cogn 2012;15(2):116-124.
33. Sulisworo D, Maniquiz N. Integrating Kano's Model and Servqual to Improve Healthcare Service Quality. In: Proceedings of the International Conference on Green World in Business and Technology. 2012 Presented at: GWBT'12; March 21-24, 2012; Yogyakarta, Indonesia.
34. Kim D, Chang H. Key functional characteristics in designing and operating health information websites for user satisfaction: an application of the extended technology acceptance model. Int J Med Inform 2007;76(11-12):790-800. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]
35. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj [accessed 2020-12-10]
36. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. J Am Med Assoc 2019 Oct 4:-epub ahead of print. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]

37. Tampa General Hospital. Tampa General Hospital and GE Healthcare Partner for Next Level Care Coordination. Tampa General Hospital. 2018. URL: https://www.tgh.org/news/media-resources/tampa-general-hospital-and-ge-healthcare-partner-next-level-care-coordination [accessed 2020-12-10]

Abbreviations

ALS: advanced life support
CPR: cardiopulmonary resuscitation
DANARREST: Danish in-hospital cardiac arrest registry
GDPR: General Data Protection Regulation

Edited by G Eysenbach; submitted 03.09.19; peer-reviewed by B Metelmann, D Dimitrov; comments to author 19.10.19; revised version received 02.11.19; accepted 14.12.19; published 05.01.21.

Please cite as:
Müller SD, Lauridsen KG, Palic AH, Frederiksen LN, Mathiasen M, Løfgren B
Mobile App Support for Cardiopulmonary Resuscitation: Development and Usability Study
JMIR Mhealth Uhealth 2021;9(1):e16114
URL: https://mhealth.jmir.org/2021/1/e16114
doi:10.2196/16114
PMID:33399539

Original Paper

Developing an Adaptive Mobile Intervention to Address Risky Substance Use Among Adolescents and Emerging Adults: Usability Study

Lara N Coughlin[1,2], PhD; Inbal Nahum-Shani[3], PhD; Meredith L Philyaw-Kotov[1], MSc; Erin E Bonar[1,2], PhD; Mashfiqui Rabbi[4], PhD; Predrag Klasnja[5], PhD; Susan Murphy[4,6], PhD; Maureen A Walton[1,2], MPH, PhD

[1]Department of Psychiatry, Addiction Center, University of Michigan, Ann Arbor, MI, United States
[2]Injury Prevention Center, University of Michigan, Ann Arbor, MI, United States
[3]Institute of Social Research, University of Michigan, Ann Arbor, MI, United States
[4]Department of Statistics, Harvard University, Cambridge, MA, United States
[5]School of Information, University of Michigan, Ann Arbor, MI, United States
[6]Computer Science, Harvard John A Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States

Corresponding Author:
Lara N Coughlin, PhD
Department of Psychiatry
Addiction Center
University of Michigan
2800 Plymouth Rd
Ann Arbor, MI, 48109
United States
Phone: 1 734 615 4774
Email: laraco@med.umich.edu

Abstract

Background: Substance use among adolescents and emerging adults continues to be an important public health problem associated with morbidity and mortality. Mobile health (mHealth) provides a promising approach to deliver just-in-time adaptive interventions (JITAIs) to prevent escalation of use and substance use–related consequences.

Objective: This pilot study aims to describe the iterative development and initial feasibility and acceptability testing of an mHealth smartphone app, called MiSARA, designed to reduce escalation in substance use.

Methods: We used social media advertisements to recruit youth (n=39; aged 16-24 years, who screened positive for past-month binge drinking or recreational cannabis use) with a waiver of parental consent. Participants used the MiSARA app for 30 days, with feasibility and acceptability data reported at a 1-month follow-up. We present descriptive data regarding behavior changes over time.

Results: The results show that most participants (31/39, 79%) somewhat liked the app at least, with most (29/39, 74%) rating MiSARA as 3 or more stars (out of 5). Almost all participants were comfortable with self-reporting sensitive information within the app (36/39, 92%); however, most participants also desired more interactivity (27/39, 69%). In addition, participants' substance use declined over time, and those reporting using the app more often reported less substance use at the 1-month follow-up than those who reported using the app less often.

Conclusions: The findings suggest that the MiSARA app is a promising platform for JITAI delivery, with future trials needed to optimize the timing and dose of messages and determine efficacy.

(JMIR Mhealth Uhealth 2021;9(1):e24424) doi:10.2196/24424

KEYWORDS

mHealth; adolescents; young adults; just-in-time adaptive intervention; alcohol misuse; cannabis; mobile phone

Introduction

Background

Advances in mobile health (mHealth) not only allow for accessible and cost-effective interventions but also offer novel opportunities for delivering personalized and adaptive interventions in the real world. A form of personalized medicine, just-in-time adaptive interventions (JITAIs) operationalize the use of real-time data collection in individualizing content and delivery of intervention strategies [1,2]. JITAIs have been developed and evaluated for a wide range of behavioral health issues (eg, cardiovascular disease [3], diabetes [3,4], mental illness [5], smoking cessation [6,7]).

mHealth approaches are lacking in the selective prevention of substance use (eg, risky use of alcohol, cannabis) among youth during this critical period when substance use is initiated and reaches peak prevalence [8-10]. For example, past-month binge drinking (eg, >4 for females or >5 drinks for males) is reported by 10.2% of youth aged 16 to 17 years, 26.2% of youth aged 18 to 20 years, and 45.4% of youth aged 21 to 25 years [11]. Such risky drinking increases the likelihood of health, academic, and social consequences and the development of alcohol use disorders [12-15] and is associated with morbidity and mortality (ie, injury, violence, suicide, overdose) [10,12-14]. Past-month use of cannabis is reported in 6.5% of adolescents and 22.1% of emerging adults [16]. The recent legalization of recreational cannabis (ages ≥21 years) in many areas heightens concern, particularly given the decreases in perceptions of risk [17,18] and effects on neuromaturational brain development (eg, compromised decision making and inhibitory control functioning) [19,20]. Furthermore, recent trends in the simultaneous use of alcohol and cannabis (defined as consumption patterns of alcohol and cannabis such that the individual reports the effects overlap), which is associated with greater consumption and consequences [21-24], highlight the urgent need for interventions to prevent risky substance use.

Critical challenges exist in developing mHealth interventions to reduce substance use among adolescents and emerging adults. Existing mHealth interventions for substance use are primarily developed for adults and/or individuals with substance use disorders (SUDs; <10% of adolescents and emerging adults) [25], which miss a larger proportion of adolescents and emerging adults with risky substance use. For example, emerging research demonstrates the benefit of relapse prevention apps (eg, A-CHESS, ChessHealth [26]) delivered after SUD treatment [27-32], underscoring the promise of mHealth tools, but they are tailored for individuals with greater readiness to change and substance use severity. In addition, these tools reinforce treatment concepts, which those with risky use have yet to receive, and include therapist support, which reduces scalability and may be unnecessary for those with lower severity substance use.

In 2015, recognizing the need for prevention-focused apps, the Society for Adolescent Health and Medicine released a health education app (including substance use) for parents of adolescents, THRIVE (Society for Adolescent Health and Medicine). Although beneficial, this app primarily consists of didactic, static information (eg, facts, conversation starters, immunization records, find a doctor), so content is not tailored in real time. In addition, a systematic review of 12 studies concluded that although mHealth shows promise to address substance use, more studies are needed to develop and test the efficacy using current platforms [29]. Furthermore, there was heterogeneity across studies in terms of length of intervention delivery, intervention dose delivered, remote delivery platform, sample severity, and content included. Therefore, research to identify optimal components of mHealth tools to reduce substance use among adolescents and emerging adults is needed.

Daily monitoring and ecological momentary assessment (EMA) studies can inform JITAIs by explaining why youth binge drink or use cannabis on some days and not others. However, prior daily monitoring and EMA studies of substance use are generally lacking among adolescents, with most conducted among college student samples [33,34]. Daily monitoring and EMA studies of college students underscore the importance of both negative affect (ie, craving and anxiety) in alcohol [35,36] and cannabis use [37,38] as well as positive affect (eg, enhancement motives) in binge drinking and cannabis use [37-39].

Adequate engagement with mHealth apps is necessary for success [1,2]; however, low engagement and attrition is common [40-42], including in those apps focusing on substance use [29]. This concern is particularly salient for not-in-treatment samples of adolescents and emerging adults who may not view their substance use as problematic (eg, lower readiness to change). Thus, engagement in mHealth tools for those not seeking behavior change is a unique challenge. Although passive data collection is rapidly advancing, self-reporting remains to be a staple for tailoring intervention delivery for substance use, as passive detection of key psychological mechanisms such as motives for use is not possible. Daily monitoring and EMA studies, including a recent meta-analysis, show modest response rates (eg, 60%-70%) [43-45], despite payments of US $2 to US $5 per day (or US $60-US $150 over a month), which creates a challenge for scalability [45]. Herein, we describe recent user-centered design work to develop the initial substance abuse research assistant (SARA) app, aimed at enhancing adolescents' and emerging adults' engagement with daily monitoring surveys. We then describe the acceptability and feasibility findings from our pilot study of a JITAI version of SARA, called MiSARA.

An Overview of Prior Work to Iteratively Develop SARA

Consistent with the preparation phase of the Multiphase Optimization Strategy (MOST) framework [46], our team's prior work [47] used an iterative process through a series of formative studies with adolescents and emerging adults to inform and refine our mHealth app (see Rabbi et al [47]). These prior studies focused on using the initial SARA app for enhancing engagement in daily assessments, which are critical for JITAI tailoring, and did not include delivery of intervention content [47]. Specifically, the SARA app integrated daily and weekly self-report surveys and tasks (eg, reaction time, spatial memory tasks) with theoretically grounded engagement strategies to improve the completion of self-report surveys with minimal financial incentive, enhancing scalability. As described

in a study by Rabbi et al [47], to develop and refine SARA, we conducted (1) a web-based survey to assess user perceptions about design features, (2) focus groups to obtain in-depth qualitative feedback on the app, and (3) a preliminary microrandomized trial (MRT; n=18) to refine app features [47-49]. The iteratively refined SARA app employed gamification through a virtual aquarium environment, which displays more fish as self-reporting increases; gamification included levels (ie, unlocking fish based on completion) and small monetary incentives.

Subsequently, as described in a study by Nahum-Shani et al (unpublished data, 2021), we conducted a larger MRT (N=68) of SARA with youth reporting past-month binge drinking or cannabis use to further refine the SARA app. Using data from this MRT, we assessed the proximal effect of delivering theory-based engagement strategies, including reciprocity and nonmonetary reinforcement of daily survey completion. In the MRT, we included much lower financial incentives (eg, 30-day incentives averaged US $6.24 per participant) than prior daily monitoring studies while yielding fairly comparable rates of self-reporting (eg, 60.3% daily surveys, 75% weekly surveys). Regarding acceptability, most youth rated SARA as at least somewhat fun (76.3%) and interesting (72.9%). The findings supported the efficacy of reciprocity strategies (ie, inspirational quote delivered before daily survey prompts regardless of engagement), with participants liking the inspirational quotes. Nonmonetary reinforcement in the form of entertaining memes or graphics interchange format (gifs) images delivered after self-reporting did not appear to increase engagement, with feedback indicating a preference for rewards that are more consistent with the aquarium environment (eg, points, fish). Finally, mixed evidence was found for employing nonmonetary reinforcement in the form of personalized feedback graphs based on daily data; participants wanted their data available at all times, rather than contingent on self-reporting. Together, prior work from our team with the SARA app involving iterative refinement and analysis of the MRT set the stage for the development of the MiSARA JITAI aimed at reducing alcohol and cannabis use among at-risk adolescents and emerging adults. The goal of this study is to establish the feasibility and acceptability of the MiSARA JITAI and to explore preliminary outcomes related to behavior change.

Methods

Study Design and Protocol

Recruitment

Participants were recruited through social media advertisements (eg, Facebook, Instagram) over a 2-week period to obtain a broad sample of adolescents and emerging adults aged 16 to 24 years (n=39) who reported binge drinking or cannabis use. When users clicked the advertisement, they were redirected to the web-based screening consent. A waiver of parental consent was obtained for minor participants (aged 16-17 years), given the minimal risk of the study and the possibility that adolescents would prefer not to participate in this selective prevention study on substance use if parental consent was required. Eligible participants were invited to provide contact information for verification purposes, which consisted of a completely automated public Turing test to tell computers and humans apart (CAPTCHA), checking for duplicate internet protocol addresses and examining social media profiles. Once verified by staff, eligible participants were invited via email or text to join the longitudinal study, where they would complete a baseline survey in Qualtrics (with baseline consent embedded and a waiver of parental consent), download, and use the MiSARA app for 1 month and complete a 1-month follow-up survey about the app functionality, design and content, and preliminary outcomes. The MiSARA study was approved by the Medical School at the University of Michigan (IRBMED ID: HUM00148393).

Inclusion or Exclusion Criteria

Screened individuals were eligible for the study if they (1) had an iPhone to download the app (as the current version was compatible only with iOS not Android), (2) screened positive for past-month binge drinking (>4 drinks for females or >5 drinks for males) or any past-month cannabis use without a medical cannabis card, and (3) met the study verification criteria described earlier.

Self-Report Measures

At baseline and 1-month follow-up, we collected descriptive data [50,51] (Table 1) as well as quantity and frequency of use (eg, Alcohol Use Disorder Identification Test-Consumption [AUDIT-C], 30-day Timeline Follow Back) [52-56]. We also assessed the consequences of use (assessed separately for alcohol and cannabis) [57,58], intention, importance [59], and confidence of change [60] (ranging from 0 [not at all] to 10 [very]), perceived risk [61] (ranging from 0 [not at all risky] to 4 [extremely risky]), reasons for alcohol and/or cannabis use [62,63] (eg, coping, enhancement, social), and past-month driving under the influence of alcohol or cannabis [64] (ranging from 0 [never] to 4 [>10 times]).

Daily surveys included single-item measures of stress, mood, loneliness, free time, fun, sensation seeking, and hopefulness (ranging from 0 [not at all] to 4 [a lot]), in addition to daily tasks [65-70]. The tasks were designed to capture the acute cognitive effects of substance use in a naturalistic setting by assessing the reaction time (ie, tapping speed) and spatial memory (ie, retrace flower patterns that briefly light up). Weekly surveys, prompted every Sunday, measured behavioral intentions to use cannabis and/or alcohol in the next week [71] (0 [not at all] to 4 [very]) and quantity of alcohol and cannabis use each day in the prior week [61,72].

During the 1-month follow-up, participants were also asked acceptability questions (eg, “Is the app content appropriate for people your age?” “Overall, how would you rate the app’s appearance?” “Did the aquarium affect the amount of time you spent using the app?”). Participants were also asked about self-reported *points* earned in the app, which is a proxy measure for app engagement. The prototype MiSARA app contained a software error, which resulted in a lack of archiving daily or weekly survey data or paradata, which prevented further examination of objective engagement data.

Table 1. Intervention content topics by day of the week.

Day of week	Content area	Example intervention messages[a]
Sunday	Coping with stress	• Taking time for yourself can help reduce the amount of stress we experience! How, if at all, could you prevent stressing? You are worth it!
Monday	Coping with negative mood	• Accepting things that happen in life can be hard. Sometimes to be strong means asking for help. Talk to someone who cares or text 741-741
Tuesday	Prosocial people and activities	• What're three things that used to make you feel excited or happy? Try doing/thinking about these things when you feel bored/down. Good luck!
Wednesday	Alternatives to address motives for substance use	• What're things you used to love doing with friends that you miss? Imagine doing them again. Who has your best interests in mind?
Thursday	Alternative leisure activities	• Doing fun or exciting activities can lift your mood! Try thinking of an exciting thing you'd like to do. Making a bucket list can be fun
Friday or Saturday	Tools and protective strategies to reduce substance use	• Exercising's a natural high. Try a new workout & challenge yourself to a fitness goal. Look in the mirror & say 3 things you like about you. • Some take breaks between drinks, so they don't overdo it. Remind yourself that you are the fun and you make the good time. Be true to you! • Think about how much money you could save by staying away from too much partying. Sounds like a great way to have extra cash. Nice going!

[a]Although daily or weekly app data were not saved because of a software error, the intervention messages were tailored based on each participant's daily responses and weekly reports of prior week substance use.

Incentives

Following app installation, participants received US $20 remuneration (eg, Amazon gift card). As a software glitch prevents the determination of survey completion, incentives were provided assuming 100% compliance with daily surveys; this along with a variable probability of reinforcement schedule could result in receipt of US $11 to US $21. Research staff contacted participants on Fridays with their weekly incentive total. Participants self-administered 1-month follow-up surveys and received US $30 in remuneration via an electronic gift card.

MiSARA JITAI

For more information about SARA intervention development, see prior publications [47-49].

Development

The MiSARA JITAI (Figure 1) was informed by SARA preliminary studies and developed for the iOS platform because most SARA participants were iPhone users and funding was insufficient for development on iOS and Android platforms. We created a more realistic aquarium interface, made personalized feedback graphs (based on within-app self-report data) available on-demand, and modified the daily survey completion window, from 6 PM to midnight in SARA and from noon to 6 PM in MiSARA. This change was made so that data from the daily survey could be used to tailor an intervention message delivered at 7 PM, potentially before a drinking event (although the timing with cannabis use is not clear). Consistent with the literature on the norm of reciprocity [73-75] and given positive findings for the daily inspirational messages (Nahum-Shani et al, unpublished data, 2021), we retained this feature, which was delivered at noon when the survey window opened. To inform refinement of the daily inspirational messages at noon, Amazon Mechanical Turk workers (n=20; mean age 21.3, SD 2.2 years; 8/20, 40% male) identified famous quotes, song lyrics, and images relevant for adolescents and emerging adults, followed by refinement by the study team to provide supporting language and by undergraduate and postgraduate students for relevance and appropriateness (eg, no images glamourizing use, no offensive jokes). We also enhanced rewards within the aquarium (eg, revamping fish artwork and delivery schedule), added a resource page with information about services and how to get help (eg, mental health, substance use), and incorporated tailored intervention content delivered at 7 PM daily (Figures 2 and 3).

Intervention content was developed and refined using an iterative participatory approach by the study team and students to fit with evidence-based intervention content (eg, motivational interviewing, mindfulness, behavioral activation) that was appealing to adolescents and emerging adults to maximize engagement. Intervention message content included supportive affirmations and tips, inspirational images to reinforce content, web links to articles (eg, 50 fun things to do without alcohol), or other web-based resources (eg, Drinkaware, YouTube meditation videos). Content areas of tailored messages varied by the day of the week (Table 1) and were tailored based on participant responses to daily surveys of factors related to substance use (eg, stress, mood, leisure activities) and to weekly surveys about substance use and motives for use [76]. For example, messages on Sunday focused on coping with stress so that participants indicating high stress on the daily stress question ("Right now, I feel stressed") received an intervention message tailored to cope with their stress level, such as "Challenging situations can take a toll on our mental or physical health. Sometimes it helps to take a moment to ground ourselves. You're resilient!" along with a link to a recentering mindfulness

meditation YouTube video. It should be noted that if surveys were not completed, then generic content was delivered on the specified topic for that day. If a daily intervention message was delivered (based on 0.33 probability), message content was directly focused on substance use only on Wednesdays, Fridays, and Saturdays (as opposed to every day), whereas intervention message content on other days focused on upstream factors associated with use (eg, stress, mood, prosocial support, and leisure activities) to prevent fatigue and minimize unintentionally priming participants to use substances (Table 1).

Figure 1. Overview of MiSARA app and content development.

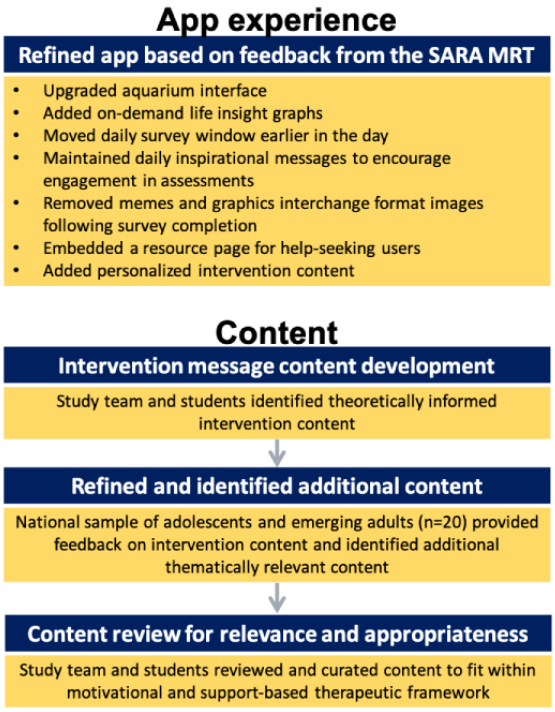


Figure 2. Screenshots of (a) MiSARA aquarium app environment and (b) resources page.

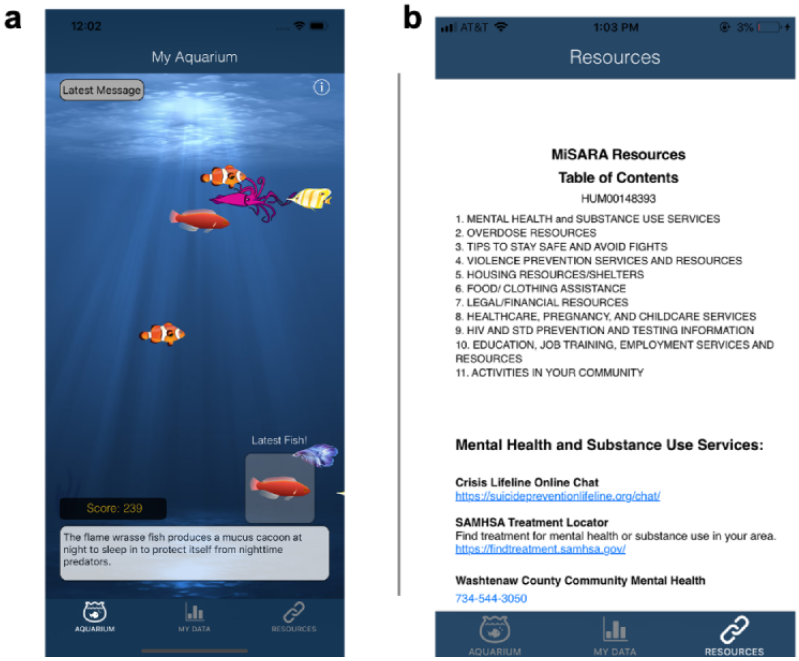
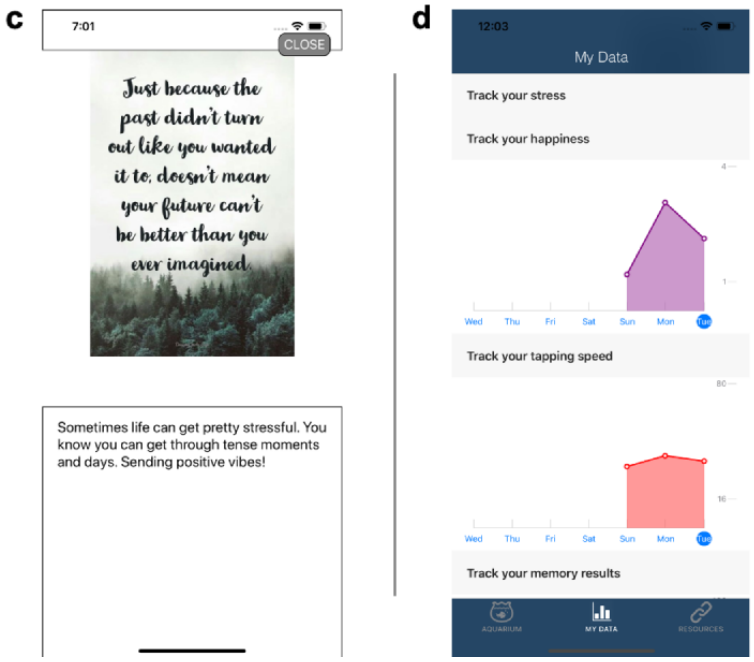


Figure 3. Screenshots of (c) example intervention message and (d) life insights graphs.



Procedure

For a period of 30 days, participants were prompted via the MiSARA app to complete: brief (2-3 min) daily surveys, short (5-10 min) weekly surveys, and daily tasks (1-2 min each). On the basis of prior work (Nahum-Shani et al, unpublished data, 2021), each day at noon, participants received one inspirational message to encourage engagement combined with a reminder to complete the daily survey by 6 PM (ie, 7 questions, 2 tasks). On Sundays, participants were prompted to complete a weekly survey (ie, 26 questions, including past-week alcohol and cannabis use). At 7 PM each day, participants were randomized (probability=0.33) to receive (1) a tailored intervention message, (2) a fun fact on a random topic (eg, "7% of American adults believe that chocolate milk comes from brown cows. 7% doesn't sound like a lot, but that works out to 16.4 million people"), or (3) no message, to reduce habituation to intervention content and consistent with the lower severity, not-in-treatment sample. We used the fun fact to infuse novelty so that participants would not be fatigued by too many intervention messages. To encourage engagement, participants earned points to unlock new fish for completing surveys and tasks. Following the 30-day app experience, participants completed a follow-up survey.

Analyses

Analyses focus on the acceptability and feasibility of the MiSARA app, as this pilot study was not intended to examine MRT outcomes (eg, not powered). Open-ended, user-centered feedback about the MiSARA app experience is described. Participant ratings of feasibility and acceptability are then presented descriptively (means, SDs, proportions), which was the primary focus of this pilot study. Next, consistent with the best practices for pilot studies [76], we report (1) the association between baseline characteristics and self-reported app engagement based on zero-order correlations and a linear regression model and (2) change in preliminary outcomes from baseline to follow-up and the association between change in outcomes and self-reported app engagement.

Results

Recruitment Feasibility

Social media advertisements resulted in 239 participants completing the screening survey over 2 weeks, with 71 (29.7%) participants eligible. Of those eligible, 39 (55%) participated in this pilot study. No meaningful differences in age, race, or alcohol use were detected between those who did elect to participate compared with those who did not elect to participate. Those who enrolled were more likely to be female (24/39, 62%) than those who were eligible but did not enroll (11/32, 34%). In addition, those who enrolled used cannabis less frequently (mean 1.8, SD 2.1 days of cannabis use in the past 30 days) than those who were eligible but did not participate (mean 3.1, SD 2.7 days). All enrolled participants completed the 1-month follow-up.

Sample Characteristics

Of 39 participants, 24 (62%) were female; 12 (31%) were aged 16 to 20 years; and 27 (69%) were aged 21-24 years (mean age 20.7 years, SD 2.1); 9 (23%) were identified as racial minorities and 5 (13%) were identified as Hispanic (Table 2). In the baseline survey, all participants reported alcohol use (AUDIT-C score: mean 5.9, SD 2.4), with 37 (95%) of 39 participants reporting past-month binge drinking, and 19 (49%) reporting drinking weekly or more. Of those reporting past-month cannabis use (23/39, 59%, of which 21 also reported past-month binge drinking), 11 (48%) reported using cannabis weekly or more, with 4 (17%) using daily. Overall, 36 participants reported substance use consequences in the past month, with an average of 8.5 (SD 6.7; range 0-45). Two-thirds reported driving under the effects of alcohol in the past month, and 6 (15%) of 39 participants reported driving under the effects of cannabis. As expected, given the non–treatment-seeking sample, intention

and importance of reducing alcohol use was low (mean 2.4, SD 2.3 and mean 2.5, SD 1.9 respectively, on a 1-10 scale), whereas confidence in the ability to reduce alcohol was high (mean 8.3, SD 2.2 on a 1-10 scale), with a similar pattern for cannabis use, and perceived risk of regular alcohol (mean 2.5, SD 0.9 on 0-4 scale, 0 indicates no risk) and cannabis use (mean 1.3, SD 0.9 on 0-3 scale) was low. The most common motive for drinking was social, followed by enhancement (ie, to increase positive affect), with coping and other motives being the least common. Similarly, the most common reason for cannabis use was enhancement, followed by other motives and coping. The most common reason for *not* drinking or not using cannabis in the past week was not wanting to drink or use.

Table 2. Sample characteristics at baseline (n=39).

Demographics	Participants, n (%)
Male sex	15 (38)
Age (years)[a]	
16-17	5 (13)
18-20	7 (18)
21-24	27 (69)
Race	
White	30 (77)
African American	3 (8)
Asian	2 (5)
Other	4 (10)
Hispanic ethnicity	5 (13)
Education	
Some high school	5 (13)
High school diploma or general educational development only	7 (18)
Some college	13 (33)
College graduate	14 (36)
Past-month substance use and consequences	
Binge-drinking only	16 (41)
Cannabis only	0 (0)
Alcohol and cannabis	23 (59)
Any substance use consequences	36 (92)
Drinking and driving	27 (69)
Cannabis and driving	6 (15)
Past-week top reasons for use[b]	
Alcohol use (n=31)	
Coping	1 (3.2)
Enhancement	11 (35.5)
Social	15 (48.4)
Other	4 (12.9)
Cannabis use (n=14)	
Coping	1 (7.1)
Enhancement	10 (71.4)
Social	0 (0.0)
Other	3 (21.4)

[a]Mean 20.7 (SD 2.1).

[b]Past-week reasons for alcohol or cannabis use questions were only asked to those participants who reported use in the past week at baseline.

MiSARA Acceptability

Most participants (31/39, 79%) at least somewhat liked MiSARA, with almost all reporting that the design was appealing (36/39, 92%; Table 3), but most participants also felt it was not interactive enough (27/39, 69%). Unsurprisingly, given the research context, ratings for *fun to use*, *interesting*, and *the aquarium was fun* were modest, with 27 (75%) of 36 participants indicating the aquarium itself did not influence the time spent using the app. Most importantly, the content was viewed as age-appropriate by nearly all participants (36/39, 92%). Data collection via the app was acceptable, with 36 (92%) of 39 participants agreeing that they were comfortable with self-reporting, and 28 (72%) of 39 participants agreed that they were comfortable with the app collecting passive data (such as geolocation). Although the app was easy to use (34/39, 87%) with clear instructions (35/39, 90%), 9 (23%) participants sometimes or regularly had technical problems (eg, crashing), 15 (38%) rarely had issues, and 15 (38%) never had trouble. Finally, most participants (29/39, 74%) rated the app 3 or more stars (24/39, 62% rated 3.5 or more stars; 19/39, 49% rated 4 or more stars).

Regarding appearance, qualitatively, participants reported liking the MiSARA aquarium and receiving fish, which was *nice* and *pleasant*; however, they recommended adding more animation (ie, increasing the type and number of fish, swimming speeds) and brighter colors and unique fonts. Next, participants requested greater interactivity, such as being able to feed the fish, tap on the glass, buy items or more fish, view a count of earned fish, and select fish to see maritime information at any time (not just when receiving new fish). They also recommended greater personalization (eg, type of tank or background, home page, and colors). Regarding features, participants consistently, positively commented on the life insights personal feedback graphs, requesting the addition of step counts and showing data trends for longer than a week. Similarly, they enjoyed the reaction time and spatial memory tasks but wanted more novelty in these tasks. Regarding content, participants commonly made positive comments about the *uplifting* inspirational quotes, messages, and images, links, or videos, requesting the ability to archive content, which disappeared, for later access. In addition, participants requested the ability to personalize the type of inspirational content topics or quotes, which were too *vague*. Despite all content being vetted by the study team and students, 2 people said they found some content to be *strange* or *patronizing*. Regarding functionality, although positive comments focused on the ease of the survey completion and appreciation of the reminders, a key recommendation for improving functionality was to add the ability to customize the timing of the daily survey window (eg, one person worked a mid-day shift that overlapped with the preset window). Additional suggestions included adding a second reminder to complete the survey before the end of the window, a clearer timeline of required activities and explanation of the app's purpose, and a *pause* feature for the spatial memory task or fewer rounds (eg, to pick up a phone call). Finally, 2 people wanted more money to complete self-report surveys, and several participants noted technical issues (eg, crashing).

Table 3. MiSARA app acceptability (n=39).

Content area and item	Participants, n (%)[a]
Appearance	
Overall	• 31 (79) rated average, good, excellent
Appealing design	• 36 (92)
Interactivity	• 26 (67) not enough • 12 (31) just right • 1 (3) too much
Enjoyment	
Fun to use	• 27 (69)
Interesting	• 25 (64)
Aquarium fun to use	• 19 (49)
Aquarium influenced time using app	• 27 (75) not at all
Content	
Messages appropriate for age	• 36 (92)
Data acceptability	
Comfort answering self-report questions	• 36 (92) agree or strongly agree
Comfort with passive data collection in app	• 28 (72) agree or strongly agree
Prefer self-reporting in MiSARA rather than by text or phone	• 32 (82)
Technical issues	
Easy to use	• 34 (87)
Clear app use instructions	• 35 (90)
Frequency of problems	• 15 (38) none • 15 (38) rarely • 9 (23) sometimes or regularly

[a]Percent rated *3* or higher (eg, somewhat) unless otherwise indicated.

Engagement Feasibility

Given an unanticipated software glitch, the prototype MiSARA app did not consistently save daily or weekly survey data or paradata, which prevented objective examination of app engagement. However, on the follow-up survey, participants self-reported *points* earned, a proxy for app engagement. Participants reported an average earning of 986.2 points (SD 688.0; range 0-2110; first quartile 324, third quartile 1591, median 1163; n=34; 5 missing). Self-reported app engagement (ie, points) was positively correlated with female sex (r=0.2) and slightly negatively correlated with age (r<−0.1) and negatively correlated with baseline markers of severity of substance use, including AUDIT-C score (r=−0.3), total number of past-month drinks (r=−0.4), days of binge drinking (r=−0.4), and days of past-month cannabis use (r=−0.16). In an adjusted regression model (Table 4) looking at reported app engagement as a function of age, gender, total number of past-month drinks, and total past-month cannabis use, alcohol use was the only baseline characteristic meaningfully associated with app engagement (estimate −10.5, SE 4.4; 95% CI −19.2 to −1.8), such that more alcohol use was associated with less app engagement.

Table 4. Linear regression of baseline predictors of app engagement.

Variable	Estimate (SE)	95% CI
Age	13.6 (53.2)	−90.7 to 117.8
Gender (referent=male)	−314.2 (216.6)	−738.7 to 110.4
Past-month alcohol use	−10.5 (4.4)	−19.2 to -1.8
Past-month cannabis use	−20.6 (23.6)	−66.9 to 25.7

Pre-Post Changes in Preliminary Outcome

Descriptively, ratings of intention (change between follow-up and baseline (Δ_(*fu–bl*)=0.4) and importance (Δ_(*fu–bl*))=0.3) to reduce alcohol use increased slightly from baseline to 1-month follow-up, but confidence in the ability to reduce use did not increase (Δ_(*fu–bl*)=−0.1); in contrast, intentions (Δ_(*fu–bl*) =−0.4) and importance (Δ_(*fu–bl*)=−0.1) to reduce cannabis use decreased slightly from baseline to follow-up and confidence (Δ_(*fu–bl*)=0.3) increased. Regarding behaviors, all variables decreased from baseline to 1-month follow-up, including weekly alcohol (Δ_(*fu–bl*)=−0.8) and cannabis use (Δ_(*fu–bl*)=−0.2), monthly binge-drinking days (Δ_(*fu–bl*)=−0.4), number of past-month substance use consequences (Δ_(*fu–bl*)=−0.8), frequency of past-month drinking and driving (Δ_(*fu–bl*)=−0.8), and frequency of past-month cannabis use and driving (Δ_(*fu–bl*)=−0.9; Table 5). Finally, more self-reported app engagement was associated with decreases from baseline to follow-up in past-month behaviors, including total number of drinks (*r*=0.4), total number of days using cannabis (*r*=−0.1), total binge-drinking days (*r*=−0.5), substance use consequences (*r*=−0.1), episodes of drinking and driving (*r*=−0.2), and using cannabis and driving (*r*=−0.2).

Table 5. Baseline and follow-up substance use–related measures.

Variables	Baseline	Follow-up
Motivation[a], confidence[a], and perceived risk, mean (SD)		
Intention to reduce alcohol	2.4 (2.3)	2.8 (2.5)
Importance of reducing alcohol	2.3 (1.9)	2.7 (1.5)
Confidence in ability to reduce alcohol use	8.3 (2.2)	8.2 (2.2)
Intention to reducing cannabis use	3.2 (3.3)	2.8 (3.1)
Importance of reducing cannabis use	2.4 (2.5)	2.3 (2.4)
Confidence in ability to reduce cannabis use	8.6 (2.1)	8.9 (2.0)
Perceived risk of regular alcohol use	2.5 (0.9)	2.5 (1.0)
Perceived risk of regular cannabis use	1.3 (0.9)	1.3 (1.0)
Substance use and consequences, mean (SD)		
Weekly number of alcoholic drinks	7.5 (6.1)	6.7 (6.4)
Weekly cannabis consumption (grams)	0.5 (1.0)	0.3 (0.7)
Number of binge-drinking episodes (past-month)	3.5 (3.3)	3.1 (3.2)
Substance use consequences (past-month)	8.5 (6.7)	7.7 (6.8)
Drinking and driving events (past-month)	2.5 (2.8)	1.7 (2.6)
Cannabis and driving events (past-month)	1.6 (3.4)	0.7 (1.3)

[a]n=38 for the alcohol importance ruler and n=18 for the cannabis importance ruler. Those who stopped drinking (n=2) or using cannabis (n=1) at follow-up and therefore did not complete ratings of intention, importance, and confidence of change at follow-up were recorded as 10 for each variable.

Discussion

Principal Findings

Here, we provide preliminary data assessing the feasibility and acceptability of the MiSARA app for addressing a significant public health problem, risky drinking, and cannabis use among adolescents and emerging adults. The MiSARA development process was informed by behavioral theories for enhancing engagement and reducing substance misuse as well as prior work [47-49] on the assessment-focused SARA app.

In terms of feasibility, web-based recruitment using social media advertisements resulted in the recruitment of 39 participants over 2 weeks to test MiSARA, underscoring the feasibility of such methods. Although representativeness in a small pilot study cannot be determined, other larger studies support the utility of web-based recruitment approaches [77]. Perhaps because of the limited funding and constrained timeline, only 55% (39/71) of those screening positive enrolled. Those who enrolled were more likely to be female and to use cannabis less frequently than those who did not enroll. In future studies, more staffing resources, extending the completion of web-based enrollment

activities beyond 2 weeks and targeting web-based recruitment toward males and those who use cannabis frequently, will likely increase enrollment rates and help ensure representativeness of the sample.

Overall, MiSARA was well-received, easy to use, and preferred over other research data collection methods (eg, website, text, telephone), with 9 out of 10 adolescents and emerging adults liking the aquarium app design. Participants were not concerned about privacy, with 9 out of 10 feeling comfortable with the self-report and 7 out of 10 being comfortable with passive data collection. Unsurprisingly, given the research purpose, ratings of the app in terms of *fun* were lower (eg, 69.2%), potentially reflecting users' expectations regarding app-based games, which may be difficult for researchers to achieve given limited budgets and may suggest that gamification is not well-suited for research software. Another challenge for research, particularly this pilot study, was reflected in the frequency of technical software issues, including the fact that daily and weekly survey data collection was not archived. This underscores the need for strong partnerships with software developers, in addition to adequate funding for research app development and ongoing support. Despite these challenges, feasibility and acceptability data were collected for the intervention messages, which was the primary purpose of the study, given prior studies examining engagement (Nahum-Shani et al, unpublished data, 2021). Participants liked the intervention messages and viewed them as appropriate for their age. Particularly helpful feedback included the recommendation that intervention content be archived by topic so that participants could refer to it later and an additional survey completion reminder part-way through the survey window. Although participants liked the theme of the aquarium, the feedback graphs, and game-like tasks, they consistently recommended greater interactivity and personalization, which was constrained by the modest resources, consistent with formative research, in the current iteration of MiSARA. Importantly, participants wanted interactivity to evolve over time to increase novelty, raising the question of how much interactivity is needed, balanced with the costs and skills needed to develop highly interactive apps, to engender adequate engagement.

The recommendation regarding customizing the survey period underscores an interesting challenge for daily monitoring and JITAIs, namely, when to assess and deliver messages each day. In the prior version of SARA, the period of assessment was later in the day, overlapping with the time of day when participants were likely to use substances (eg, evenings). In MiSARA, the survey period was the middle of the day, with intervention messages in the evening to increase relevance to potential use periods. Another alternative would be to have a morning assessment period, assessing substance use–related factors and/or use from the prior day, followed by delivering intervention content earlier in the day. Some of these decisions will likely be influenced by the type of questions asked, the focus of the JITAI (eg, mood, substance use), and the within-day or day of the week variability in the construct (eg, binge drinking on weekend evenings).

Regarding behavior change over time, all substance use–related behaviors decreased from baseline to 1-month follow-up. Furthermore, greater app usage (as measured by self-reported points earned in the app aquarium environment) was associated with less self-reported substance use, including fewer days drinking alcohol, binge drinking, and using cannabis and fewer consequences of use and episodes of driving after drinking or using cannabis. Although these data support the potential of the MiSARA app and the motivational intervention framework, the design and scope of this single-arm pilot study, in which all participants received some intervention content (0.33 probability), precluded causal conclusions. Alternatively, these findings may reflect self-selection bias, as people motivated to make changes to their use may have used the app more often. Indeed, adolescents and emerging adults who reported using the app more had lower initial alcohol use, potentially indicating a need to broaden app features to facilitate engagement (eg, incentives, inspirational quotes) among those who report higher initial alcohol use. Regardless, these preliminary findings justify further refinement and testing of MiSARA in a fully powered trial.

Future Research

A key future direction is the integration of passive data collection methods with self-reporting to optimize future JITAIs. For example, a recent study used passive cell phone data among 30 emerging adults over 28 days to develop a machine learning algorithm to predict nondrinking, drinking, and heavy drinking episodes (>95% accuracy), with significant predictors including features such as device usage (eg, longer keyboard press time) and movement features (eg, more changes in activities) [78,79]. The most robust predictor of drinking was time of the day (eg, evening), with the day of the week also being predictive, underscoring the timing of alcohol intervention content (eg, weekends). However, our prior studies with SARA suggest that such timing is not as useful for cannabis, which is just as likely on weekdays as weekends. Next, the Bae et al [78] study focused on identifying when drinking was already in progress, as opposed to this study, which conceptually focused on intervening before a drinking episode commences. Challenges for passive data collection include the acceptability of passive mining of cell phone data among adolescents and emerging adults (eg, in a study by Bae et al [78], 16% disabled the sensors) and feasibility across iPhone and Android devices, where privacy restrictions limit data accessibility, coupled with concerns about increased mobile phone battery use. Importantly, passive data collection is not possible to measure some important constructs that meaningfully inform tailoring of JITAIs, such as motives that are associated with use [80]. For example, Shrier et al [81] found that negative affect the day before cannabis use was a marker of next day use, as were motives. Thus, future JITAIs could integrate both passive and self-report data.

Another key direction for JITAIs is related to the timing of assessment and intervention delivery. In an ideal world, there would be an assessment of substance use patterns at the daily level (eg, mornings, afternoons, evenings, on weekdays and weekends), with assessment and intervention timing personalized over time using machine learning. However, this assessment and intervention delivery complexity requires sophisticated design and analysis methodology [82] to keep pace with the development of mHealth JITAIs.

In terms of clinical implications, if demonstrated as efficacious, the MiSARA app could be provided to youth as a universal prevention strategy, such as during a health care visit, or as part of prevention efforts at high schools or universities. Alternatively, as a selective prevention approach to alter risk trajectories and prevent the development of SUDs, the MiSARA app could be provided to adolescents and emerging adults screening positive for binge drinking or cannabis use as part of interactions in academic, primary care, or other community settings. Although our purpose was to develop an app to address primary and secondary prevention, MiSARA could also be used as an adjunct in treatment programs to provide clinicians with clinically relevant information between sessions.

Limitations

The findings are limited by the formative nature (eg, small sample size) of this pilot study, which focused on feasibility and acceptability. Replication is required with a larger sample to determine the efficacy of intervention content on proximal substance use (eg, same day, next day, next week) and consequences from use. The limited funding for app development resulted in programming bugs, which precluded archiving of the daily data in our initial MiSARA prototype, highlighting the challenge of obtaining sufficient funding for mHealth intervention development. Thus, self-report data regarding app use (points earned) could not be verified and, despite this being a proxy for dosage, is subject to errors in recall, underscoring the need for additional testing of a refined app. Given the increasing recreational cannabis legalization, future testing should examine the efficacy of JITAIs for cannabis in states with and without such laws (eg, particularly for ages 21 years and older, who have legal access to purchase cannabis). Finally, the pilot nature of this study precludes causal determination of JITAI efficacy and necessitates future research testing of MiSARA compared with a control condition receiving only daily surveys. Nonetheless, the data collected provide clues for subsequent refinement of MiSARA and underscore the need for careful development of mHealth JITAIs.

Conclusions

The development of real-time, real-world, personalized, and scalable JITAI approaches to reduce substance use among adolescents and emerging adults is a growing area of mHealth research. The MiSARA app is promising, as most users liked the app experience and found the content to be highly appropriate. Adolescents and emerging adult users were comfortable providing personal information on this platform and preferred it to other data collection modalities, with future refinements focusing on improving interactivity and technical functionality to sustain interest over time. Although future studies are needed to examine the efficacy of personalized JITAI content in reducing risky substance use among adolescents and emerging adults who are not in treatment, formative data showed decreases in substance use over time in the single-arm pilot study, which along with acceptability data support future testing in a full-scale trial. Consistent with the MOST framework, recommendations for JITAI development also include the use of an iterative feedback process across a series of studies to optimize efficacy.

Acknowledgments

This study was supported by a grant from the Michigan Institute for Clinical and Health Research (MICHR #003558). LC was supported by the National Institute of Alcoholism and Alcohol Abuse (NIAAA T32AA007477 and K23AA028232).

Conflicts of Interest

MW is a minor shareholder in Facebook and has a conflict of interest plan approved by the University of Michigan.

References

1. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
2. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34 Suppl:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
3. Mirowski M, Reid PR, Mower MM, Watkins L, Gott VL, Schauble JF, et al. Termination of malignant ventricular arrhythmias with an implanted automatic defibrillator in human beings. N Engl J Med 1980 Aug 07;303(6):322-324. [doi: 10.1056/NEJM198008073030607] [Medline: 6991948]
4. Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group, Tamborlane WV, Beck RW, Bode BW, Buckingham B, Chase HP, et al. Continuous glucose monitoring and intensive treatment of type 1 diabetes. N Engl J Med 2008 Oct 02;359(14):1464-1476. [doi: 10.1056/NEJMoa0805017] [Medline: 18779236]
5. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
6. Riley W, Obermayer J, Jean-Mary J. Internet and mobile phone text messaging intervention for college smokers. J Am Coll Health 2008;57(2):245-248. [doi: 10.3200/JACH.57.2.245-248] [Medline: 18809542]

7. 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]
8. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
9. Arnett JJ. The developmental context of substance use in emerging adulthood. Journal Drug Issues 2005 Apr 1;35(2):235-254. [doi: 10.1177/002204260503500202]
10. Kuntsche E, Kuntsche S, Thrul J, Gmel G. Binge drinking: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):976-1017. [doi: 10.1080/08870446.2017.1325889] [Medline: 28513195]
11. Substance Abuse and Mental Health Service Administration. Results from the 2016 National Survey on Drug Use and Health: Detailed Tables. 2016. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.htm [accessed 2020-06-15]
12. Behrendt S, Wittchen H, Höfler M, Lieb R, Beesdo K. Transitions from first substance use to substance use disorders in adolescence: is early onset associated with a rapid escalation? Drug Alcohol Depend 2009 Jan 01;99(1-3):68-78. [doi: 10.1016/j.drugalcdep.2008.06.014] [Medline: 18768267]
13. White A, Hingson R. The burden of alcohol use: excessive alcohol consumption and related consequences among college students. Alcohol Res 2013;35(2):201-218 [FREE Full text] [Medline: 24881329]
14. Fiellin LE, Tetrault JM, Becker WC, Fiellin DA, Hoff RA. Previous use of alcohol, cigarettes, and marijuana and subsequent abuse of prescription opioids in young adults. J Adolesc Health 2013 Feb;52(2):158-163 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.010] [Medline: 23332479]
15. Stockwell T, Zhao J, Macdonald S. Who under-reports their alcohol consumption in telephone surveys and by how much? An application of the 'yesterday method' in a national Canadian substance use survey. Addiction 2014 Oct;109(10):1657-1666. [doi: 10.1111/add.12609] [Medline: 24825591]
16. Bose J, Hedden S, Lipari R, Park-Lee E. Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.htm [accessed 2020-10-15]
17. Sarvet AL, Wall MM, Keyes KM, Cerdá M, Schulenberg JE, O'Malley PM, et al. Recent rapid decrease in adolescents' perception that marijuana is harmful, but no concurrent increase in use. Drug Alcohol Depend 2018 May 01;186:68-74 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.12.041] [Medline: 29550624]
18. Miech R, Johnston L, O'Malley PM. Prevalence and attitudes regarding marijuana use among adolescents over the past decade. Pediatrics 2017 Dec;140(6) [FREE Full text] [doi: 10.1542/peds.2017-0982] [Medline: 29109106]
19. Gruber SA, Sagar KA, Dahlgren MK, Racine M, Lukas SE. Age of onset of marijuana use and executive function. Psychol Addict Behav 2012 Sep;26(3):496-506 [FREE Full text] [doi: 10.1037/a0026269] [Medline: 22103843]
20. Gruber SA, Dahlgren MK, Sagar KA, Gönenc A, Killgore WDS. Age of onset of marijuana use impacts inhibitory processing. Neurosci Lett 2012 Mar 09;511(2):89-94 [FREE Full text] [doi: 10.1016/j.neulet.2012.01.039] [Medline: 22306089]
21. Patrick ME, Kloska DD, Terry-McElrath YM, Lee CM, O'Malley PM, Johnston LD. Patterns of simultaneous and concurrent alcohol and marijuana use among adolescents. Am J Drug Alcohol Abuse 2018;44(4):441-451 [FREE Full text] [doi: 10.1080/00952990.2017.1402335] [Medline: 29261344]
22. Patrick ME, Terry-McElrath YM, Lee CM, Schulenberg JE. Simultaneous alcohol and marijuana use among underage young adults in the United States. Addict Behav 2019 Jan;88:77-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.08.015] [Medline: 30170141]
23. Patrick ME, Fairlie AM, Cadigan JM, Abdallah DA, Larimer ME, Lee CM. Daily motives for alcohol and marijuana use as predictors of simultaneous use among young adults. J Stud Alcohol Drugs 2019 Jul;80(4):454-461 [FREE Full text] [Medline: 31495383]
24. Patrick ME, Fairlie AM, Lee CM. Motives for simultaneous alcohol and marijuana use among young adults. Addict Behav 2018 Jan;76:363-369 [FREE Full text] [doi: 10.1016/j.addbeh.2017.08.027] [Medline: 28915500]
25. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (NSDUH). Substance Abuse and Mental Health Service Administration. Rockville: Center for Behavioral Health Statistics and Quality; 2019. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2020-10-15]
26. First Evidence-Based Platform Automating Addiction Management Life Cycle. CHESS Health. URL: www.chess.health/ [accessed 2020-09-25]
27. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
28. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]

29. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
30. Gonzales R, Hernandez M, Murphy DA, Ang A. Youth recovery outcomes at 6 and 9 months following participation in a mobile texting recovery support aftercare pilot study. Am J Addict 2016 Jan;25(1):62-68 [FREE Full text] [doi: 10.1111/ajad.12322] [Medline: 26689171]
31. Scott CK, Dennis ML, Johnson KA, Grella CE. A randomized clinical trial of smartphone self-managed recovery support services. J Subst Abuse Treat 2020 Oct;117:108089. [doi: 10.1016/j.jsat.2020.108089] [Medline: 32811628]
32. Dennis ML, Scott CK, Funk RR, Nicholson L. A pilot study to examine the feasibility and potential effectiveness of using smartphones to provide recovery support for adolescents. Subst Abus 2015;36(4):486-492 [FREE Full text] [doi: 10.1080/08897077.2014.970323] [Medline: 25310057]
33. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
34. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
35. Wray TB, Merrill JE, Monti PM. Using Ecological Momentary Assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. Alcohol Res 2014;36(1):19-27 [FREE Full text] [Medline: 26258997]
36. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015 Mar 1;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
37. Buckner JD, Crosby RD, Silgado J, Wonderlich SA, Schmidt NB. Immediate antecedents of marijuana use: an analysis from ecological momentary assessment. J Behav Ther Exp Psychiatry 2012 Mar;43(1):647-655 [FREE Full text] [doi: 10.1016/j.jbtep.2011.09.010] [Medline: 21946296]
38. Buckner JD, Crosby RD, Wonderlich SA, Schmidt NB. Social anxiety and cannabis use: an analysis from ecological momentary assessment. J Anxiety Disord 2012 Mar;26(2):297-304 [FREE Full text] [doi: 10.1016/j.janxdis.2011.12.006] [Medline: 22246109]
39. Piasecki TM, Cooper ML, Wood PK, Sher KJ, Shiffman S, Heath AC. Dispositional drinking motives: associations with appraised alcohol effects and alcohol consumption in an ecological momentary assessment investigation. Psychol Assess 2014 Jun;26(2):363-369 [FREE Full text] [doi: 10.1037/a0035153] [Medline: 24274049]
40. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
41. Saberi P, Johnson MO. Correlation of internet use for health care engagement purposes and HIV clinical outcomes among HIV-positive individuals using online social media. J Health Commun 2015;20(9):1026-1032 [FREE Full text] [doi: 10.1080/10810730.2015.1018617] [Medline: 26120890]
42. Dobson R, Whittaker R, Murphy R, Khanolkar M, Miller S, Naylor J, et al. The use of mobile health to deliver self-management support to young people With type 1 diabetes: a cross-sectional survey. JMIR Diabetes 2017 Feb 15;2(1):e4 [FREE Full text] [doi: 10.2196/diabetes.7221] [Medline: 30291057]
43. Singh NB, Björling EA. A review of EMA assessment period reporting for mood variables in substance use research: expanding existing EMA guidelines. Addict Behav 2019 Jul;94:133-146. [doi: 10.1016/j.addbeh.2019.01.033] [Medline: 30712774]
44. Gajecki M, Andersson C, Rosendahl I, Sinadinovic K, Fredriksson M, Berman AH. Skills training via smartphone app for university students with excessive alcohol consumption: a randomized controlled trial. Int J Behav Med 2017 Dec;24(5):778-788 [FREE Full text] [doi: 10.1007/s12529-016-9629-9] [Medline: 28224445]
45. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619. [doi: 10.1111/add.14503] [Medline: 30461120]
46. Collins LM, Kugler KC, editors. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics. Switzerland: Springer International Publishing; 2018.
47. Rabbi M, Philyaw-Kotov M, Lee J, Mansour A, Dent L, Wang X, et al. SARA: a mobile app to engage users in health data collection. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:781-789 [FREE Full text] [doi: 10.1145/3123024.3125611] [Medline: 29503985]
48. Rabbi M, Philyaw Kotov M, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, et al. Toward increasing engagement in substance use data collection: development of the Substance Abuse Research Assistant app and protocol for a microrandomized trial using adolescents and emerging adults. JMIR Res Protoc 2018 Jul 18;7(7):e166. [doi: 10.2196/resprot.9850] [Medline: 30021714]
49. Rabbi M, Philyaw-Kotov M, Li J. Translating Behavioral Theory into Technological Interventions: Case Study of an mHealth App to Increase Self-reporting of Substance-Use Related Data. Cornell University. 2020 Mar 30. URL: https://arxiv.org/abs/2003.13545 [accessed 2020-12-20]

50. Carolina Population Center. URL: http://www.cpc.unc.edu/projects/addhealth [accessed 2020-12-20]
51. Youth Risk Behavior Survey. Centers for Disease Control and Prevention. 2005. URL: https://www.cdc.gov/healthyyouth/data/yrbs/index.htm [accessed 2020-12-21]
52. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
53. Humeniuk R, Henry-Edwards S, Ali R, Poznyak V, Monteiro M. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Manual for Use in Primary Care. Geneva: World Health Organization; 2010.
54. McNeely J, Wu L, Subramaniam G, Sharma G, Cathers LA, Svikis D, et al. Performance of the tobacco, alcohol, prescription medication, and other substance use (taps) tool for substance use screening in primary care patients. Ann Intern Med 2016 Nov 15;165(10):690-699. [doi: 10.7326/M16-0317] [Medline: 27595276]
55. Ali R, Meena S, Eastwood B, Richards I, Marsden J. Ultra-rapid screening for substance-use disorders: the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST-Lite). Drug Alcohol Depend 2013 Sep 01;132(1-2):352-361. [doi: 10.1016/j.drugalcdep.2013.03.001] [Medline: 23561823]
56. Sobell L, Sobell M. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Measuring Alcohol Consumption. Totowa, NJ: Humana Press; 1992.
57. Kahler CW, Hustad J, Barnett NP, Strong DR, Borsari B. Validation of the 30-day version of the brief young adult alcohol consequences questionnaire for use in longitudinal studies. J Stud Alcohol Drugs 2008 Jul;69(4):611-615 [FREE Full text] [Medline: 18612578]
58. Simons JS, Dvorak RD, Merrill JE, Read JP. Dimensions and severity of marijuana consequences: development and validation of the Marijuana Consequences Questionnaire (MACQ). Addict Behav 2012 May;37(5):613-621 [FREE Full text] [doi: 10.1016/j.addbeh.2012.01.008] [Medline: 22305645]
59. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
60. Breslin FC, Sobell LC, Sobell MB, Agrawal S. A comparison of a brief and long version of the Situational Confidence Questionnaire. Behav Res Ther 2000 Dec;38(12):1211-1220. [doi: 10.1016/s0005-7967(99)00152-7] [Medline: 11104185]
61. Johnston L, O'Malley P, Bachman J, Schulenberg J. Monitoring the Future National Survey Results on Drug Use-2005. Volume II. National Institute on Drug Abuse. Bethesda, MD; 2006. URL: https://files.eric.ed.gov/fulltext/ED494061.pdf [accessed 2020-12-25]
62. Simons J, Correia CJ, Carey KB, Borsari BE. Validating a five-factor marijuana motives measure: relations with use, problems, and alcohol motives. J Counsel Psychol 1998;45(3):265-273. [doi: 10.1037/0022-0167.45.3.265]
63. Grant VV, Stewart SH, O'Connor RM, Blackwell E, Conrod PJ. Psychometric evaluation of the five-factor Modified Drinking Motives Questionnaire--revised in undergraduates. Addict Behav 2007 Nov;32(11):2611-2632. [doi: 10.1016/j.addbeh.2007.07.004] [Medline: 17716823]
64. Donovan JE. Young adult drinking-driving: behavioral and psychosocial correlates. J Stud Alcohol 1993 Sep;54(5):600-613. [doi: 10.15288/jsa.1993.54.600] [Medline: 8412150]
65. Pollak J, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: SIGCHI conference on Human factors in computing systems; May 2011; Vancouver URL: https://dl.acm.org/doi/10.1145/1978942.1979047 [doi: 10.1145/1978942.1979047]
66. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [doi: 10.1037//0022-3514.39.3.472] [Medline: 7431205]
67. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Pers Soc Psychol Bull 2006 Jul;32(7):917-929 [FREE Full text] [doi: 10.1177/0146167206287721] [Medline: 16738025]
68. Ramirez-Valles J, Zimmerman MA, Newcomb MD. Sexual risk behavior among youth: modeling the influence of prosocial activities and socioeconomic factors. J Health Soc Behav 1998 Sep;39(3):237-253. [Medline: 9785696]
69. Hoyle RH, Stephenson MT, Palmgreen P, Lorch EP, Donohew R. Reliability and validity of a brief measure of sensation seeking. Pers Individ Dif 2002 Feb;32(3):401-414. [doi: 10.1016/s0191-8869(01)00032-0]
70. Lippman L, Moore K, Guzman L. Flourishing Children: Defining and Testing Indicators of Positive Development. Netherlands: Springer; 2014.
71. Ellickson P, Bell R. Prospects for Preventing Drug Use Among Young Adolescents. Santa Monica, CA: RAND Corp; 1990.
72. Stephens RS, Babor TF, Kadden R, Miller M, Marijuana Treatment Project Research Group. The Marijuana Treatment Project: rationale, design and participant characteristics. Addiction 2002 Dec;97 Suppl 1:109-124. [doi: 10.1046/j.1360-0443.97.s01.6.x] [Medline: 12460133]
73. Falk A, Fischbacher U. A theory of reciprocity. Games Econ Behav 2006 Feb;54(2):293-315. [doi: 10.1016/j.geb.2005.03.001]

74. Surma J. Social exchange in online social networks. The reciprocity phenomenon on Facebook. Comput Commun 2016 Jan;73:342-346. [doi: 10.1016/j.comcom.2015.06.017]
75. Cropanzano R, Mitchell MS. Social exchange theory: an interdisciplinary review. J Manag 2016 Jul;31(6):874-900. [doi: 10.1177/0149206305279602]
76. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
77. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838 [FREE Full text] [doi: 10.15288/jsad.2012.73.834] [Medline: 22846248]
78. Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun 30;1(2):1-36. [doi: 10.1145/3090051]
79. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2017 Nov 27. [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
80. Bonar EE, Goldstick JE, Collins RL, Cranford JA, Cunningham RM, Chermack ST, et al. Daily associations between cannabis motives and consumption in emerging adults. Drug Alcohol Depend 2017 Sep 01;178:136-142 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.006] [Medline: 28647681]
81. Shrier LA, Ross CS, Blood EA. Momentary positive and negative affect preceding marijuana use events in youth. J Stud Alcohol Drugs 2014 Sep;75(5):781-789 [FREE Full text] [doi: 10.15288/jsad.2014.75.781] [Medline: 25208196]
82. Riley WT, Oh A, Aklin WM, Wolff-Hughes DL. National Institutes of Health support of digital health behavior research. Health Educ Behav 2019 Dec;46(2_suppl):12-19. [doi: 10.1177/1090198119866644] [Medline: 31742453]

Abbreviations

AUDIT-C: Alcohol Use Disorder Identification Test-Consumption
EMA: ecological momentary assessment
JITAI: just-in-time adaptive intervention
SARA: substance abuse research assistant
SUD: substance use disorders
mHealth: mobile health
MOST: Multiphase Optimization Strategy

Edited by L Buis; submitted 18.09.20; peer-reviewed by D Litt, MH Yong; comments to author 22.10.20; revised version received 11.11.20; accepted 30.11.20; published 15.01.21.

Please cite as:
Coughlin LN, Nahum-Shani I, Philyaw-Kotov ML, Bonar EE, Rabbi M, Klasnja P, Murphy S, Walton MA
Developing an Adaptive Mobile Intervention to Address Risky Substance Use Among Adolescents and Emerging Adults: Usability Study
JMIR Mhealth Uhealth 2021;9(1):e24424
URL: http://mhealth.jmir.org/2021/1/e24424/
doi:10.2196/24424
PMID:33448931

Original Paper

Habits Heart App for Patient Engagement in Heart Failure Management: Pilot Feasibility Randomized Trial

Kevin S Wei[1,2], BSc; Nasrien E Ibrahim[1,3], MD; Ashok A Kumar[4], PhD; Sidhant Jena[4], MBA; Veronica Chew[4], MBA; Michal Depa[4], MSc; Namrata Mayanil[4], MSc; Joseph C Kvedar[3,5], MD; Hanna K Gaggin[1,3], MD, MPH

[1]Cardiology Division, Massachusetts General Hospital, Boston, MA, United States

[2]University of California, Irvine, School of Medicine, Irvine, CA, United States

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

[4]Jana Care, Boston, MA, United States

[5]Department of Dermatology, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Hanna K Gaggin, MD, MPH
Cardiology Division
Massachusetts General Hospital
55 Fruit Street, Yawkey 5B
Boston, MA, 02114
United States
Phone: 1 617 726 2709
Email: HGAGGIN@mgh.harvard.edu

Abstract

Background: Due to the complexity and chronicity of heart failure, engaging yet simple patient self-management tools are needed.

Objective: This study aimed to assess the feasibility and patient engagement with a smartphone app designed for heart failure.

Methods: Patients with heart failure were randomized to intervention (smartphone with the Habits Heart App installed and Bluetooth-linked scale) or control (paper education material) groups. All intervention group patients were interviewed and monitored closely for app feasibility while receiving standard of care heart failure management by cardiologists. The Atlanta Heart Failure Knowledge Test, a quality of life survey (Kansas City Cardiomyopathy Questionnaire), and weight were assessed at baseline and final visits.

Results: Patients (N=28 patients; intervention: n=15; control: n=13) with heart failure (with reduced ejection fraction: 15/28, 54%; male: 20/28, 71%, female: 8/28, 29%; median age 63 years) were enrolled, and 82% of patients (N=23; intervention: 12/15, 80%; control: 11/13, 85%) completed both baseline and final visits (median follow up 60 days). In the intervention group, 2 out of the 12 patients who completed the study did not use the app after study onboarding due to illnesses and hospitalizations. Of the remaining 10 patients who used the app, 5 patients logged ≥1 interaction with the app per day on average, and 2 patients logged an interaction with the app every other day on average. The intervention group averaged 403 screen views (per patient) in 56 distinct sessions, 5-minute session durations, and 22 weight entries per patient. There was a direct correlation between duration of app use and improvement in heart failure knowledge (Atlanta Heart Failure Knowledge Test score; ρ=0.59, *P*=.04) and quality of life (Kansas City Cardiomyopathy Questionnaire score; ρ=0.63, *P*=.03). The correlation between app use and weight change was ρ=–0.40 (*P*=.19). Only 1 out of 11 patients in the control group retained education material by the follow-up visit.

Conclusions: The Habits Heart App with a Bluetooth-linked scale is a feasible way to engage patients in heart failure management, and barriers to app engagement were identified. A larger multicenter study may be warranted to evaluate the effectiveness of the app.

Trial Registration: ClinicalTrials.gov NCT03238729; http://clinicaltrials.gov/ct2/show/NCT03238729

(JMIR Mhealth Uhealth 2021;9(1):e19465) doi:10.2196/19465

KEYWORDS

heart failure; smartphone application; heart failure management

Introduction

The COVID-19 pandemic has disrupted chronic disease management [1]. In particular, patients with heart failure are at high risk of complications from disruption of care. Heart failure management is complex; it requires taking multiple medications, a close monitoring of total daily fluid and sodium intake, and monitoring for volume and symptoms. Clinical changes may warrant timely intervention in medication regimen to avoid life-threatening complications. Despite the importance of adherence to medical recommendations, in patients living with heart failure, patient engagement, adherence, and self-management are suboptimal and may limit the benefits of heart failure recommendations and existing treatments [2-4]. Smartphones and mobile health apps are ubiquitous even in older adult populations [5,6]. Smartphones and mobile health apps have the potential for bettering patient engagement and self-management. While there are a number of commercially available smartphone heart failure apps, few have been designed by clinicians or specifically designed to address heart failure self-management; only 41% had weight management features [7]. Even fewer have published their methodology and feasibility studies [8,9] in peer-reviewed journals.

In this context, we employed a heart failure–specific smartphone-based app intervention with content designed by clinicians with the following features: daily to-do lists for heart failure–related activities including patient education, symptom monitoring, wireless weight tracking, sodium intake tracking, exercise logs, and action prompts based on weight and symptom severity. There was also a messaging platform for patients to reach out to study staff and physicians. Our goal was to test feasibility and usability in this pilot randomized proof-of-concept study.

Methods

All study procedures were approved by the Partners Healthcare Institutional Review Board and carried out in accordance with the Declaration of Helsinki, and the study was registered (NCT03238729). The Habits Heart App was developed in conjunction with cardiologists who care for patients with heart failure at Massachusetts General Hospital and software engineers at Jana Care. Cardiologists designed the main concepts (daily To-Do list, Learn, Track, Coach) and content in writing and worked with Jana Care to develop the app. While the main goal of Track was to monitor weight changes, tracking sodium content (in diet) and exercising logging were also included, in an exploratory fashion, using a publicly available database of exercises and a database from University of Minnesota of the sodium content in food items. The Habits Heart App was downloaded onto a study Android smartphone.

Patients with heart failure aged ≥18 years with smartphones were recruited from cardiology clinics and inpatient units at Massachusetts General Hospital. Heart failure diagnosis by a cardiologist in the electronic medical records with ≥1 hospitalization or emergency department visit for acute heart failure or any outpatient escalation of diuretic therapy in the past 12 months was required. *Heart failure with reduced ejection fraction* was defined as symptomatic heart failure with a left ventricular ejection fraction ≤40%, and *heart failure with preserved ejection fraction* was defined as symptomatic heart failure with left ventricular ejection fraction ≥50% by any imaging modality. A randomized list with the order of intervention or control group allocations corresponding to the order of enrollment was generated at the start of the study. After informed consent was obtained, patients were then allocated to either the intervention group or the control group (Figure 1). Those randomized to the intervention group were given a 30-minute onboarding session to get acquainted with the Habits Heart App and the study smartphone. At the end of the 30-minute session, patients were asked to demonstrate their app use competency by performing actions in 5 key areas including opening and watching the introduction patient education video, logging weight on the app by using the Bluetooth-linked scale, inputting sodium content entries, inputting exercise entries, and messaging the study team through the app.

All patients received the standard of care for heart failure management by cardiologists. Patients in the control group received written heart failure education materials covering the same topics as the patient education portion of the app and were instructed to read the patient education materials until the final visit.

Intervention group patients were provided a smartphone with the Habits Heart App. The daily To-Do list prompted patients to watch the educational heart failure video of the day, exercise, weigh themselves, or complete other healthy activities in addition to 3 self-management functionalities: Track, Learn, and Coach (Figure 2). Track synced automatically with a Bluetooth-linked digital scale to record daily weight. Weight gain ≥3 lbs (1.4 kg) in 1 to 2 days or ≥5 lbs (2.3 kg) in a week triggered a symptoms survey. Depending on symptom severity, the app had an automated risk-based algorithm to instruct patients to contact study staff or contact their doctor (or dial 911, ie, emergency services, if after hours). If patients were instructed to call a physician or dial 911, the study team would follow up with patients by phone within 1 day if during the week or on Monday of the next week if during the weekend. Diet and physical activities were tracked in an exploratory fashion. Patients picked foods consumed at each meal from a list of available food items with corresponding sodium content, while patients entered length and mode of exercise performed each day. The exercise database had not previously been validated. Learn featured 13 interactive lessons including videos with topics recorded and narrated by cardiologists. Coach included a messaging platform where patients could communicate with the study team and cardiologist. All patients in the intervention group were contacted through the app messaging platform by the study staff within 1 week of starting the study to obtain feasibility feedback regarding the study app. The study coordinator tracked each patient's messages sent to

the study team along with weight, sodium content, and exercise entries through a secure online dashboard with deidentified patient data accessible only to the study staff. At the final study visit, all patients were interviewed for feasibility feedback.

Figure 1. Study flow diagram. AHFKT-V2: Atlanta HF Knowledge Test, Version 2; HF: heart failure; KCCQ-12: Kansas City Cardiomyopathy Questionnaire.

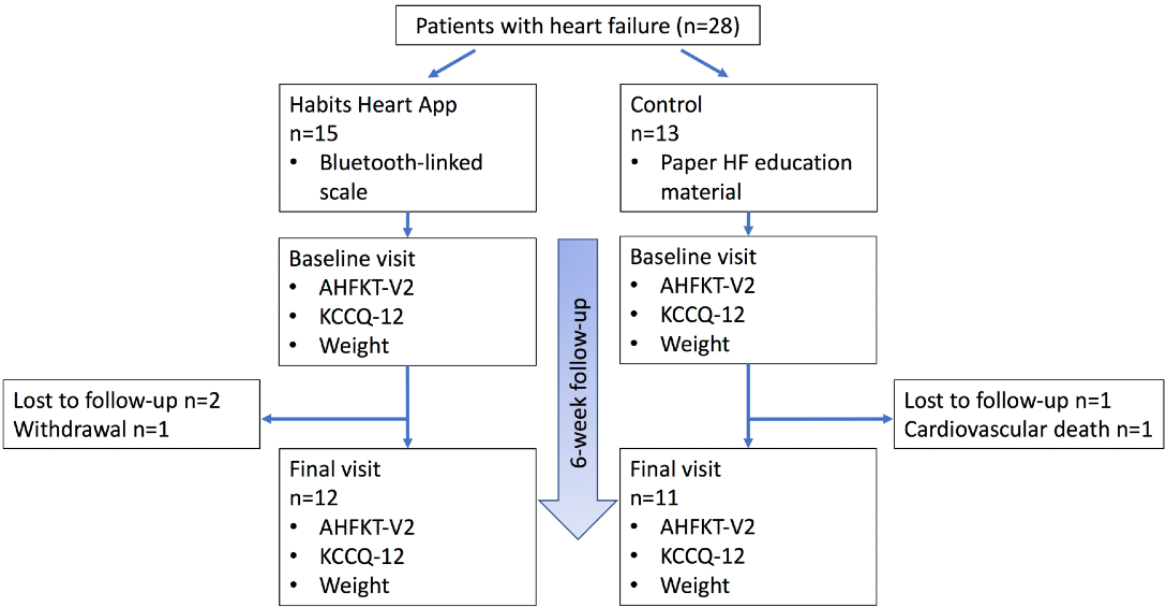
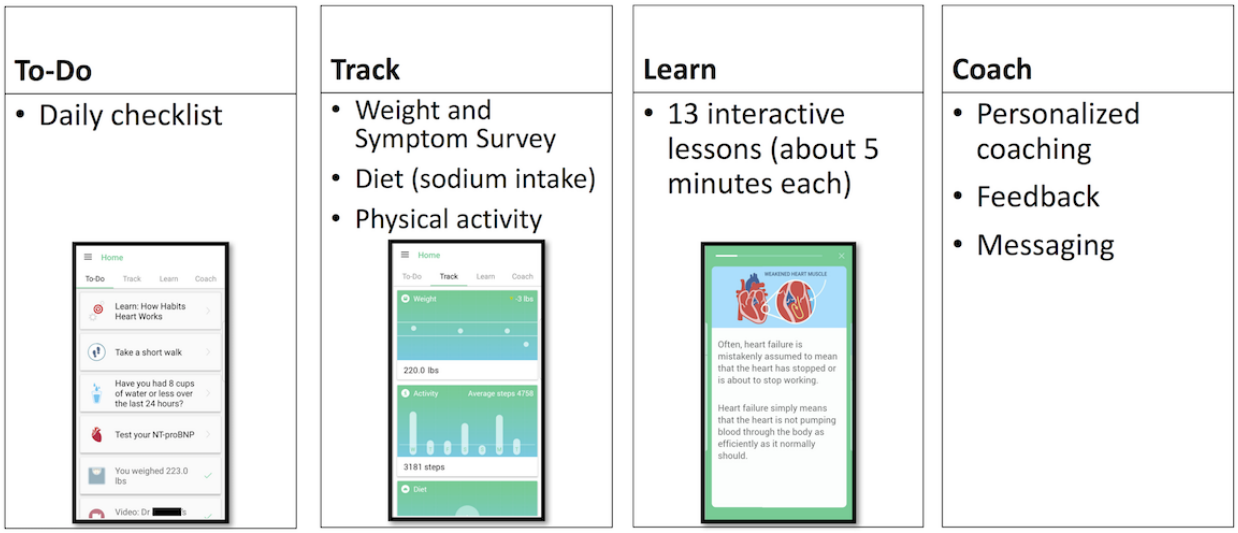


Figure 2. Layout of the Habits Heart App: To-Do List, Track, Learn, Coach.



Baseline characteristics were obtained from patient electronic medical records. At baseline and final visits (goal of at least 6 weeks in-between), the Atlanta Heart Failure Knowledge Test (AHFKT-V2) [10] and a quality of life survey (Kansas City Cardiomyopathy Questionnaire, KCCQ-12) [11] were filled out by patients along with recording of weight for both intervention and control groups. The KCCQ-12 scores were normalized to a 0-to-100-point scale. Higher AHFKT-V2 or KCCQ-12 scores indicated better heart failure-specific education or health status. Patients who did not have hospital follow-up visits already scheduled 6 weeks after study onboarding were asked to schedule an appointment for the earliest availability after 6 weeks. Intervention group patients were asked to return study smartphones and Bluetooth scales upon study completion. Control group patients were asked if they retained paper patient education material.

Spearman correlations were used to determine the relationships between the average session duration of app use and the AHFKT-V2 score, the KCCQ-12 score, and weight loss.

Results

Patient Characteristics

A total of 28 patients were enrolled in the study (intervention: n=15; control: n=13). Of the patients enrolled, 54% (15/28) had heart failure with reduced ejection fraction, 71% (20/28) were male, and the median age was 63 years. The median follow-up was 60 days (range 50-66 days); 80% of patients (12/15) in the

intervention and 85% of patients (11/13) in the control group completed both baseline and final visits (Table 1). In the intervention group, 2 patients were lost to follow-up, and 1 patient withdrew from the study. In the control group, 1 patient was lost to follow-up, and 1 patient died.

Table 1. Baseline characteristics of patients with heart failure.

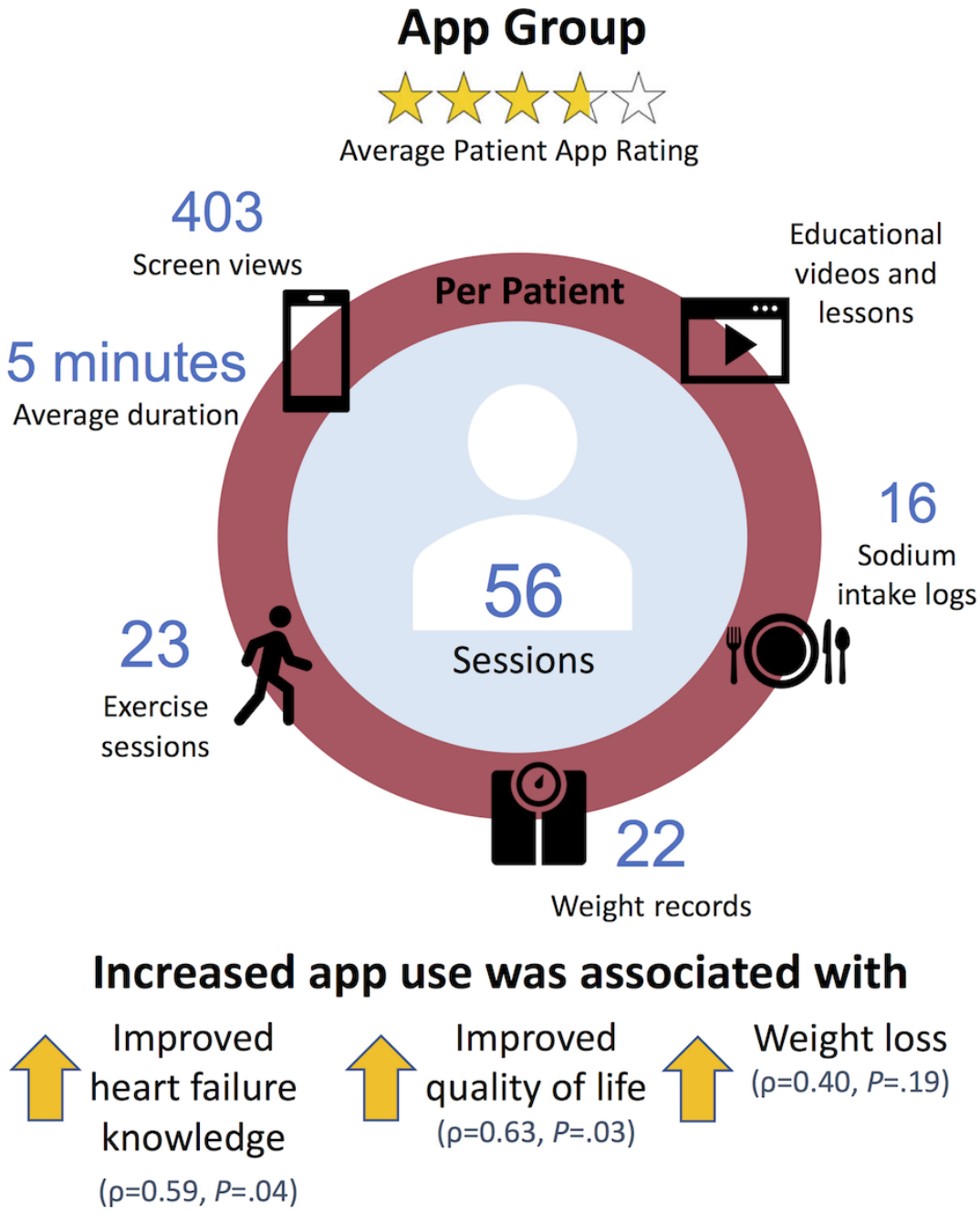
Characteristics	App (n=12)	Control (n=11)
Demographic		
Age (years), median	63	62
Sex, n (%)		
Male	9 (75)	7 (64)
Female	3 (25)	4 (36)
Race, n (%)		
African-American	2 (17)	0 (0)
Asian	1 (8)	0 (0)
Caucasian	8 (67)	11 (100)
Hispanic	1 (8)	0 (0)
Weight (kg), median	98.1	93.4
With reduced ejection fraction, n (%)	5 (42)	6 (55)
New York Heart Association class[a], n (%)		
Class I	3 (33)	1 (13)
Class II	6 (67)	3 (38)
Class III	0 (0)	3 (38)
Class IV	0 (0)	1 (13)
Medical history, n (%)		
Atrial fibrillation/flutter	8 (67)	2 (18)
Hypertension	12 (100)	9 (82)
Coronary artery disease	9 (75)	5 (45)
Chronic obstructive pulmonary disease	2 (17)	0 (0)
Type 1 or type 2 diabetes	6 (50)	5 (45)
Hyperlipidemia	9 (75)	8 (73)
Cerebrovascular accident/transient ischemic attack	4 (33)	0 (0)
Chronic kidney disease	6 (50)	3 (27)
Medication, n (%)		
Angiotensin converting enzyme inhibitor	3 (25)	3 (27)
Angiotensin II receptor blocker	3 (25)	4 (36)
Sacubitril/valsartan	4 (33)	1 (9)
Beta blocker	11 (92)	9 (82)
Mineralocorticoid receptor antagonist	5 (42)	3 (27)
Loop diuretics	12 (100)	10 (91)
Echocardiography results		
Left ventricular ejection fraction (%)	47.4	40.9
Biomarkers, median		
N-terminal pro–B-type natriuretic peptide (pg/mL)	669	1910

[a]App: n=9; control: n=8. Note: As a result of rounding, percentages may not add to 100.

Intervention Feasibility

All patients in the intervention group (15/15, 100%) were able to demonstrate their competency using the app by performing actions in 5 key areas at the end of the 30-minute onboarding session; however, 2 out of the 12 patients (17%) who completed both study visits were unable to use the app after study onboarding due to intermittent illnesses and hospitalizations. Of the 10 patients who were able to use the app, 7 patients (70%) logged ≥1 interaction with the app every other day on average throughout the study duration, and 5 patients (50%) logged ≥1 interaction with the app per day on average throughout the study duration. At the final visit, 9 out of 12 patients (75%) in the intervention group subjectively reported regularly interacting with the app (Figure 3).

Figure 3. App feasibility and patient rating.



There was an average of 403 screen views per person over 56 distinct sessions (duration per session, average: 5 minutes; documentation per person, average—weight: 22; sodium intake: 16; exercise: 23). The patients who reported using the app regularly (n=9) averaged 527 screen views per person over 73 distinct sessions (duration per session, average: 5 minutes; documentation per person, average—weight: 29; sodium intake: 21; exercise: 30).

As of the 6-week follow-up visit, 0 out of 12 patients had problems opening and watching education sessions on Learn. Only 1 out of 12 patients (8%) reported synching issues between the scale and the app, 1 out of 12 patients (8%) reported not

being able to send additional messages on the app messaging platform for several minutes after sending the first message, and 2 out of 12 patients (17%) had problems with the smartphone intermittently disconnecting from the internet. There were 28 app crashes among the entire group. Only 1 out of 11 patients (9%) in the control group retained reading materials by the follow-up visit.

Patients rated the app (average 3.8 out of 5) and described enjoying the app's educational lessons and videos most, followed by the weight recording feature, and the messaging platform. Patients reported that sometimes they could not find the exact food that they had consumed in the diet database, so they input the next-best option. Throughout the study, 48 symptoms were logged by app users, of which 25 triggered a prompt to contact the study team, and 4 required that patients call their doctors or call 911. In total, 11 messages were sent by app users to the study coordinator through the messaging platform regarding symptoms and trouble-shooting of the app.

App Use and Heart Failure Education, Quality of Life, and Weight

The more a patient interacted with the app (higher the average session duration), the greater the improvement in AHFKT-V2 (ρ=0.59, *P*=.04) and KCCQ-12 (ρ=0.63, *P*=.03) scores from baseline to follow-up (Figure 3). The correlation between app use and weight loss was ρ=–0.40 (*P*=.19).

The median AHFKT-V2 score was 80% at both baseline and final visits for the control arm while the median score was 83.3% at baseline and 85% at the final visit for the intervention arm. The median KCCQ-12 score was 55 at baseline and 41 at the final visit for the control arm, while the median score was 74 at baseline and 70 at the final visit for the intervention arm. Patients in the control arm gained 2.5 lbs (1.1 kg) bodyweight during the study while patients in the app arm lost 3.8 lbs (1.7 kg) bodyweight.

Discussion

In this pilot proof-of-concept study in a sample of patients with heart failure, we found that the Habits Heart App with a Bluetooth-linked scale was feasible and easily implemented. All of the patients in the intervention group were able to demonstrate that they could use the 5 key areas of the app on their own after a 30-minute onboarding session, and almost all of the patients were able to use the app without major stability, performance, or standards adherence issues. The Habits Heart App with scale is unique in that it was designed by clinicians who care for patients with heart failure to address a patient's complex self-management needs for patient education, tracking, and coaching without requiring resource-intensive telemonitoring. While intensive monitoring may help to decrease adverse events in patients with heart failure, telemonitoring can require a lot of resources and be financially taxing [12]. Resource-intensive heart failure management may not be practical for the vast majority of patients with heart failure and may be best suited for high-risk patients [13]. As our experience has shown (patients with frequent hospitalizations not being able to use the app consistently), lower risk patients with heart failure may benefit from the Habits Heart App. Compared with a real-life control group, intervention group patients appeared to have better heart failure knowledge, quality of life, and weight. We also showed that greater involvement with the intervention was associated with greater improvement in heart failure knowledge and quality of life and may be associated with weight loss rather than gain over time.

Compared with a previously published study [8], our control group closely resembled the situation of real-life patients with heart failure; Athilingam and colleagues [8] used an active wait-listed control group with patients who were given an app with heart failure education features that incentivized the patients with the activation of additional features upon reaching the end of the study, which may have motivated patients in the control group to engage more than they typically would in real life. On a separate note, only 1 out of 11 control patients (9%) in our study retained patient reading materials by the follow-up visit. This statistic supports the notion that the usual standard of care of reading materials and handouts may not be engaging for patients with heart failure [14]. Heart failure education delivered through an app, such as the Habits Heart App, would be at the patients' (literal) fingertips and has the potential to adapt to the individual's learning level or style.

While there are several patient-centered heart failure apps available, few have been designed by clinicians who care for patients with heart failure or have been evaluated with a scientific assessment or publication in a medical journal to date [8,9,15-17]. Several well-designed, previously published feasibility studies utilized a more resource-intensive telemonitoring approach (such as a chest strap that monitors physiological data and accelerometer or blood pressure monitoring) in combination with a smartphone app [8,9]. In our study, we focused on the essentials that would not require intensive monitoring or devices for the patient or health care staff. We believe that such an approach would be more easily applied to broad clinical practice where resources may be scarce. The Habits Heart App features a comprehensive interactive tool for patients with heart failure that allows tracking of healthy habits, education of heart failure disease management, and direct contact with study staff and cardiologists for questions regarding app use or medical symptoms. Additionally, the Bluetooth-linked scale allows patients to measure weight directly and visualize trends in a centralized setting. Having customizable, prespecified automatic triggers for involving patients and notifying clinicians (rather than feeding raw data) may reduce the burden on already busy clinicians while empowering patients.

There are several areas identified in the feasibility assessment to address for the future iteration of the app including Bluetooth connectivity with the app, messenger delays, and sodium and exercise logging issues. Enabling app download onto patients' own smartphones, whether iPhone or Android, will solve the internet connectivity issue and decrease the inconvenience of carrying 2 smartphones. An ability to sync summary data with electronic health records may further reduce administrative burden. Addition of mobile laboratory capabilities to check B-type natriuretic peptides, N-terminal pro–B-type natriuretic

peptides, and basic metabolic panel may expand the utility of the app.

Despite randomization, there were differences in patient characteristics between the app and control groups, likely due to the small sample size of this feasibility study. The control group appeared to have more advanced heart failure (higher proportion of patients with New York Heart Association Class III or IV and higher N-terminal pro–B-type natriuretic peptide values), while the intervention group had a higher burden of comorbidities such as atrial fibrillation, coronary artery disease, chronic obstructive pulmonary disease, and chronic kidney disease. The differences between the two groups may potentially result in more improvement in the app group compared with that in the control group; patients with less advanced heart failure may score better. However, in order to minimize the potential impact of baseline differences in the two groups, we looked at the relative change in heart failure knowledge, quality of life, and weight loss rather than the final or absolute values. Nonetheless, any definitive conclusions about the app's effects on quality of life or heart failure knowledge should be evaluated with a large multicenter trial.

The Habits Heart App is a feasible way to dynamically engage patients in heart failure management and equip patients with effective self-management tools to potentially improve heart failure knowledge, quality of life, and adherence to medical recommendations. A larger multicenter trial is needed to further test the use of the app and clinical outcomes.

Acknowledgments

Research reported in this publication was indirectly supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB015408 (awarded to Jana Care Inc). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support for this study was provided by Jana Care Inc. HKG and NEI were supported by the Dennis and Marilyn Barry Fellowship in Cardiology, and HKG is supported in part by the Clark Fund for Cardiac Research Innovation.

Authors' Contributions

KSW, AAK, SJ, VC, MD, and NM contributed to the analysis and interpretation of data and drafting of the manuscript. NEI and HKG contributed to the design of the study, analysis and interpretation of data, and drafting of the manuscript. JCK contributed to the drafting of the manuscript.

Conflicts of Interest

The authors report the following relationships with industry: NEI—Jana Care research grant recipient; AAK, SJ, VC, MD, and NM are Jana Care employees; JCK—advisor for MD Revolution, Medtronic, Mint Health, Res App, and LuminDx, board member for b.well Connected Health and shareholder of b.well, Res app, MD Revolution, and LuminDx; HKG—recipient of research grant support from Roche Diagnostics, Jana Care, Ortho Clinical, Novartis, Pfizer, Alnylam, and Akcea and consulting income from Amgen, Eko, Merck, Roche Diagnostics, Radiometer, and Pfizer, ownership of Eko stocks; recipient of research payments for clinical endpoint committees from Radiometer and Baim Institute for Clinical Research (for Abbott, Siemens, and Beckman Coulter). All other authors have no conflicts to declare.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1207 KB - mhealth_v9i1e19465_app1.pdf]

References

1. Bromage DI, Cannatà A, Rind IA, Gregorio C, Piper S, Shah AM, et al. The impact of COVID-19 on heart failure hospitalization and management: report from a Heart Failure Unit in London during the peak of the pandemic. Eur J Heart Fail 2020 Jun 04;22(6):978-984 [FREE Full text] [doi: 10.1002/ejhf.1925] [Medline: 32478951]
2. Loop MS, van Dyke MK, Chen L, Safford MM, Kilgore ML, Brown TM, et al. Low utilization of beta-blockers among medicare beneficiaries hospitalized for heart failure with reduced ejection fraction. J Card Fail 2019 May;25(5):343-351 [FREE Full text] [doi: 10.1016/j.cardfail.2018.10.005] [Medline: 30339796]
3. Greene SJ, Butler J, Albert NM, DeVore AD, Sharma PP, Duffy CI, et al. Medical therapy for heart failure with reduced ejection fraction: the CHAMP-HF registry. J Am Coll Cardiol 2018 Jul 24;72(4):351-366 [FREE Full text] [doi: 10.1016/j.jacc.2018.04.070] [Medline: 30025570]
4. Reilly CM, Butler J, Culler SD, Gary RA, Higgins M, Schindler P, et al. An economic evaluation of a self-care intervention in persons with heart failure and diabetes. J Card Fail 2015 Sep;21(9):730-737 [FREE Full text] [doi: 10.1016/j.cardfail.2015.06.382] [Medline: 26164214]
5. Silver L, Smith A, Johnson C, Jiang J, Anderson M, Rainie L. Mobile connectivity in emerging economies: use of smartphones and social media is common across most emerging economies. Pew Research Center. 2019 Mar 07. URL: https://www.

pewresearch.org/internet/2019/03/07/use-of-smartphones-and-social-media-is-common-across-most-emerging-economies/ [accessed 2020-11-13]
6. Mobile fact sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-11-12]
7. Wali S, Demers C, Shah H, Wali H, Lim D, Naik N, et al. Evaluation of heart failure apps to promote self-care: systematic app search. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e13173 [FREE Full text] [doi: 10.2196/13173] [Medline: 31710298]
8. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
9. Park C, Otobo E, Ullman J, Rogers J, Fasihuddin F, Garg S, et al. Impact on readmission reduction among heart failure patients using digital health monitoring: feasibility and adoptability study. JMIR Med Inform 2019 Nov 15;7(4):e13353 [FREE Full text] [doi: 10.2196/13353] [Medline: 31730039]
10. Reilly CM, Higgins M, Smith A, Gary RA, Robinson J, Clark PC, et al. Development, psychometric testing, and revision of the Atlanta Heart Failure Knowledge Test. The Journal of Cardiovascular Nursing 2009;24(6):500-509. [doi: 10.1097/jcn.0b013e3181aff0b0]
11. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):469-476. [doi: 10.1161/circoutcomes.115.001958]
12. Gensini GF, Alderighi C, Rasoini R, Mazzanti M, Casolo G. Value of telemonitoring and telemedicine in heart failure management. Card Fail Rev 2017 Nov;3(2):116-121 [FREE Full text] [doi: 10.15420/cfr.2017:6:2] [Medline: 29387464]
13. Xiang R, Li L, Liu SX. Meta-analysis and meta-regression of telehealth programmes for patients with chronic heart failure. J Telemed Telecare 2016 Jul 10;19(5):249-259. [doi: 10.1177/1357633x13495490]
14. Veroff DR, Sullivan LA, Shoptaw E, Venator B, Ochoa-Arvelo T, Baxter JR, et al. Improving self-care for heart failure for seniors: the impact of video and written education and decision aids. Popul Health Manag 2012 Feb;15(1):37-45. [doi: 10.1089/pop.2011.0019] [Medline: 22004181]
15. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
16. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
17. Heiney SP, Donevant SB, Arp Adams S, Parker PD, Chen H, Levkoff S. A smartphone app for self-management of heart failure in older African Americans: feasibility and usability study. JMIR Aging 2020 Apr 03;3(1):e17142 [FREE Full text] [doi: 10.2196/17142] [Medline: 32242822]

Abbreviations

AHFKT-V2: Atlanta Heart Failure Knowledge Test, Version 2
KCCQ-12: Kansas City Cardiomyopathy Questionnaire

Edited by C Dias; submitted 07.05.20; peer-reviewed by M Marcolino, M Pijl, H Ross, K Radhakrishnan, N Mohammadzadeh; comments to author 23.06.20; revised version received 24.11.20; accepted 29.11.20; published 20.01.21.

Please cite as:
Wei KS, Ibrahim NE, Kumar AA, Jena S, Chew V, Depa M, Mayanil N, Kvedar JC, Gaggin HK
Habits Heart App for Patient Engagement in Heart Failure Management: Pilot Feasibility Randomized Trial
JMIR Mhealth Uhealth 2021;9(1):e19465
URL: http://mhealth.jmir.org/2021/1/e19465/
doi:10.2196/19465
PMID:33470941

Original Paper

Utilization of Smartphone Depth Mapping Cameras for App-Based Grading of Facial Movement Disorders: Development and Feasibility Study

Johannes Taeger[1], MD; Stefanie Bischoff[1], MD; Rudolf Hagen[1], Prof Dr; Kristen Rak[1], MD

Department of Otorhinolaryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery, University Hospital Würzburg, Würzburg, Germany

Corresponding Author:
Johannes Taeger, MD
Department of Otorhinolaryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery
University Hospital Würzburg
Josef-Schneider-Straße 11
Würzburg, 97080
Germany
Phone: 49 931 201 21489
Email: taeger_j@ukw.de

Abstract

Background: For the classification of facial paresis, various systems of description and evaluation in the form of clinician-graded or software-based scoring systems are available. They serve the purpose of scientific and clinical assessment of the spontaneous course of the disease or monitoring therapeutic interventions. Nevertheless, none have been able to achieve universal acceptance in everyday clinical practice. Hence, a quick and precise tool for assessing the functional status of the facial nerve would be desirable. In this context, the possibilities that the TrueDepth camera of recent iPhone models offer have sparked our interest.

Objective: This paper describes the utilization of the iPhone's TrueDepth camera via a specially developed app prototype for quick, objective, and reproducible quantification of facial asymmetries.

Methods: After conceptual and user interface design, a native app prototype for iOS was programmed that accesses and processes the data of the TrueDepth camera. Using a special algorithm, a new index for the grading of unilateral facial paresis ranging from 0% to 100% was developed. The algorithm was adapted to the well-established Stennert index by weighting the individual facial regions based on functional and cosmetic aspects. Test measurements with healthy subjects using the app were performed in order to prove the reliability of the system.

Results: After the development process, the app prototype had no runtime or buildtime errors and also worked under suboptimal conditions such as different measurement angles, so it met our criteria for a safe and reliable app. The newly defined index expresses the result of the measurements as a generally understandable percentage value for each half of the face. The measurements that correctly rated the facial expressions of healthy individuals as symmetrical in all cases were reproducible and showed no statistically significant intertest variability.

Conclusions: Based on the experience with the app prototype assessing healthy subjects, the use of the TrueDepth camera should have considerable potential for app-based grading of facial movement disorders. The app and its algorithm, which is based on theoretical considerations, should be evaluated in a prospective clinical study and correlated with common facial scores.

(JMIR Mhealth Uhealth 2021;9(1):e19346) doi:10.2196/19346

KEYWORDS

facial nerve; facial palsy; app development; medical informatics; eHealth; mHealth; Stennert's index; depth mapping camera; smartphone sensors

Introduction

Due to significant functional and cosmetic restrictions, facial palsy is a clinical symptom that is associated with a very high degree of psychological strain for those affected [1]. The incidence of peripheral facial palsy is 23-35 cases per 100,000 people; the incidence of the central form remains unknown. Both sexes are often affected equally, and those in the age groups of 30-50 years and 60-70 years are especially affected [2]. Affected patients can have severe functional problems

ranging from asymmetry at rest, decreased facial movement, incomplete eye closure, problems eating and drinking, and a decreased sense of taste [3,4]. The fact that no underlying disease can be identified in about three-quarters of the cases (currently designated as idiopathic) also represents an unsatisfactory diagnostic situation [2]. There is still a great need for research in this area in order to ultimately be able to offer targeted therapy for these cases. Restoring facial function to the greatest possible extent can lead to improved self-confidence and a higher quality of life [5-7]. However, not only the diagnosis but also the quantification of facial palsy are still problematic.

Current Methods for Quantifying Facial Palsy

Various systems have been established for the scientific and clinical classification of the degree of facial palsy, assessment of therapeutic interventions, and spontaneous disease course. These standardized systems facilitate effective communication with patients and professionals. The most widely established clinician-graded scoring systems are the Stennert index [8], House-Brackmann scale [9], and Sunnybrook Facial Grading System [10]. Numerous efforts to generally improve facial grading systems have been described in the literature. For example, an approach was published by Banks et al [11] with the electronic, clinician-graded facial function scale (eFACE), which was subsequently evaluated internationally [11-13]. Nevertheless, like other clinician-graded scoring systems, eFACE is also subject to bias and human error [14]. While a comparatively simple measuring method with a conventional hand ruler was propagated by Manktelow et al [15] in 2008, the immense technological progress of the past few years has led to software-based methods. Here, the measurement of facial expressions with Adobe Photoshop [16], for example, but also semiautomatic photographic assessment tools [17] were developed. Katsumi et al [18] described an advanced 3-dimensional (3D) facial motion measurement system, consisting of a color charge-coupled device camera and 2 laser scanners in combination with their proprietary software. In this experimental setting, the test subjects were asked to perform various facial movements. The authors describe that the measurement time was less than 5 minutes in each case. After the measurement, the data were transferred to a computer and then evaluated and compared with the clinical results from the Yanagihara and House-Brackmann grading scales, whereby a good correlation could be demonstrated [18]. Furthermore, several machine learning approaches [19-21] have been developed. Despite all efforts in the description and validation of various systems, none could achieve universal acceptance in everyday clinical practice. This is due to the fact that many systems are characterized as being very time-consuming or requiring expensive and complicated hardware, which is why some can hardly be used in routine practice [14].

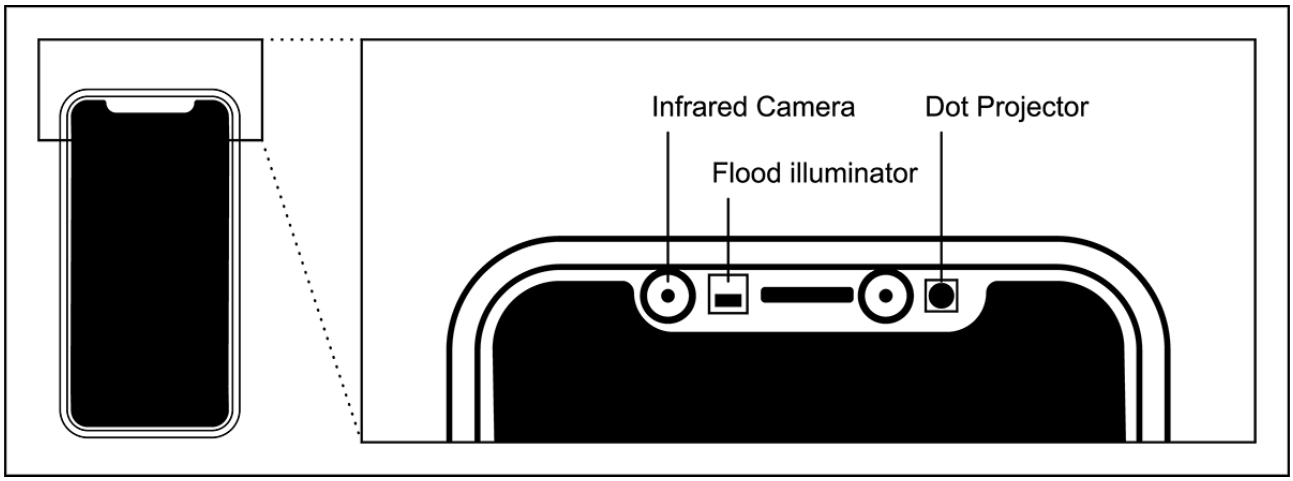
Hence, it would be desirable to have a quick and precise tool for assessing the functional status of the facial nerve. This would improve and streamline the data situation for symptom severity in the case of facial palsy and the effectiveness of different treatment methods like the application of prednisolone or acyclovir [4]. In addition, it would enable patients without any required training to perform a reproducible measurement of the extent and course of their own disease, which might lead to an increase in motivation for performing therapeutic activities like facial exercises. For the latter, positive effects in the treatment of facial paralysis have been demonstrated [6,22-27].

Smartphone Sensors as Sophisticated Medical Tools

Since the first smartphone was announced by Apple in 2007, the technical development of smartphones has progressed rapidly. In addition to an immense increase in computing power and storage options, the devices have been gradually equipped with high-precision sensors. For example, newer smartphone generations have an acceleration sensor, 3-axis gyroscope, barometer, proximity sensor, ambient light sensor, GPS, digital compass, microphone(s), and sophisticated camera system(s) [28]. In addition, high speed wireless communication platforms, such as long-term evolution, allow connection to health care providers at any time [29,30].

With regard to measurements of different facial muscles, one feature of recent smartphones is of high interest: the TrueDepth camera system. It enables user authentication via face recognition with the Face ID system and has been implemented since the iPhone generation X. The exact registration of facial data works via the projection of 30,000 infrared points, which are recorded at 60 times per second via an infrared camera and processed by a neural engine of the smartphone chip. An infrared flood illuminator adds infrared light to enable measurements also in the dark (Figure 1). An accurate 3D depth map of the face is calculated from the measurement data, which is converted together with the infrared image to a mathematical representation of the user's facial features. According to the manufacturer, the technology works so precisely that the probability of false-positive detection of a random person is 1:1,000,000. In addition, the technology functions even with modifiers of the user's appearance such as makeup or beard growth as well as hats, glasses, contact lenses, or scarves [31,32].

Figure 1. Hardware components of the TrueDepth camera system, which is integrated in the upper part of the smartphone. A Dot Projector throws over 30,000 infrared dots onto the face of the user, and the dots and an infrared image are captured via an infrared camera to create a depth map of the face (modified from [20]).



Apple's ARKit framework enables developers to use face tracking with the TrueDepth camera for applications that go beyond Face ID. It is possible to access 52 predefined facial movement features, the extent of which is expressed in a numerical value ranging from 0.0 (neutral) to 1.0 (maximum movement) [33]. These features are also calculated 60 times per second, which results in a very high temporal resolution of the recorded facial movements.

The aim of the present study was to develop an app prototype that objectively and reproducibly quantifies facial asymmetries with the data from the iPhone's TrueDepth camera system. With the developed app, it was possible to perform the measurement within a few seconds and without any additional external devices or the need for calibration. The extracted data were used to calculate a new index for the grading of unilateral facial palsy, called the Digital Facial Index (DFI), which is represented in a numerical value of 0%-100% for intuitive understanding.

Conception of the App

During the content-related conception, the following features of the app were defined:

- Start screen with navigation elements and a setting option for the duration of the actual measurement to meet the needs of different users (for inexperienced users, for example, a time window for the measurement of 10 seconds might be too short, whereas for other users, 5 seconds is sufficient)
- Measuring mode with a simple instruction on 3 functionally relevant facial movements ("raise your eyebrows," "close your eyes," "pull up the corners of your mouth") and a countdown timer over the previously set duration of the measurement
- Evaluation screen with an easily understandable representation of the individual measurement parameters and the calculated output of the newly developed index (DFI)
- Buttons to save the DFI measurement results within the app for simple statistics and repeating the measurement
- Raw data output for scientific evaluation

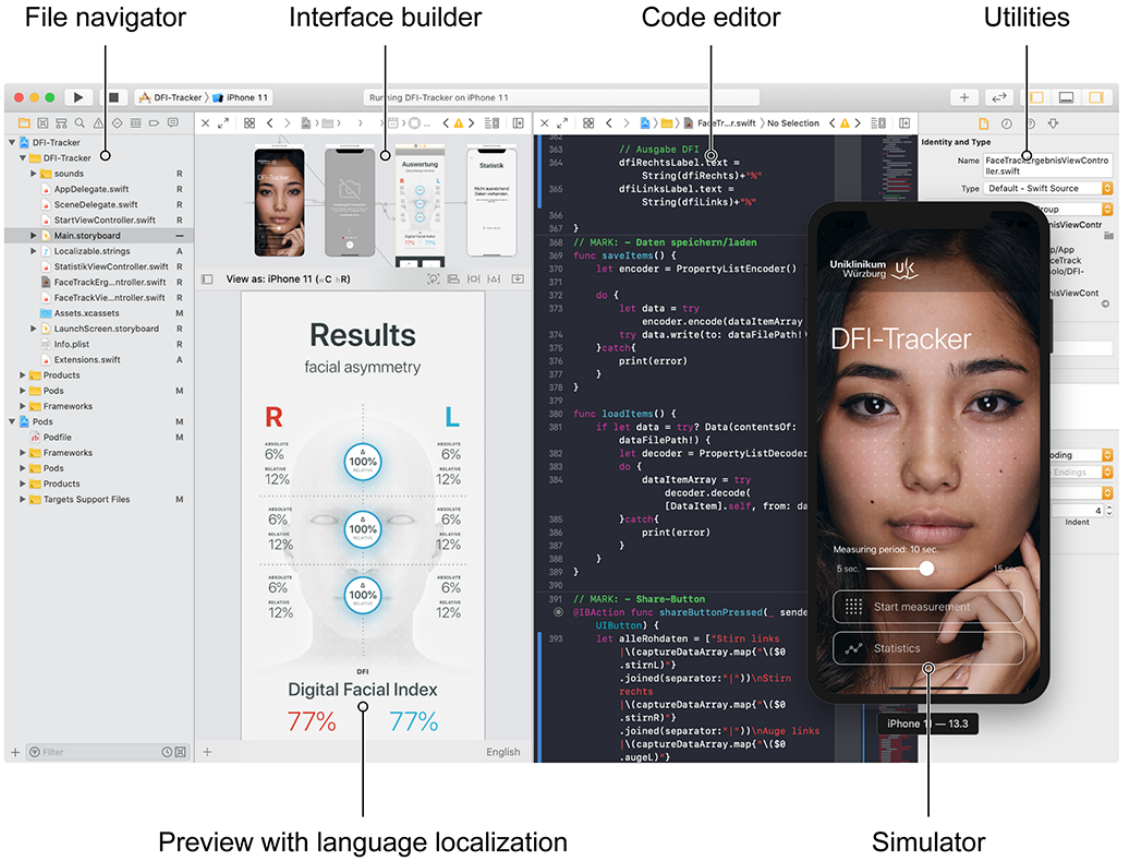
Methods

App Development

A native app architecture under iOS (Apple Inc, Cupertino, CA) was chosen to develop the app. One of the reasons why Apple devices and Apple's ecosystem were chosen is that the Face ID technology, which is based on 3D data collection, is considered unrivalled compared to the 2-dimensional technology of its competitors [34]. In addition, iOS is primarily suitable for the development of medical apps due to its high security standards, compatibility, usability, and stability , even if it is not possible to prove these factors objectively in their full extent.

The user interface was designed with Adobe Creative Suite products (Photoshop 2020, Illustrator 2020; Adobe Inc, San José, CA) and the Sketch App (Version 61.2; Bohemian Coding, London, UK), which is specially optimized for the design of mobile apps. Then, the app was implemented using the Swift programming language in the latest version of the Xcode development environment (Apple Inc; Figure 2). An open-source library (ScrollableGraphView [35]) was integrated with the dependency manager CocoaPods [36] to graphically display the stored data in the form of a graph. As far as possible with a prototype, the Mobile App Rating Scale was used as a template for compliance with established quality criteria [37].

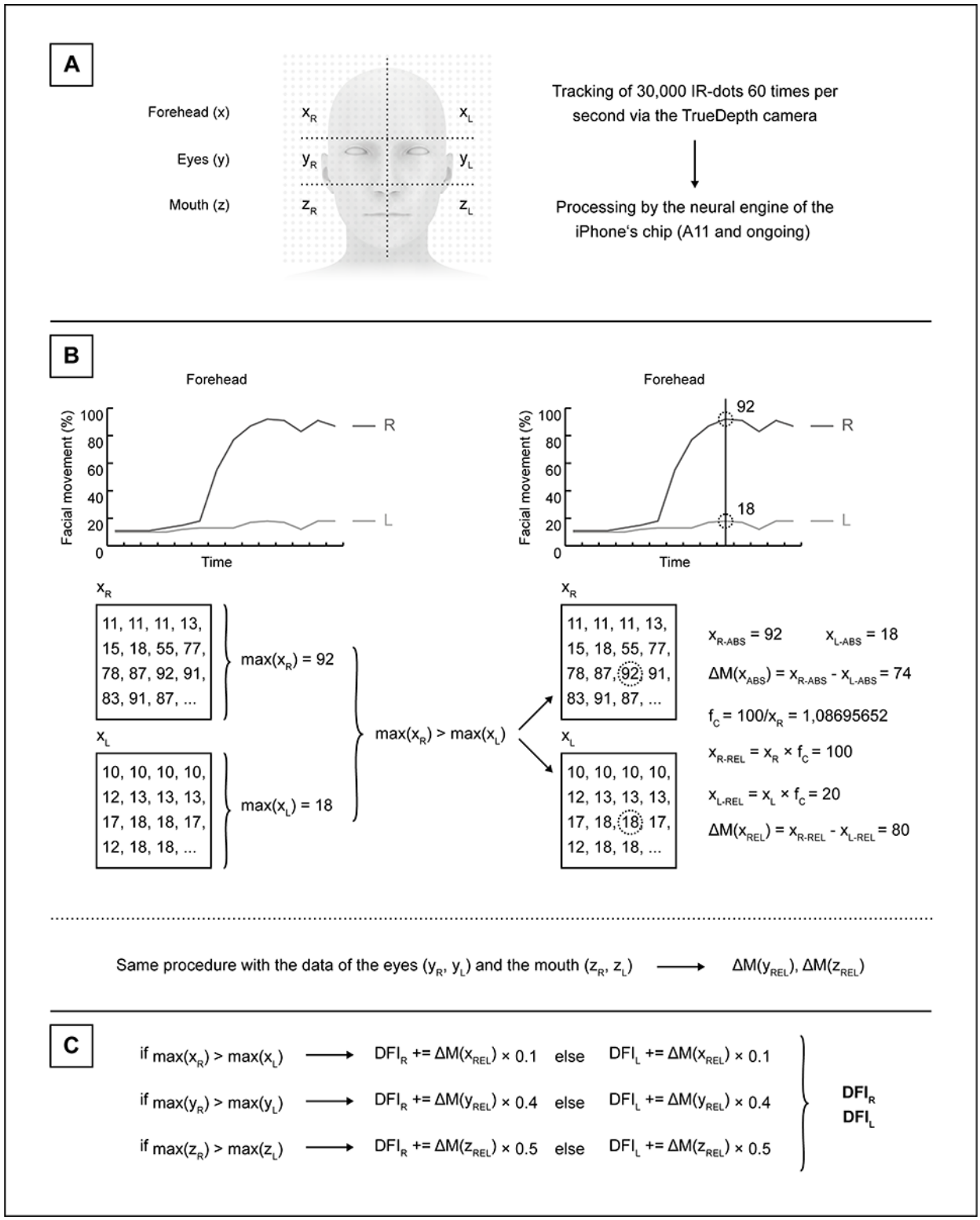
Figure 2. Screenshot of the app developed in the Apple Xcode Development environment. All associated project files are listed in the file navigator. The graphical user interface is designed with the interface builder and then linked to the code, which is written in the code editor. A utility area allows further settings. For the implementation of different regions, the user interface can be checked via the preview in the corresponding language. The simulator enables testing of the app on various virtual devices with the iOS operating system.



Calculation of the Digital Facial Index

Since the data output for the specific facial features via the ARKit framework has a value ranging from 0.0 to 1.0, the expression of the movement extent in the form of an easily understandable percentage value seemed rational. Similar to the description by Stennert et al [8], individual facial regions should be matched with a defined weighting based on functional and cosmetic aspects, namely with the proportions of 10% for the forehead, 40% for the eye area, and 50% for the mouth region. The newly developed DFI should adhere exactly to this weighting for optimal follow-up correlation using the Stennert index that applies an easily comparable scaling (0-10). The algorithm for calculating the DFI is shown in Figure 3: The so-called blend shape coefficients, which describe the movement of the eyebrows, closure of the eyelids, and upward movement of the corners of the mouth (in accordance with the Apple Developer Documentation) are recorded at 60 times per second and saved as an array. From this data record, the respective maximum for the individual feature is determined. For each facial region (forehead: x; eyes: y; mouth: z), it is assumed that in a patient suffering from unilateral facial palsy, the side with the greater maximum corresponds to the healthy side. The numerical value of the facial movement of the opposite side in each specific region is used to compare with the maximum of the stronger (healthy) side. The difference in facial motility $\Delta M(x_{ABS}, y_{ABS}, z_{ABS})$ for the 3 regions is calculated by subtracting these 2 numerical values. As it was observed in test runs with the prototype in healthy subjects, the respective absolute values often do not total 100% despite a complete raising of the eyebrows, a full closure of the eyelids, and a maximal lifting of the corners of the mouth; therefore, a correction factor f_C was introduced. This scales the value of the half of the face with higher values in 1 of the 3 areas to 100% and the half of the face with the lower values proportionally. This aims to obtain a realistic evaluation of facial symmetry, derived from theoretical considerations. Thus, relative values and the corresponding relative differences $\Delta M(x_{REL}, y_{REL}, z_{REL})$ are calculated, which are ultimately used in the calculation of the DFI. For the 3 regions, the relative differences are summed up for the left and the right depending on the identified side of facial palsy by comparison of the maximum values for each region. A patient affected by facial palsy will initially have higher differences calculated in this way in terms of facial asymmetry. In most patients, these differences become smaller [38]. In order to represent the course of the disease in a positive trajectory, the maximum extent of the facial palsy was set to 0% and a perfectly symmetrical face to 100%. Graphically represented, an increasing rather than a decreasing curve should result in most cases. Thus, this procedure is reciprocal to Stennert´s index, in which the highest value (10) corresponds to complete facial palsy.

Figure 3. Algorithm for the calculation of the Digital Facial Index (DFI): (A) The TrueDepth camera tracks 30,000 infrared dots projected by the iPhone's dot projector, and the data are rendered as so-called blend shape coefficients for the (B) measurement time window for the forehead, eyes, and mouth (exemplary data shown), and the maxima of both sides are compared for each facial region; the half of the face with the larger maximum corresponds to 100% (relative value). (C) The DFI is calculated by adding the relative differences (ΔM based on absolute and relative values) of the 3 facial regions, with a weighting, similar to Stennert´s index.



Data Display and Sharing Options

Following a successful measurement, a screen opens automatically displaying the analyzed data. It shows all calculated values of the 3 facial regions including the resulting DFI values. There are also buttons for repeating the measurement and storing the current data, as well as a button that enables raw data output for scientific purposes. This was implemented using the *UIActivityViewController* class from Apple's UIKit, which allows the sharing of data (eg, via AirDrop, email, or text message).

Test Measurements

Applying the prototype developed as described in the previous section, all 4 authors of this publication carried out 10 consecutive self-measurements using an iPhone X_R to evaluate the intertest variability. In this setting, all test subjects looked straight ahead into the smartphone's front camera or display while holding the device in their hands.

To check the robustness under unfavorable conditions, the results were compared with those obtained from one individual to determine whether different angles of the smartphone front camera relative to the face (reference point: median of the head at eye level; distance: 37.5 cm, which is within the distance of 25-50 cm recommended by the manufacturer and roughly corresponds to a natural holding position) leads to different results. During this measurement, the positions of both the smartphone and the subject's face were fixed with a tripod. In addition, we checked whether incorrect measurements occur if the smartphone is held deliberately and intermittently leaves the camera's field of view.

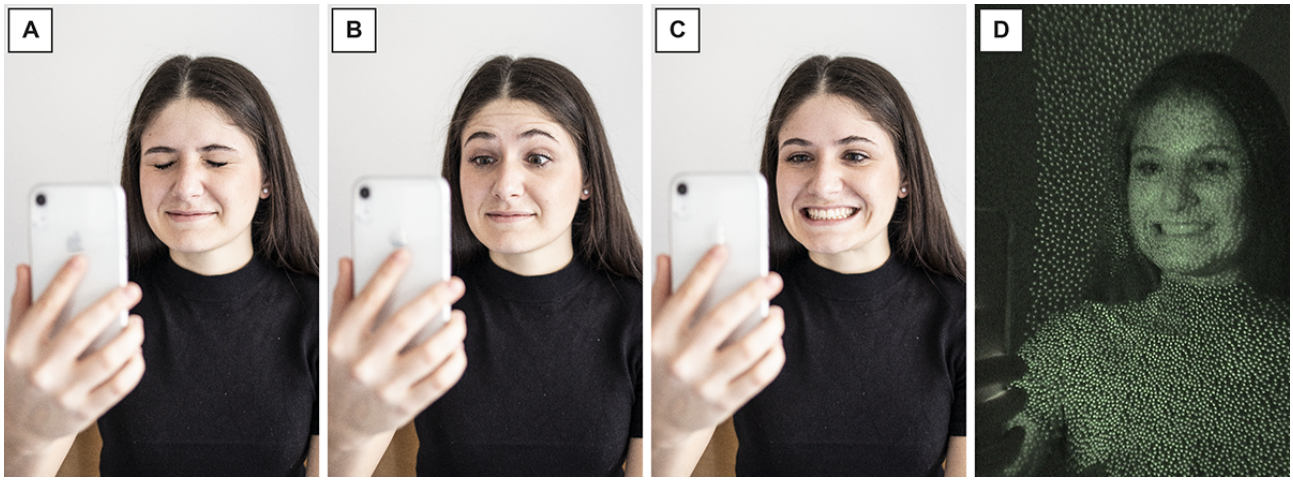
No ethic committee approval at a named institution was needed and applied for, since only data from the authors were used. All authors provided written informed consent for their data and pictures used in the study.

Statistical analysis was performed using one-way analysis of variance in GraphPad Prism (Version 8.2.1; GraphPad Software Inc, San Diego, CA). Differences were considered significant when $P \leq .05$.

The raw data of the measurements were imported as a text file in Microsoft Excel (Version 16.34; Microsoft Corporation, Redmond, WA) for rendering as a graph and to double check that the calculations within the app were correct. The measurements were also partially documented in the form of a screen recording with QuickTime Player (Version 10.5; Apple Inc) and the iPhone connected to a Mac computer (Apple Inc) using a lightning cable.

For demonstration purposes, conventional photographs of test sequences were taken with a Canon 80 D digital single-lens reflex camera (Canon Inc, Tokyo, Japan), with a Canon 35 mm f2.0 lens attached (Figures 4A-4C). In addition, an infrared image was recorded with a Sony DCR-TRV25E Camcorder (Sony Corporation, Tokyo, Japan) in “NightShot” mode during the measurement (Figure 4D).

Figure 4. Test sequence using the Digital Facial Index (DFI) tracker app, with conventional photos during which the user performed 3 facial movements during a defined measurement time window, including (A) “close your eyes,” (B) “raise your eyebrows,” and (C) “pull up the corners of your mouth,” and (D) an infrared photograph that shows part of the 30,000 points thrown at the user by the smartphone’s dot projector to capture depth data of the face. The depicted author provided written informed consent for her photos to be used for this paper.

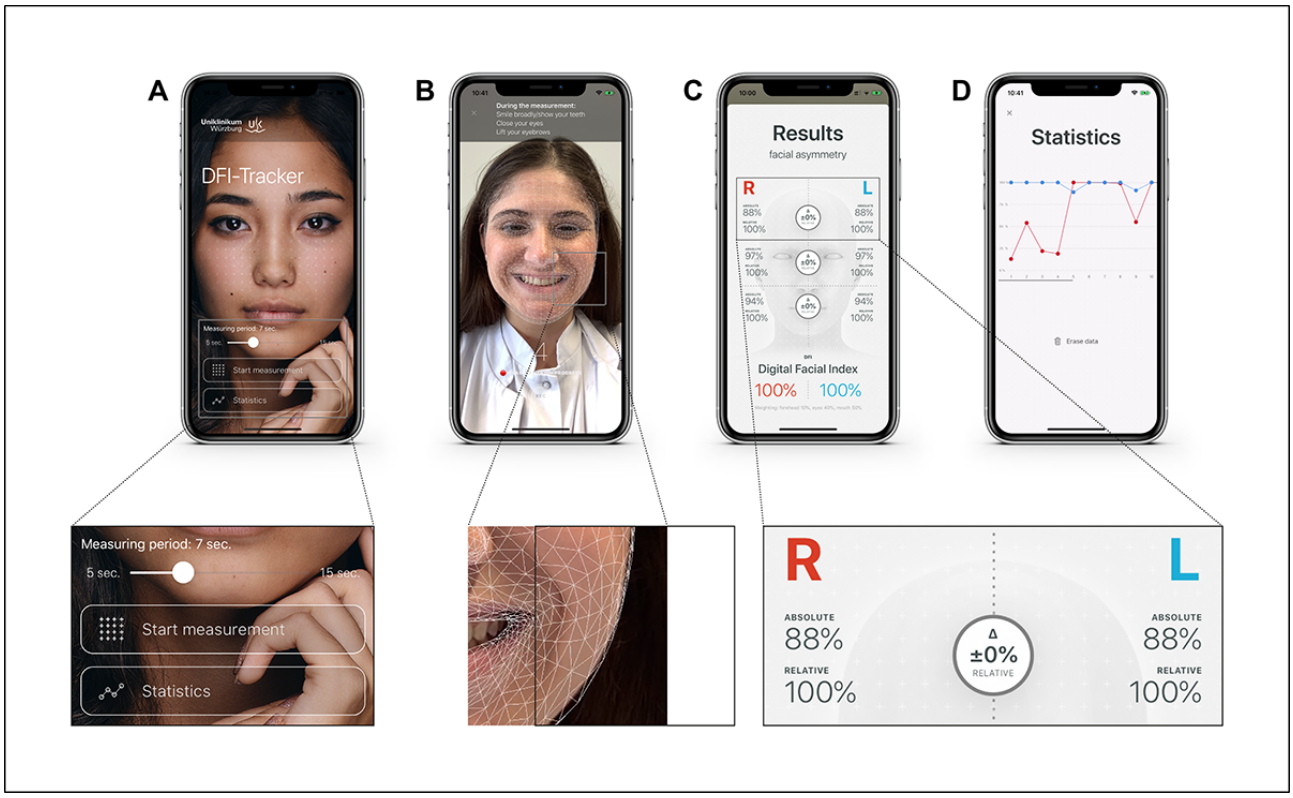


Results

On the basis of the described theoretical considerations, a functioning, stable app prototype could be developed that utilizes the TrueDepth camera system and the ARKit framework to measure the extent of facial motions fast and objectively. In the latest version of the prototype, no runtime nor buildtime errors occurred. The requirements for the app that were initially defined could be fully met.

The complex underlying technology was successfully integrated into a user-friendly interface (Figure 5; Multimedia Appendix 1). When using the measurement mode, a real-time overlay of a polygon grid over the user’s face provides immediate feedback on the precise registration of the face.

Figure 5. User interface of the Digital Facial Index (DFI) Tracker App, including the (A) start screen with a setting for the measurement duration (5 to 15 seconds) and buttons to get to the measurement mode or statistics; (B) measurement mode with a real-time polygon grid overlay for immediate feedback on the precise registration of the user's face; (C) results showing absolute and relative values, their difference, and calculated DFI values; and (D) statistics with a simple graphical representation of the last measured values. The depicted author provided written informed consent for her photos to be used for this paper.



With the newly defined DFI, the measurement result was an easily understandable numerical value expressed as a percentage for each half of the face. After a brief explanation (<1 minute; the authors were initially not exactly familiar with the measuring principle of the app prototype), measurement with the app prototype was easily and independently carried out by the respective test person within a time window of less than 10 seconds each (Multimedia Appendix 1). For scientific analysis, the raw data can be easily transferred to a computer via the "share raw data" button as a text file. The import into Microsoft Excel, for example, for statistical analysis was also an easy and efficient process due to the text file's raw data structure, which was optimized for this purpose.

The measurements were reproducible and showed no statistically significant intertest variability (Figure 6; Multimedia Appendix 2). The DFI was ≥99% for both sides in all measurements and in all subjects with just one exception in a single measurement (97% on one side). Hence, all healthy test persons were identified correctly as having no facial movement disorder.

When comparing the calculated absolute and relative values and the resulting DFI values, there were no discrepancies in the visual comparison with the video recordings and the curves derived from the raw data (Figure 7A). Simulating facial asymmetries by grimacing also gave consistent results (Figure 7B). It should be noted here that a healthy test person is not able to correctly simulate all aspects of facial palsy. With maximum voluntary movement of the facial muscles on one side, the opposite side also moves to a small extent. In the example shown, our subject also found it more difficult to raise the right eyebrow as high as the left as part of the movements combined here. This results in a barely noticeable deflection of the absolute measured values for the right eyebrow.

Figure 6. Intertest variability and SD (bars) of measurements with the app prototype and 4 healthy subjects (10 measurements each); no asymmetry of facial motor skills was found.

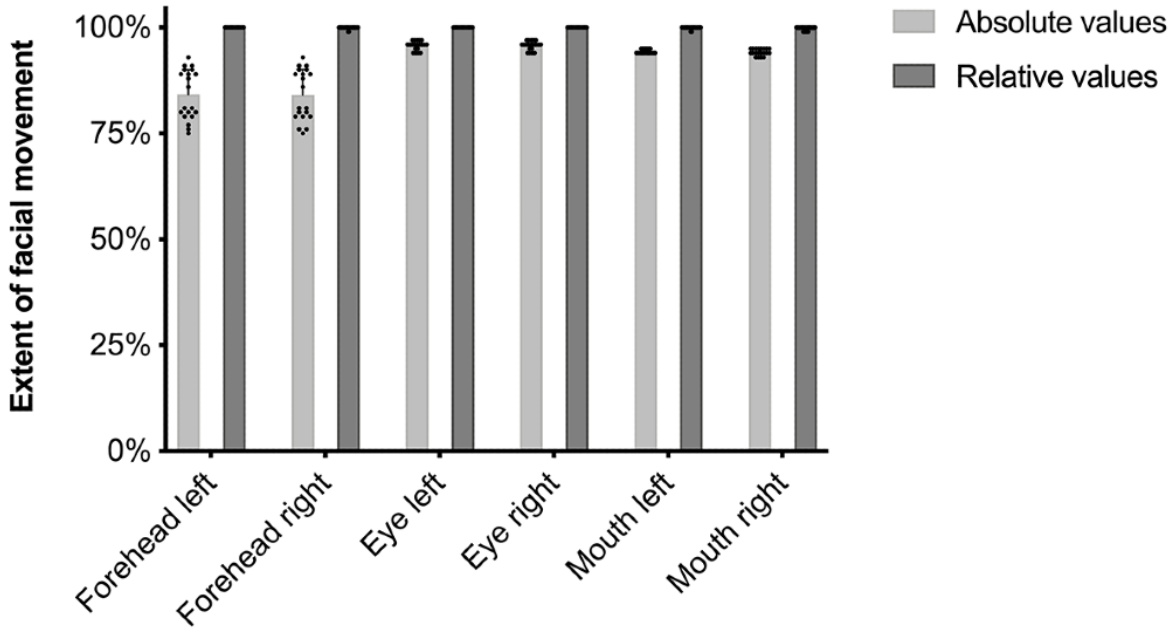
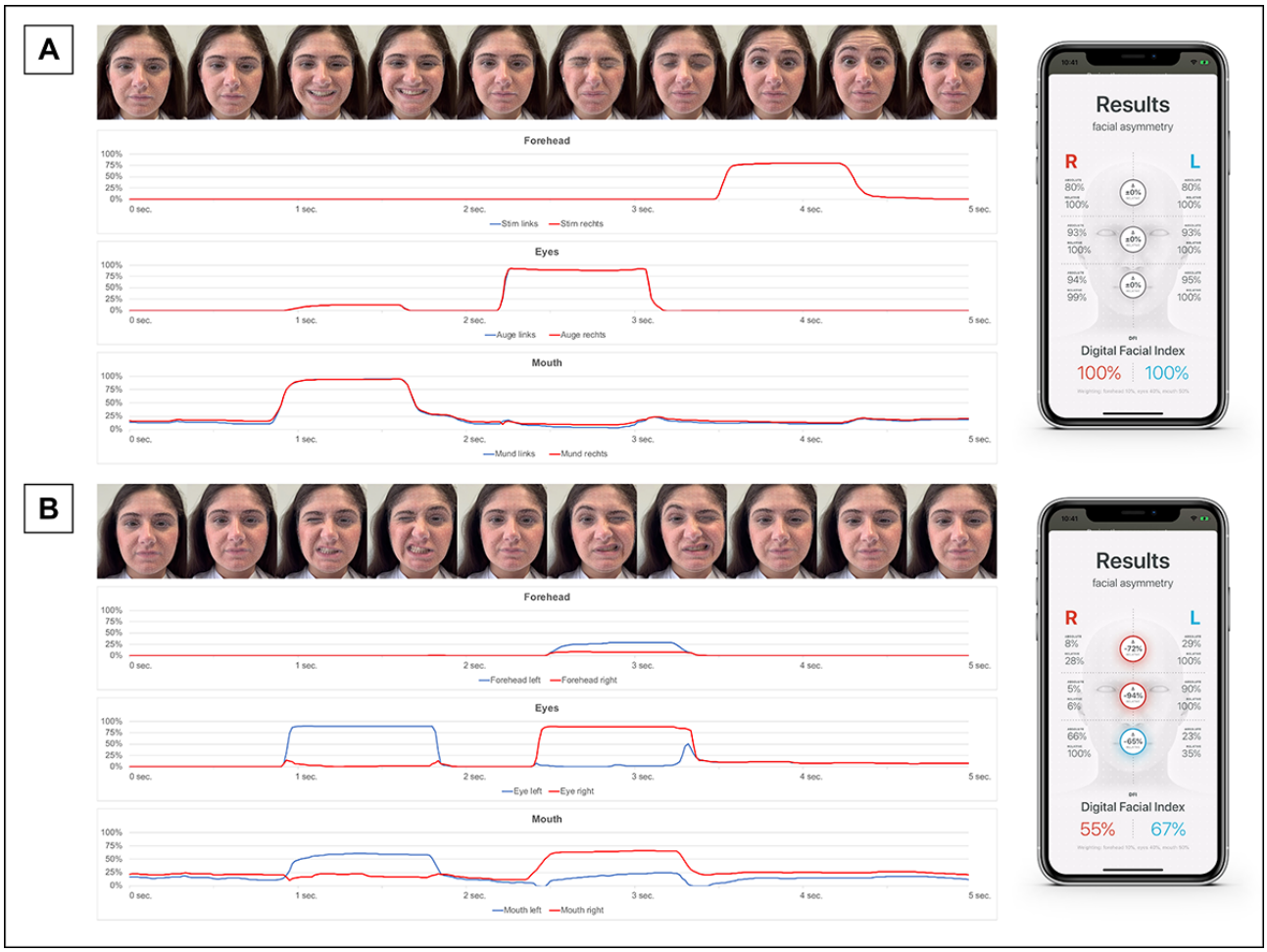


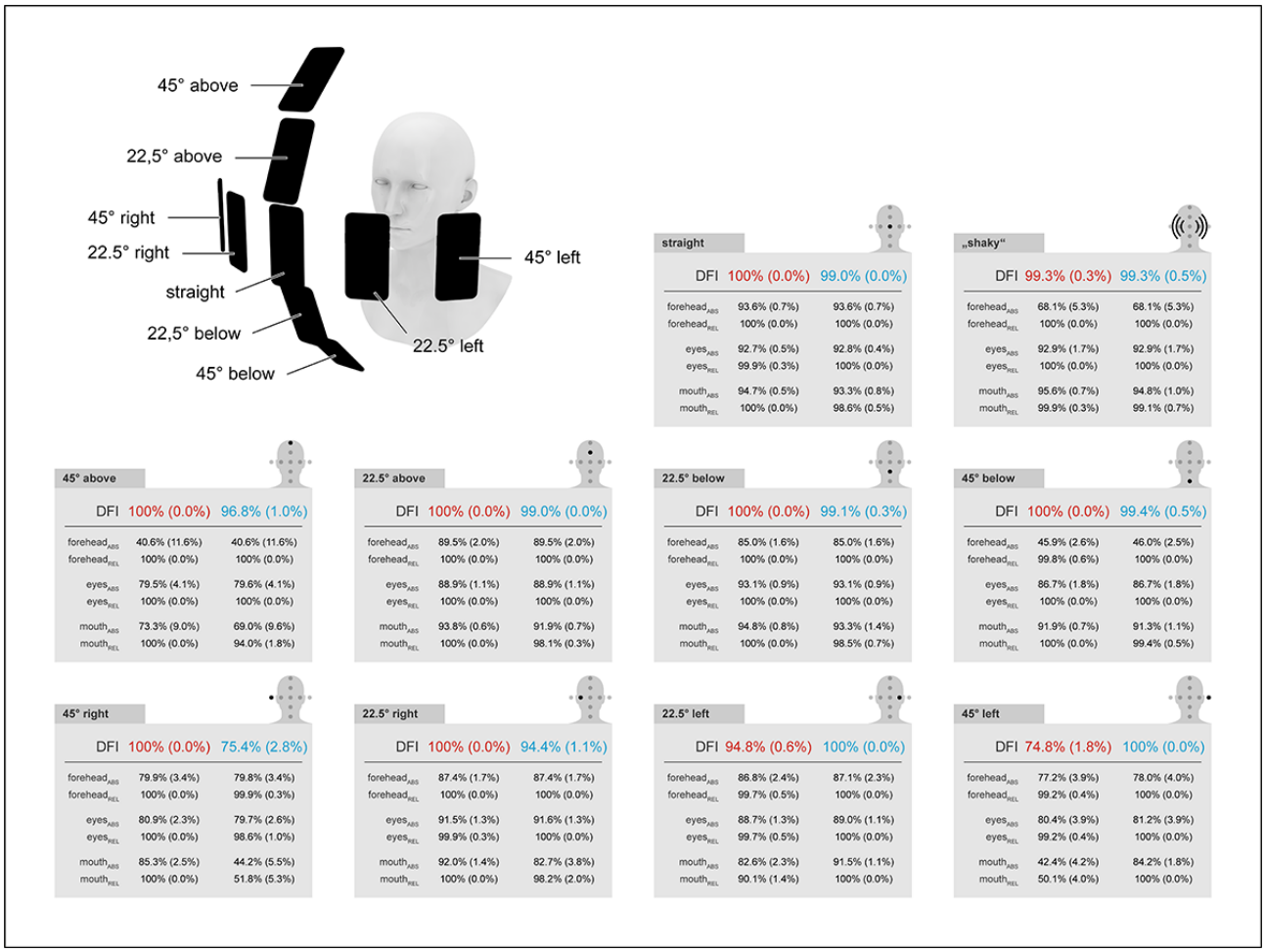
Figure 7. Test measurements with one of the (healthy) authors, showing the raw data (absolute values) over time, which are recorded 60 times per second, for a (A) normal measurement sequence, in which the test subject makes 3 grimaces that result in symmetrical facial motor skills as Digital Facial Index (DFI) values of 100% for both sides, and (B) simulation of asymmetrical facial features alternately for each side. The video stills shown here are mirrored horizontally, as is usual when shooting with the front camera of smartphones. The depicted author provided written informed consent for her photos to be used for this paper.



Nevertheless, there is a symmetrical lifting of the eyebrows with higher measured values during the regular measurement process (Figure 7A). Manually checking the app's internal calculations with the raw data in Microsoft Excel did not identify any detectable errors.

The measurement from different angles (Figure 8) relative to the face gave consistent results when looking straight into the camera (DFI_R 100% [SD 0.0%], DFI_L 99.0% [SD 0.0%]), 22.5 ° from above (DFI_R 100% [SD 0.0%], DFI_L 99.0% [SD 0.0%]), 22.5 ° from below (DFI_R 100% [SD 0.0%], DFI_L 99.1% [SD 0.3%]), and 45 ° from below (DFI_R 100% [SD 0.0%], DFI_L 99.4% [SD 0.5%]). The measurements with a lateral deviation of 22.5 ° resulted in side differences for the opposite half of the face: The measurements 22.5° from the right resulted in DFI_R of 100% (SD 0.0%) and DFI_L of 94.4% (SD 1.1%) and 22.5° from the left resulted in DFI_R of 94.8% (SD 0.6%) and DFI_L of 100.0% (SD 0.0%). There were significant deviations at angles of 45 ° from above (DFI_R 100% [SD 0.0%], DFI_L 96.8% [SD 1.0%]) and 45 ° from the right (DFI_R 100% [SD 0.0%], DFI_L 75.4% [SD 2.8%]) and the left (DFI_R 74.8% [SD 1.8%], DFI_L 100% [SD 0.0%]). Interestingly, with these large angles, the mouth component is decisive for the measurement error, whereas no relevant errors occur for the forehead and eye regions. If measurements were made with deliberately shaky hand movements and occasionally leaving the face out of the camera's field of view, there were no relevant measurement errors (DFI_R 99.3% [SD 0.3%], DFI_L 99.3% [SD 0.5%]).

Figure 8. To simulate unfavorable measurement conditions, 10 measurements (mean [SD] shown) were carried out on a (healthy) test person at different angles relative to the median of the head at eye level and at 37.5 cm, with tripods to fix the position of the smartphone and the head of the test person; in addition, a measurement was made with deliberately shaky hand movements and occasionally leaving the face out of the camera's field of view.



Discussion

Smartphones, tablets, and related widely used devices such as smartwatches offer immense potential for the development of medical applications thanks to highly advanced sensors and computing power, as well as high storage capacities. For health monitoring and diagnosis, smartphone sensors have already proven to be highly potent tools in many medical specialties. To name just a few examples, smartphones are suitable for measuring heart rate and its variability [39-41]. Even the accurate automated detection of atrial fibrillation is possible [42-44]. For pulmonary medicine, low-cost spirometers were developed using the microphones of smartphones [45,46]. Smartphone cameras are suitable for retinal fundus imaging for assessing ophthalmic diseases [47-49], as well as for laboratory applications, like the detection of specific DNA sequences in combination with specific assays [50,51]. Other applications are for skin health monitoring [52-55], detection of anemia [56], sleep monitoring [57-59], and mental health assessment [60-66].

In addition, smartphones are highly available worldwide and are an integral part of everyday life. In September 2019, the number of active smartphones reached 3.6 billion. Although this work only describes an implementation for iPhones, which limits the potential target group to Apple's smartphone market share of 22.3% [67], a comparable solution for iPads and Android devices would be technically easy to implement. It is currently assumed that other manufacturers such as Samsung will also integrate a similar depth mapping camera system in future smartphones [68].

Regarding possible security concerns with respect to the TrueDepth data, Apple states that the data are protected by the so-called Secure Enclave (a coprocessor that provides cryptographic operations for data security) and do not leave the device or are backed up in an online storage medium without user permission [31,69]. The measurement results collected in the app prototype described are also not assigned to any user data nor are saved in a cloud without any active user action (eg, via the button for raw data export that is currently intended for further scientific analysis). In a future user version, features such as integration into electronic patient records would be desirable while maintaining correspondingly high security standards.

One limitation of this study is the fact that the newly developed app is only aimed for patients with unilateral facial palsy. However, bilateral facial palsy is very rare, with an annual incidence rate of 1:5,000,000 and only accounts for 0.3%-3.0% of all cases with facial palsy [70-72].

Selecting 3 of 52 possible facial features that can be extracted simultaneously using ARKit was intended to test movements that are also carried out when the motility criteria of Stennert's index are assessed. The presence of facial synkinesis as well as the resting tone are not determined using the method described. However, as Stennert et al [8] stated in their description of the Stennert index in 1977, motility reveals more about the degree of residual innervation or reinnervation of the facial nerve than resting tone.

The formulas described here for interpreting the TrueDepth data and for calculating the DFI were created on the basis of theoretical considerations and aim for a symmetry of the facial motor skills by matching the face in 3 areas. Since the range of motion of the healthy half of the face—with the idea of symmetry as the ideal state—represents a reasonable measure for the sick side, the absolute values of the healthy side are scaled to 100% and compared with the proportionally scaled opposite side. However, it has to be stated that this approach was not chosen to achieve ultimate scientific accuracy, but to allow for a simple and realistic evaluation of facial symmetry. In combination with the existing functionality of the TrueDepth camera, there is no need for calibration. At present, it is not possible to estimate to what extent the newly designed app reflects the "reality" in patients with facial palsy and the course of their disease, which is currently expressed by established clinician-graded scoring systems. Nevertheless, based on the first experiences with the app prototype and with no statistically significant intertest variability in healthy subjects, it seems possible to advance a precise classification of the severity of facial palsy by the use of this app.

If you compare our approach with clinical scoring systems such as the Stennert Index, House-Brackmann scale, Sunnybrook Facial Grading System, or eFACE, we see our method clearly having an advantage in terms of intraobserver and interobserver variability and the duration of the measurement (a few seconds for the DFI measurement compared to a few minutes such as when collecting the 19 eFACE parameters). When compared with technical solutions, the solution described by Katsumi et al [73] has broad similarities in terms of the functional principle. However, this requires a relatively complex hardware setup. Since smartphones are ubiquitous and global and area-wide distribution of software via app stores is possible, we consider app-based solutions as described here to be much more practical and cheaper. The, from our experience, simple, fast, and practically self-explanatory measuring process should also make it possible for laypeople to measure themselves for self-assessment.

The DFI measurements from different angles and with deliberately shaky hands show that the measurement method is sufficiently robust in realistic scenarios. Since the user, from the experience of our test runs, intuitively looks into the camera during the measurement so that his face is completely shown in the display, relevant lateral axis deviations and the associated measurement errors are unlikely. Nonetheless, in a future end-user solution, it could make sense to provide the user with appropriate information about the correct measurement process.

Another positive aspect of the measurement method is that changes in external appearance, such as facial hair or glasses, should not have any effect on the results [32,33]. Weight increases or decreases in the time interval in which measurements of facial symmetry are made (eg, 6 months) should not have any relevant effects, except in extreme cases, due to the pure assessment of the maximum range of facial movements.

On the other hand, acquired or congenital asymmetries such as scars, deformities, or grafts on the face could of course contribute to a relevant falsification of the measurable range of motion. However, these can also influence established rating systems such as the Stennert's index, based on our everyday clinical experience.

The decision to configure the DFI reciprocal to Stennert´s index was based on clinical considerations. Following a regular course of healing, an increasing numerical value is graphically represented as an increasing curve. Although this is only a small detail, maybe it can contribute to a positive psychological effect for those affected.

The possibilities of the TrueDepth camera in everyday life, such as authorization of payments or unlocking of the mobile phone, are very popular among users. However, it is currently not possible to anticipate whether mobile phone manufacturers might cease the integration of depth mapping cameras into its products or replace it by another technology. Nevertheless, the fact that Apple is currently considering integrating TrueDepth technology into notebooks [74] makes it likely that TrueDepth

cameras will also be installed in future device generations. In general, the focus in Apple's marketing concept has recently been placed intensely on privacy. FaceID, which is based on TrueDepth camera technology, is an integral part of this concept [75]. In addition, even rear-facing depth mapping cameras may be added to enable improved augmented reality applications, for example. To what extent such a camera would then also be suitable for facial measurements (eg, by the physician at the patient's bedside for better documentation of the facial nerve status) is currently unclear. In summary, the medical application described here should therefore also be feasible in the future.

With the app prototype described in this work, a highly precise, easy-to-use solution for the grading of facial movement disorders via the 3D-based measurements of the TrueDepth camera system was successfully designed. It does not require any expert knowledge to operate and delivers reproducible, easy-to-understand results within a few seconds. This empowers patients suffering from facial palsy to independently determine the progress of their therapy for this functionally and cosmetically severely impairing disease. The necessary hardware is widely available, and no external equipment is needed. Nevertheless, to our knowledge, this is the first time that the TrueDepth camera system has been used for this type of medical usage.

A prospective clinical study is necessary to check the validity by correlation with established clinician-graded scoring systems and to objectively assess the usability of the app. Depending on the result of such a study, the underlying algorithm might have to be adjusted to achieve an optimal fit. The app-based design enables large-scale, decentralized data collection for multicenter clinical studies, which can advance different fields in medical science.

In addition, the export of raw data with a temporal resolution of 60 times per second opens the interesting possibility of further scientific evaluation and development of other medical fields of application. It is conceivable that the evaluation of TrueDepth data could be used for purposes such as facial training methods with real-time biofeedback or the surveillance of dysfunctions of facial movement associated with conditions such as Parkinson's disease, where spontaneous facial expressions appear to be selectively affected [76], or motor tic disorders with facial affection [77]. Automated, fast, objective tracking of facial motor skills could generate valuable data on the individual course of these diseases and on the response to different treatments.

Another possible application area would be in studies of affective psychology. For example, depressed individuals show fewer Duchenne smiles and less facial animation. This implies that automated facial analysis may prove to be useful in mental health screening too [78]. In a similar context, a smartphone-based solution has already been demonstrated with the Loki app as a proof-of-concept, which was developed in a Canadian hackathon project. By training a simple neural network with two hidden layers, it was able to map TrueDepth facial data to 4 emotions in real-time (happy, sad, angry, surprised) [79].

Furthermore, a comparable integration of machine learning into the current algorithm might even enable automated stroke screening using a smartphone in the future, as facial palsy is a very common consequence of strokes [80].

Acknowledgments

The custom Swift code used in this study is available from the corresponding author upon reasonable request. This code is to be used only for educational and research purposes. Any commercial use including the distribution, sale, lease, license, or other transfer of the code to a third party is prohibited.

Authors' Contributions

JT conceived and designed the study, provided data, performed the analysis, and wrote and edited the paper. SB provided data and edited the paper. RH provided data and edited the paper. KR provided data and wrote and edited the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exemplary measurement sequence with the app prototype. After the measurement, the absolute and relative values for the different facial regions as well as the calculated DFI score are displayed. The time course of the DFI scores can be shown in a diagram. The depicted author provided written informed consent for the video to be used for this paper.
[MP4 File (MP4 Video), 13128 KB - mhealth_v9i1e19346_app1.mp4]

Multimedia Appendix 2
Absolute and relative values measured and calculated by the app prototype and four healthy subjects.
[DOCX File , 39 KB - mhealth_v9i1e19346_app2.docx]

References

1. Saadi R, Shokri T, Schaefer E, Hollenbeak C, Lighthall JG. Depression Rates After Facial Paralysis. Ann Plast Surg 2019 Aug;83(2):190-194. [doi: 10.1097/SAP.0000000000001908] [Medline: 31232802]
2. Finkensieper M, Volk GF, Guntinas-Lichius O. [Facial nerve disorders]. Laryngorhinootologie 2012 Feb;91(2):121-142. [doi: 10.1055/s-0031-1300965] [Medline: 22318445]
3. Bascom DA, Schaitkin BM, May M, Klein S. Facial nerve repair: a retrospective review. Facial Plast Surg 2000;16(4):309-313. [doi: 10.1055/s-2000-15545] [Medline: 11460296]
4. Sullivan FM, Swan IR, Donnan PT, Morrison JM, Smith BH, McKinstry B, et al. Early Treatment with Prednisolone or Acyclovir in Bell's Palsy. N Engl J Med 2007 Oct 18;357(16):1598-1607. [doi: 10.1056/nejmoa072006]
5. Diels HJ, Combs D. Neuromuscular retraining for facial paralysis. Otolaryngol Clin North Am 1997 Oct;30(5):727-743. [Medline: 9295250]
6. Lindsay RW, Robinson M, Hadlock TA. Comprehensive facial rehabilitation improves function in people with facial paralysis: a 5-year experience at the Massachusetts Eye and Ear Infirmary. Phys Ther 2010 Mar;90(3):391-397. [doi: 10.2522/ptj.20090176] [Medline: 20093325]
7. Lindsay RW, Bhama P, Hadlock TA. Quality-of-life improvement after free gracilis muscle transfer for smile restoration in patients with facial paralysis. JAMA Facial Plast Surg 2014 Nov;16(6):419-424. [doi: 10.1001/jamafacial.2014.679] [Medline: 25275339]
8. Stennert E, Limberg CH, Frentrup KP. [An index for paresis and defective healing--an easily applied method for objectively determining therapeutic results in facial paresis (author's transl)]. HNO 1977 Jul;25(7):238-245. [Medline: 893151]
9. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg 1985 Apr;93(2):146-147. [doi: 10.1177/019459988509300202] [Medline: 3921901]
10. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg 1996 Mar;114(3):380-386. [doi: 10.1016/s0194-5998(96)70206-1] [Medline: 8649870]
11. Banks CA, Bhama PK, Park J, Hadlock CR, Hadlock TA. Clinician-Graded Electronic Facial Paralysis Assessment: The eFACE. Plast Reconstr Surg 2015 Aug;136(2):223e-230e. [doi: 10.1097/PRS.0000000000001447] [Medline: 26218397]
12. Banks CA, Jowett N, Azizzadeh B, Beurskens C, Bhama P, Borschel G, et al. Worldwide Testing of the eFACE Facial Nerve Clinician-Graded Scale. Plast Reconstr Surg 2017 Feb;139(2):491e-498e. [doi: 10.1097/PRS.0000000000002954] [Medline: 28121888]
13. Banks CA, Jowett N, Hadlock TA. Test-Retest Reliability and Agreement Between In-Person and Video Assessment of Facial Mimetic Function Using the eFACE Facial Grading System. JAMA Facial Plast Surg 2017 May 01;19(3):206-211 [FREE Full text] [doi: 10.1001/jamafacial.2016.1620] [Medline: 28006048]
14. Dusseldorp JR, van Veen MM, Mohan S, Hadlock TA. Outcome Tracking in Facial Palsy. Otolaryngol Clin North Am 2018 Dec;51(6):1033-1050. [doi: 10.1016/j.otc.2018.07.003] [Medline: 30172554]
15. Manktelow RT, Zuker RM, Tomat LR. Facial paralysis measurement with a handheld ruler. Plast Reconstr Surg 2008 Feb;121(2):435-442. [doi: 10.1097/01.prs.0000297648.90717.bd] [Medline: 18300959]
16. Pourmomeny AA, Zadmehr H, Hossaini M. Measurement of facial movements with Photoshop software during treatment of facial nerve palsy. J Res Med Sci 2011 Oct;16(10):1313-1318 [FREE Full text] [Medline: 22973325]
17. Hadlock TA, Urban LS. Toward a universal, automated facial measurement tool in facial reanimation. Arch Facial Plast Surg 2012;14(4):277-282. [doi: 10.1001/archfacial.2012.111] [Medline: 22508895]
18. Katsumi S, Esaki S, Hattori K, Yamano K, Umezaki T, Murakami S. Quantitative analysis of facial palsy using a three-dimensional facial motion measurement system. Auris Nasus Larynx 2015 Aug;42(4):275-283. [doi: 10.1016/j.anl.2015.01.002] [Medline: 25655983]
19. Guarin DL, Dusseldorp J, Hadlock TA, Jowett N. A Machine Learning Approach for Automated Facial Measurements in Facial Palsy. JAMA Facial Plast Surg 2018 Jul 01;20(4):335-337. [doi: 10.1001/jamafacial.2018.0030] [Medline: 29543955]
20. Mothes O, Modersohn L, Volk GF, Klingner C, Witte OW, Schlattmann P, et al. Automated objective and marker-free facial grading using photographs of patients with facial palsy. Eur Arch Otorhinolaryngol 2019 Dec;276(12):3335-3343. [doi: 10.1007/s00405-019-05647-7] [Medline: 31535292]
21. Barbosa J, Lee K, Lee S, Lodhi B, Cho J, Seo W, et al. Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier. BMC Med Imaging 2016 Mar 12;16:23 [FREE Full text] [doi: 10.1186/s12880-016-0117-0] [Medline: 26968938]
22. Baugh RF, Basura GJ, Ishii LE, Schwartz SR, Drumheller CM, Burkholder R, et al. Clinical practice guideline: Bell's palsy. Otolaryngol Head Neck Surg 2013 Nov;149(3 Suppl):S1-27. [doi: 10.1177/0194599813505967] [Medline: 24189771]
23. Beurskens CHG, Heymans PG. Positive effects of mime therapy on sequelae of facial paralysis: stiffness, lip mobility, and social and physical aspects of facial disability. Otol Neurotol 2003 Jul;24(4):677-681. [doi: 10.1097/00129492-200307000-00024] [Medline: 12851564]
24. Butler DP, Grobbelaar AO. Facial palsy: what can the multidisciplinary team do? J Multidiscip Healthc 2017;10:377-381 [FREE Full text] [doi: 10.2147/JMDH.S125574] [Medline: 29026314]
25. Pereira LM, Obara K, Dias JM, Menacho MO, Lavado EL, Cardoso JR. Facial exercise therapy for facial palsy: systematic review and meta-analysis. Clin Rehabil 2011 Jul;25(7):649-658. [doi: 10.1177/0269215510395634] [Medline: 21382865]

26. Ross B, Nedzelski JM, McLean JA. Efficacy of feedback training in long-standing facial nerve paresis. Laryngoscope 1991 Jul;101(7 Pt 1):744-750. [doi: 10.1288/00005537-199107000-00009] [Medline: 2062155]
27. Teixeira LJ, Soares BGDO, Vieira VP, Prado GF. Physical therapy for Bell s palsy (idiopathic facial paralysis). Cochrane Database Syst Rev 2008 Jul 16(3):CD006283. [doi: 10.1002/14651858.CD006283.pub2] [Medline: 18646144]
28. iPhone 11 Pro - Technical Specifications. Apple Inc. 2019. URL: https://www.apple.com/de/iphone-11-pro/specs/ [accessed 2020-01-27]
29. Majumder S, Aghayi E, Noferesti M, Memarzadeh-Tehran H, Mondal T, Pang Z, et al. Smart Homes for Elderly Healthcare-Recent Advances and Research Challenges. Sensors (Basel) 2017 Oct 31;17(11) [FREE Full text] [doi: 10.3390/s17112496] [Medline: 29088123]
30. Majumder S, Mondal T, Deen M. Wearable Sensors for Remote Health Monitoring. Sensors (Basel) 2017 Jan 12;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
31. About Face ID advanced technology. Apple Inc. 2020. URL: https://support.apple.com/en-us/HT208108 [accessed 2020-01-21]
32. What is the TrueDepth camera system? iPhoneFAQ. 2019. URL: https://www.iphonefaq.org/archives/976228 [accessed 2021-01-10]
33. ARKit: Apple Developer Documentation. Apple Inc. 2021. URL: https://developer.apple.com/documentation/arkit [accessed 2021-01-10]
34. Face ID weiter unerreicht: Wie Samsung zum wiederholten Mal bei der Gesichtserkennung patzt. MacTechNews. 2019. URL: https://www.mactechnews.de/news/article/Face-ID-weiter-unerreicht-Wie-Samsung-zum-wiederholten-Mal-bei-der-Gesichtserkennung-patzt-171838.html [accessed 2021-01-10]
35. philackm / ScrollableGraphView. Github. 2017. URL: https://github.com/philackm/ScrollableGraphView [accessed 2021-01-10]
36. CocoaPods. 2021. URL: https://cocoapods.org [accessed 2021-01-10]
37. 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]
38. Adour KK, Byl FM, Hilsinger RL, Kahn ZM, Sheldon MI. The true nature of Bell's palsy: analysis of 1,000 consecutive patients. Laryngoscope 1978 May;88(5):787-801. [doi: 10.1002/lary.1978.88.5.787] [Medline: 642672]
39. Bánhalmi A, Borbás J, Fidrich M, Bilicki V, Gingl Z, Rudas L. Analysis of a Pulse Rate Variability Measurement Using a Smartphone Camera. J Healthc Eng 2018;2018:4038034 [FREE Full text] [doi: 10.1155/2018/4038034] [Medline: 29666670]
40. Koenig N, Seeck A, Eckstein J, Mainka A, Huebner T, Voss A, et al. Validation of a New Heart Rate Measurement Algorithm for Fingertip Recording of Video Signals with Smartphones. Telemed J E Health 2016 Aug;22(8):631-636. [doi: 10.1089/tmj.2015.0212] [Medline: 26938673]
41. Siddiqui SA, Zhang Y, Feng Z, Kos A. A Pulse Rate Estimation Algorithm Using PPG and Smartphone Camera. J Med Syst 2016 May 11;40(5):126. [doi: 10.1007/s10916-016-0485-6] [Medline: 27067432]
42. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
43. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. Heart 2020 May 07;106(9):665-670. [doi: 10.1136/heartjnl-2019-316004] [Medline: 31911507]
44. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
45. Goel M, Saba E, Stiber M, Whitmire E, Fromm J, Larson EC, et al. SpiroCall: Measuring Lung Function over a Phone Call. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16) 2016:5675-5685. [doi: 10.1145/2858036.2858401]
46. Larson EC, Goel M, Boriello G, Heltshe S, Rosenfeld M, Patel SN. SpiroSmart: using a microphone to measure lung function on a mobile phone. Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12) 2012:280-289. [doi: 10.1145/2370216.2370261]
47. Bastawrous A. Smartphone fundoscopy. Ophthalmology 2012 Feb;119(2):432-433.e2; author reply 433. [doi: 10.1016/j.ophtha.2011.11.014] [Medline: 22305322]
48. Haddock LJ, Kim DY, Mukai S. Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. J Ophthalmol 2013;2013:518479-518475 [FREE Full text] [doi: 10.1155/2013/518479] [Medline: 24171108]
49. Lord RK, Shah VA, San Filippo AN, Krishna R. Novel uses of smartphones in ophthalmology. Ophthalmology 2010 Jun;117(6):1274-1274.e3. [doi: 10.1016/j.ophtha.2010.01.001] [Medline: 20522335]

50. Ho NRY, Lim GS, Sundah NR, Lim D, Loh TP, Shao H. Visual and modular detection of pathogen nucleic acids with enzyme-DNA molecular complexes. Nat Commun 2018 Aug 13;9(1):3238 [FREE Full text] [doi: 10.1038/s41467-018-05733-0] [Medline: 30104566]
51. Kühnemund M, Wei Q, Darai E, Wang Y, Hernández-Neuta I, Yang Z, et al. Targeted DNA sequencing and in situ mutation analysis using mobile phone microscopy. Nat Commun 2017 Jan 17;8:13913 [FREE Full text] [doi: 10.1038/ncomms13913] [Medline: 28094784]
52. Börve A, Dahlén Gyllencreutz J, Terstappen K, Johansson Backman E, Aldenbratt A, Danielsson M, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190 [FREE Full text] [doi: 10.2340/00015555-1906] [Medline: 24923283]
53. Hubiche T, Valério L, Boralevi F, Mahe E, Bodemer Skandalis C, Phan A, Research Group of the French Society of Pediatric Dermatology Groupe de Recherche de la Société Française de Dermatologie Pédiatrique. Visualization of Patients' Skin Lesions on Their Smartphones: A New Step During Dermatology Visits. JAMA Dermatol 2016 Jan 01;152(1):95-97. [doi: 10.1001/jamadermatol.2015.2977] [Medline: 26443967]
54. Kim S, Cho D, Kim J, Kim M, Youn S, Jang JE, et al. Smartphone-based multispectral imaging: system development and potential for mobile skin diagnosis. Biomed. Opt. Express 2016 Nov 28;7(12):5294. [doi: 10.1364/boe.7.005294]
55. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: A Portable Library for Melanoma Detection on Handheld Devices. Proc IEEE Int Symp Biomed Imaging 2011 Mar 30;2011:133-136 [FREE Full text] [doi: 10.1109/ISBI.2011.5872372] [Medline: 21892382]
56. Mannino RG, Myers DR, Tyburski EA, Caruso C, Boudreaux J, Leong T, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. Nat Commun 2018 Dec 04;9(1):4924 [FREE Full text] [doi: 10.1038/s41467-018-07262-2] [Medline: 30514831]
57. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss 'n' turn: smartphone as sleep and sleep quality detector. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Canada p. 477-486. [doi: 10.1145/2556288.2557220]
58. Nandakumar R, Gollakota S, Watson N. Contactless sleep apnea detection on smartphones. 2015 Presented at: 13th Annual International Conference on Mobile Systems, Applications, and Services; May 18-22, 2015; Florence, Italy p. 45-57. [doi: 10.1145/2742647.2742674]
59. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Martoni M. Monitoring sleep with a smartphone accelerometer. Sleep and Biological Rhythms 2012;10(4):287-292.
60. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
61. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
62. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
63. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
64. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile Phone-Based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
65. Ma Y, Xu B, Bai Y, Sun G, Zhu R. Infer Daily Mood using Mobile Phone Sensing. Ad Hoc & Sensor Wireless Networks 2013;20(1):133-152.
66. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
67. Gu T. Top Five Smartphone Brands by Active Devices: Samsung Is Still Top, But Chinese Brands Are on the Rise. Newzoo. 2019 Dec 4. URL: https://newzoo.com/insights/articles/top-five-smartphone-brands-by-active-devices-samsung-is-still-top-but-chinese-brands-are-on-the-rise/ [accessed 2021-01-10]
68. Lee G. Sunny Optical to supply ToF modules to next-gen Galaxy S11. TheElec. 2019 Oct 23. URL: http://en.thelec.kr/news/articleView.html?idxno=567 [accessed 2021-01-10]
69. Security certifications for SEP: Secure Key Store. Apple Inc. 2020 Sep 18. URL: https://support.apple.com/en-us/HT209632 [accessed 2021-01-10]
70. Jain V, Deshmukh A, Gollomp S. Bilateral facial paralysis: case presentation and discussion of differential diagnosis. J Gen Intern Med 2006 Jul;21(7):C7-10 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00466.x] [Medline: 16808763]

71. Gaudin RA, Jowett N, Banks CA, Knox CJ, Hadlock TA. Bilateral Facial Paralysis: A 13-Year Experience. Plast Reconstr Surg 2016 Oct;138(4):879-887. [doi: 10.1097/PRS.0000000000002599] [Medline: 27307336]
72. Stahl N, Ferit T. Recurrent bilateral peripheral facial palsy. J Laryngol Otol 1989 Jan;103(1):117-119. [doi: 10.1017/s0022215100108199] [Medline: 2921543]
73. Katsumi S, Esaki S, Hattori K, Yamano K, Umezaki T, Murakami S. Quantitative analysis of facial palsy using a three-dimensional facial motion measurement system. Auris Nasus Larynx 2015 Aug;42(4):275-283 [FREE Full text] [doi: 10.1016/j.anl.2015.01.002] [Medline: 25655983]
74. MacBook mit Face ID: Apple reicht Patentantrag ein. MacTechNews. 2020 Jan 15. URL: https://www.mactechnews.de/news/article/MacBook-mit-Face-ID-Apple-reicht-Patentantrag-ein-174166.html [accessed 2021-01-10]
75. Privacy. Apple Inc. 2021. URL: https://www.apple.com/privacy/ [accessed 2021-01-10]
76. Smith MC, Smith MK, Ellgring H. Spontaneous and posed facial expression in Parkinson's disease. J Int Neuropsychol Soc 1996 Sep 26;2(5):383-391. [doi: 10.1017/s1355617700001454] [Medline: 9375163]
77. Singer HS. Tourette's syndrome: from behaviour to biology. The Lancet Neurology 2005 Mar;4(3):149-159. [doi: 10.1016/S1474-4422(05)01012-4] [Medline: 15721825]
78. Cohn JF, De La Torre F. Automated face analysis for affective computing. In: Calvo R, D'Mello S, Gritch J, Kappas A, editors. The Oxford Handbook of Affective Computing. Oxford, England: Oxford University Press; Jan 2015.
79. Chu L, Yap K, Tannar N, Huber P. Proof-of-concept of emotion-targeted content delivery using machine learning and ARKit. GitHub. 2018. URL: https://github.com/nwhacks-loki/loki [accessed 2021-01-10]
80. Kang J, Chun MH, Choi SJ, Chang MC, Yi YG. Effects of Mirror Therapy Using a Tablet PC on Central Facial Paresis in Stroke Patients. Ann Rehabil Med 2017 Jun;41(3):347-353 [FREE Full text] [doi: 10.5535/arm.2017.41.3.347] [Medline: 28758071]

Abbreviations

3D: 3-dimensional
DFI: Digital Facial Index
eFACE: electronic, clinician-graded facial function scale

Edited by L Buis; submitted 15.04.20; peer-reviewed by O Guntinas-Lichius, A Davoudi, A Behmanesh, DX Pastor; comments to author 04.09.20; revised version received 30.10.20; accepted 30.11.20; published 26.01.21.

Please cite as:
Taeger J, Bischoff S, Hagen R, Rak K
Utilization of Smartphone Depth Mapping Cameras for App-Based Grading of Facial Movement Disorders: Development and Feasibility Study
JMIR Mhealth Uhealth 2021;9(1):e19346
URL: http://mhealth.jmir.org/2021/1/e19346/
doi:10.2196/19346
PMID:33496670

Original Paper

Privacy Policy Compliance of Chronic Disease Management Apps in China: Scale Development and Content Evaluation

Zhenni Ni[1,2,3*], BM; Yiying Wang[4*], LLB; Yuxing Qian[1,2,3*], MSc

[1]School of Information Management, Wuhan University, Wuhan, China

[2]Center for Studies of Information Resources, Wuhan University, Wuhan, China

[3]Big Data Institute, Wuhan University, Wuhan, China

[4]School of International Law, Shanghai University of Political Science and Law, Shanghai, China

[*]all authors contributed equally

Corresponding Author:
Zhenni Ni, BM
School of Information Management
Wuhan University
LuoJiaShan
WuChang district
Wuhan
China
Phone: 86 13667195159
Email: Jennie_N@whu.edu.cn

Abstract

Background: With the development of mobile health (mHealth), chronic disease management apps have brought not only the possibility of reducing the burden of chronic diseases but also huge privacy risks to patients' health data.

Objective: The purpose of the study was to analyze the extent to which chronic disease management apps in China comply with the Personal Information Security Specification (PI Specification).

Methods: The compliance of 45 popular chronic disease management apps was evaluated from the perspective of the information life cycle. To conduct a fine-grained evaluation, a scale based on the PI Specification was developed. Finally, 6 level 1 indicators, 22 level 2 indicators, and 61 level 3 indicators were defined.

Results: There were 33/45 apps (73%) with a privacy policy, and the average score of these apps was 40.4 out of 100. Items of level 1 indicators with high scores included general characteristics (mean 51.9% [SD 28.1%]), information collection and use (mean 51.1% [SD 36.7%]), and information sharing and transfer (mean 50.3% [SD 33.5%]). Information storage and protection had the lowest compliance with PI Specification (mean 29.4% [SD 32.4%]). Few personal information (PI) controllers have stated how to handle security incidents, including security incident reporting (7/33, 21%), security incident notification (10/33, 30%), and commitment to bear corresponding legal responsibility for PI security incidents (1/33, 3%). The performance of apps in the stage of information destruction (mean 31.8% [SD 40.0%]) was poor, and only 21% (7/33) apps would notify third parties to promptly delete PI after individuals cancelled their accounts. Moreover, the scoring rate for rights of PI subjects is generally low (mean 31.2% [SD 35.5%]), especially for obtaining copies of PI (15%) and responding to requests (25%).

Conclusions: Although most chronic disease management apps had a privacy policy, the total compliance rate of the policy content was low, especially in the stage of information storage and protection. Thus, the field has a long way to go with regard to compliance around personal privacy protection in China.

(JMIR Mhealth Uhealth 2021;9(1):e23409) doi:10.2196/23409

KEYWORDS

mHealth; noncommunicable diseases; content analysis

Introduction

Background

Chronic diseases, such as diabetes and hypertension, are a major global health issue affecting many countries [1]. Fortunately, the booming of mobile health (mHealth) offers opportunities for chronic diseases prevention, treatment, and daily self-management. The health benefits of mHealth interventions for patients with chronic diseases have been demonstrated [2]; mHealth apps can be used to collect and monitor health data [3,4], promote and support self-management [5,6], and provide medication and appointment reminders [7]. Different from other types of mHealth apps, such as online registration and online consultation, chronic disease management apps allow individuals to generate large quantities of data about their lifestyle, introducing risks to the security and privacy of patient data.

Considering the potential negative effects of security breaches of health data systems, such as social stigma, damage to reputation, and fraud in the health system [8], privacy has become an important factor discouraging patients from using mHealth apps for health care [9-11]. Different from other kinds of disease management, the care of chronic diseases requires patients to regularly track key health indicators. It means that protecting the safety and privacy of personal information (PI) is crucial for chronic disease management apps. A few studies were related to the privacy policy of chronic disease management apps. These studies mainly involved 3 focal aspects: the quality assessment [12,13], the complexity analysis of app privacy policies [14], and the security analysis [15]. Although the above 3 aspects involved privacy policies, the evaluations were relatively rough.

As for the evaluation criteria, various standards were used to evaluate the privacy of mHealth apps. Most papers established evaluation indicators based on the existing literature [16,17] or authors' criteria [18-20]. The most common items in the evaluation criteria included stating processing purposes, determining the recipient of personal data, the existence of the data rights of the individuals, and the existence of privacy policies. Although a few papers on the privacy assessment of mHealth apps were based on laws or regulations, such as General Data Protection Regulation (GDPR), Fair Information Practices (FIPS) [18,21], some of them proposed a set of items to check the compliance of laws or regulations [20,22].

In China, the Information Security Technology–Personal Information Security Specification (GB/*t* 35273-2020) (PI Specification) came into effect on October 1, 2020 [23]. This specification, also as the standard basis for apps' privacy certification, lays out granular guidelines for how personal data should be collected, used, and shared. Besides, it provides a template of PI protection policy in the form of attachments. Although PI Specification is a national voluntary standard instead of a mandatory standard, it provides a reference for the industry. However, the compliance with PI Specification of mHealth apps remains unclear. In each step of the information life cycle, the patient's PI is at risk of leakage, such as collection, storage, usage, sharing, destruction, and so on. Therefore, it is necessary to review the compliance of the privacy policy of mHealth apps based on PI Specification from various stages of the information life cycle, especially for chronic disease management apps that have insufficient privacy assessment.

Objectives

This study aimed to evaluate the compliance of privacy policies of chronic disease apps with the PI Specification from the perspective of the information life cycle. Specifically, this study can provide answers to the following 2 research questions: (1) To what extent do chronic disease apps comply with PI Specification 2020? (2) Among the various stages of the information life cycle, which stage has the weakest privacy policy protection?

Methods

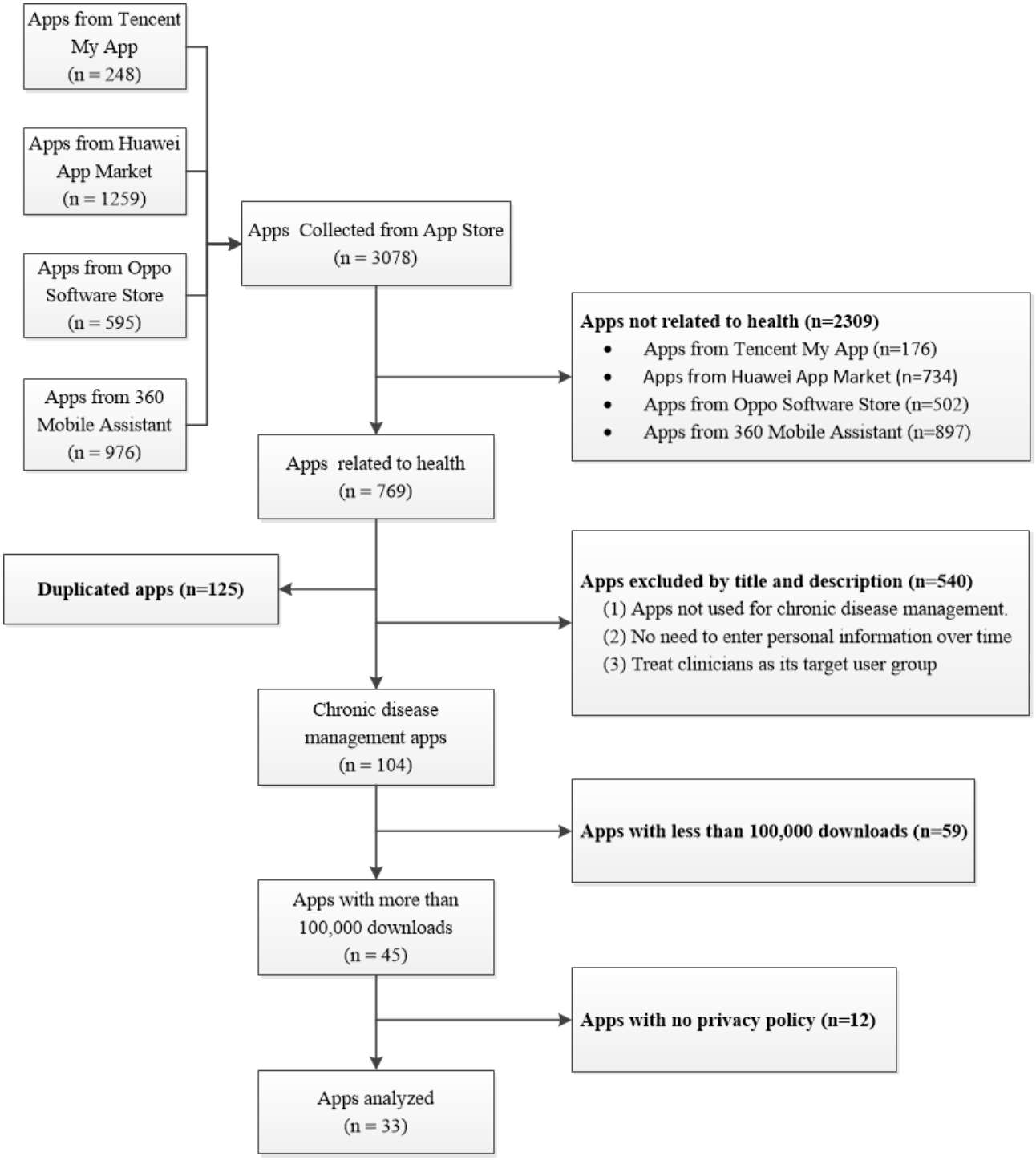
Apps Selection

Considering the popularity of Android in China [24], this study investigated mHealth apps in Android app stores. The top 4 Android app stores were selected, which accounts for 61.0% of the Chinese Android market [25], including Tencent My App (26.0%) [26], Huawei App Market (15.1%) [27], Oppo Software Store (10.2%) [28], and 360 Mobile Assistant (9.7%) [29]. The apps returned by queries for "noncommunicable diseases," "chronic disease," "diabetes," "blood pressure," "hypertension," "heart disease," "kidney," "cardiovascular," "asthma," "respiratory disease," or "cancer" were included in the set of chronic diseases management apps.

This search was conducted on October 2, 2020. Our sample was filtered based on the title and description in the app stores. The app met inclusion criteria if it (1) was in Chinese; (2) required the input of PI over time; (3) had the general public as its target user group rather than clinicians; and (4) had over 100,000 downloads. The authors saved all privacy policies as text files and recorded the downloads, update time, and disease category.

A total of 45 apps met the inclusion criteria (Figure 1). Among them, 12/45 apps (27%) had no privacy policy. Excluding apps without a privacy policy, the remaining 33 privacy policies were analyzed.

Figure 1. Flow chart of the search strategy.



Scale Development and Scoring

The level 1 and level 2 evaluation indicators are shown in Textbox 1. Level 3 evaluation indicators for privacy policies are listed in Multimedia Appendix 1. Based on the information life cycle, 6 level 1 indicators were developed, including information collection and use, information storage and protection, information sharing and transfer, information destruction, general characteristics of privacy policies, and rights of PI subjects. There are 22 items on level 2 indicators and 61 items on level 3 indicators. For each level 3 indicator, a brief explanation, example sentences, and corresponding clauses of PI Specification are listed in Multimedia Appendix 2.

Each level 3 indicator was scored as 1 point if the privacy policy complies with the indicator and 0 otherwise. Scoring rate of each level 3 indicator was defined as the percentage of the number of apps scored 1 point in the total sample. Scoring rate of each level 2 indicator was the average of all level 3 indicators under that level 2 indicator. Scoring rate of each level 1 indicator, which indicated the compliance of apps in the corresponding stage of the information life cycle, was the average of all level 2 indicators under that level 1 indicator. For each app, the sum of all level 3 indicators scores was converted into a percentage system as a final score; the final score represented the compliance of the app. Bar graphs are used to

visualize the degree of policy compliance. The ordinate of bar graphs is the scoring indicators, including level 3 and level 2 indicators, and the scoring rate of level 2 indicators. The abscissa is the scoring rate of level 3 indicators. In order to more intuitively reflect the scores of the level 2 indicators, we use different colors to visualize each level 2 indicator; if the scoring rate is close to the average score, it is yellow; if the scoring rate is close to the minimum value, it is red; If the scoring rate is close to the maximum value, it is green.

Initially, 2 raters (ZN and YW) independently reviewed 21% (7/33) of randomly selected apps to assess the level of agreement; the Kappa-Cohen Index was 0.87, which denoted an almost perfect agreement. Then, 2 raters (ZN and YW) discussed indicators with inconsistent scores, and each rater analyzed half of the remaining apps after the standard was unified.

Textbox 1. Level 1 and level 2 evaluation indicators for privacy policies.

1. General characteristics

- App scope
- Policy disclosure
- Policy updates

2. Information collection and use

- Information collection and usage rules for business functions
- Personal sensitive information

3. Information storage and protection

- Storage security
- The handling of security incidents

4. Information sharing and transfer

- Entrusted processing
- Sharing of PI
- Transfer of PI
- Public disclosure of PI
- Cross-border transmission

5. Information destruction

- Storage time limit
- Data deletion and anonymization

6. Rights of PI subjects

Results

Sample Distribution

The basic characteristics of these apps are presented in Table 1. The types of chronic diseases targeted by apps mainly include diabetes (11/45, 24%), hypertension (4/45, 9%), heart disease (4/45, 9%), cancer (2/45, 4%), and comprehensive chronic disease management (19/45,42%). The comprehensive chronic disease management app referred to providing users with long-term, multifaceted chronic disease prevention and treatment services that were not targeted at specific chronic disease. Besides, it included a small number of apps for other types of chronic diseases (5/45, 11%), such as asthma, chronic kidney disease, and chronic skeletal muscle diseases. Most apps (30/45, 67%) had between 100,000 and 1,000,000 downloads; 73% (33/45) of apps were updated in 2020.

Table 1. Sample distribution of chronic disease management apps (N=45).

Category	Count, n (%)
Disease category	
Diabetes	11 (24)
Hypertension	4 (9)
Heart disease	4 (9)
Cancer	2 (4)
Comprehensive	19 (42)
Others	5 (11)
Downloads	
100,000-1,000,000	30 (67)
1,000,000-10,000,000	11 (24)
10,000,000	4 (9)
Updated	
2014-2016	2 (4)
2017-2019	10 (22)
2020	33 (73)

Compliance Evaluation

The average score of 33 apps was 40.4 out of 100, and the degree of dispersion was very high (SD 31.2). The evaluation results on level 1 indicators of privacy policies are presented in Figure 2. The most complied-with items in level 1 indicators were the following: general characteristics (mean 51.9% [SD 28.1%]), information collection and use (mean 51.1% [SD 36.7%]), and information sharing and transfer (mean 50.3% [SD 33.5%]). However, some indicators had a low degree of overall compliance, such as information storage and protection (mean 29.4% [SD 32.4%]), information destruction (mean 31.8% [SD 40.0%]), and rights of PI subjects (mean 31.2% [SD 35.5%]). The name and evaluation results of each app are listed in Multimedia Appendix 3.

The scoring rate for level 2 indicators ranged from 15.2% to 75.8%, with an average of 40.4%. We visualized the evaluation results with bar graphs, in which the color of bars indicates the scoring rate of level 2 indicators (the value in parentheses) and the length of bars indicates the scoring rate of level 3 indicators.

The general characteristics of privacy policy reflect its openness, readability, and timeliness of updates. Compliance evaluation results of the privacy policies general characteristics are shown in Figure 3. Some level 2 indicators scored high, such as policy updates (59%) and disclosure (58%). More than one-half of the apps promised to notify users (19/33, 58%) and obtain the explicit consent of PI subjects again (17/33, 52%) if the policy was updated. As for policy disclosure, although most apps provided independent (20/33, 61%) and easily accessible (27/33, 82%) privacy policies, only a few apps (10/33, 30%) had a clear logical structure and provided a directory summary. In terms of scope, a few apps (9/33, 27%) marked the update date or effective time of the privacy policy, which indicated that the timeliness of policy updates was low.

Figure 2. The scoring rate of chronic disease management apps on level 1 indicators. PI: personal information.

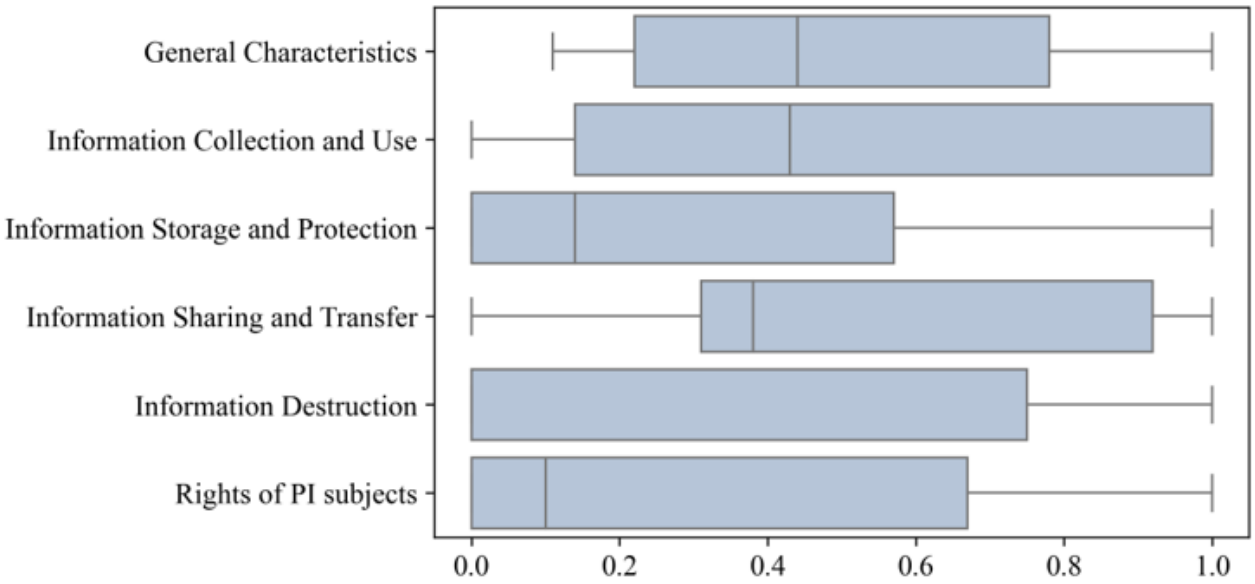
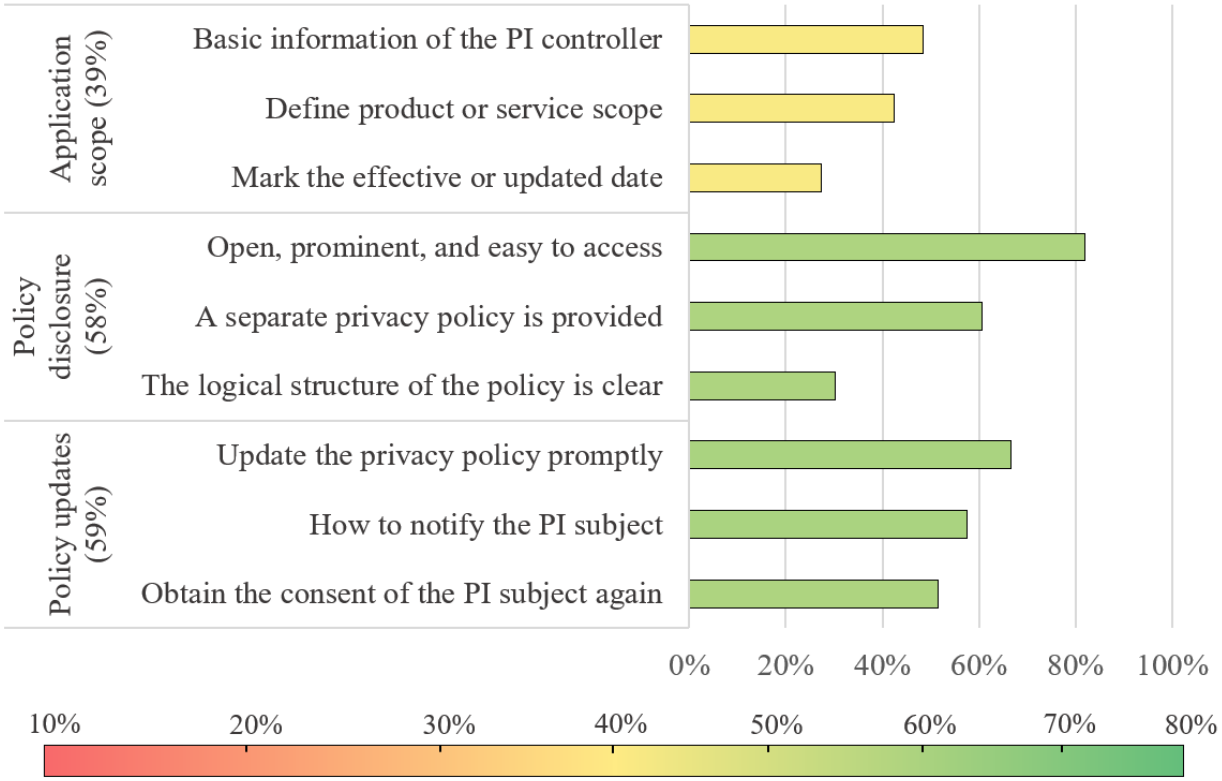


Figure 3. Compliance evaluation results of the privacy policies general characteristics. PI: personal information.



Compliance evaluation results in the stage of information collection and use, and the stage of information storage and protection are presented in Figure 4. In the information collection and use stage, the scoring rate of all level 2 indicators reached the average, and the overall compliance degree was relatively high. Because the research object of this article was chronic disease management apps, all apps in this research involved the collection and processing of personally sensitive information. However, in terms of personal sensitive information, only 30% (10/33) of apps marked personal sensitive information prominently.

Although the compliance level of storage security was close to the average (38%), most apps (28/33, 85%) did not inform PI subjects the security agreement they followed and the certification they obtained. The compliance level of the handling of security incidents (18%) was far below the average. Among the 33 apps, only 1 app (3%) promised to bear corresponding responsibilities if a security incident occurred. In addition, no more than one-third of apps described how to inform PI subjects after a security incident (10/33, 30%), and whether they would report it truthfully to government organizations (7/33, 21%).

Compliance evaluation results in the stage of information sharing and transfer, and the stage of information destruction are shown in Figure 5. Only 24% (8/33) of apps informed the type of shared information and 33% (11/33) of apps informed the security measures taken before sharing, such as anonymization or deidentification. Entrusted processing scored low; only 27% (9/33) of apps stated that they would supervise the entrusted party by establishing the third-party's responsibilities and duties through contract or other such means. The 2 level 2 indicators of the information destruction stage, namely, storage time limit (35%) and data deletion and anonymization (29%), were all lower than the average scoring rate. Especially if PI subjects request to delete user data, only 21% (7/33) of apps would notify third parties to promptly delete their PI.

Most privacy policies had a low scoring rate for the indicators related to rights of PI subjects (Figure 6), especially the right to obtain a copy of PI, which was only 15% (5/33). Scores for level 2 indicators such as complaint management (29%) and responding to requests (26%) were far below the average, which meant that most apps did not pay attention to the handling of user requests and complaints.

Figure 4. Evaluation results in the stage of information collection and use, and the stage of information storage and protection. PI: personal information.

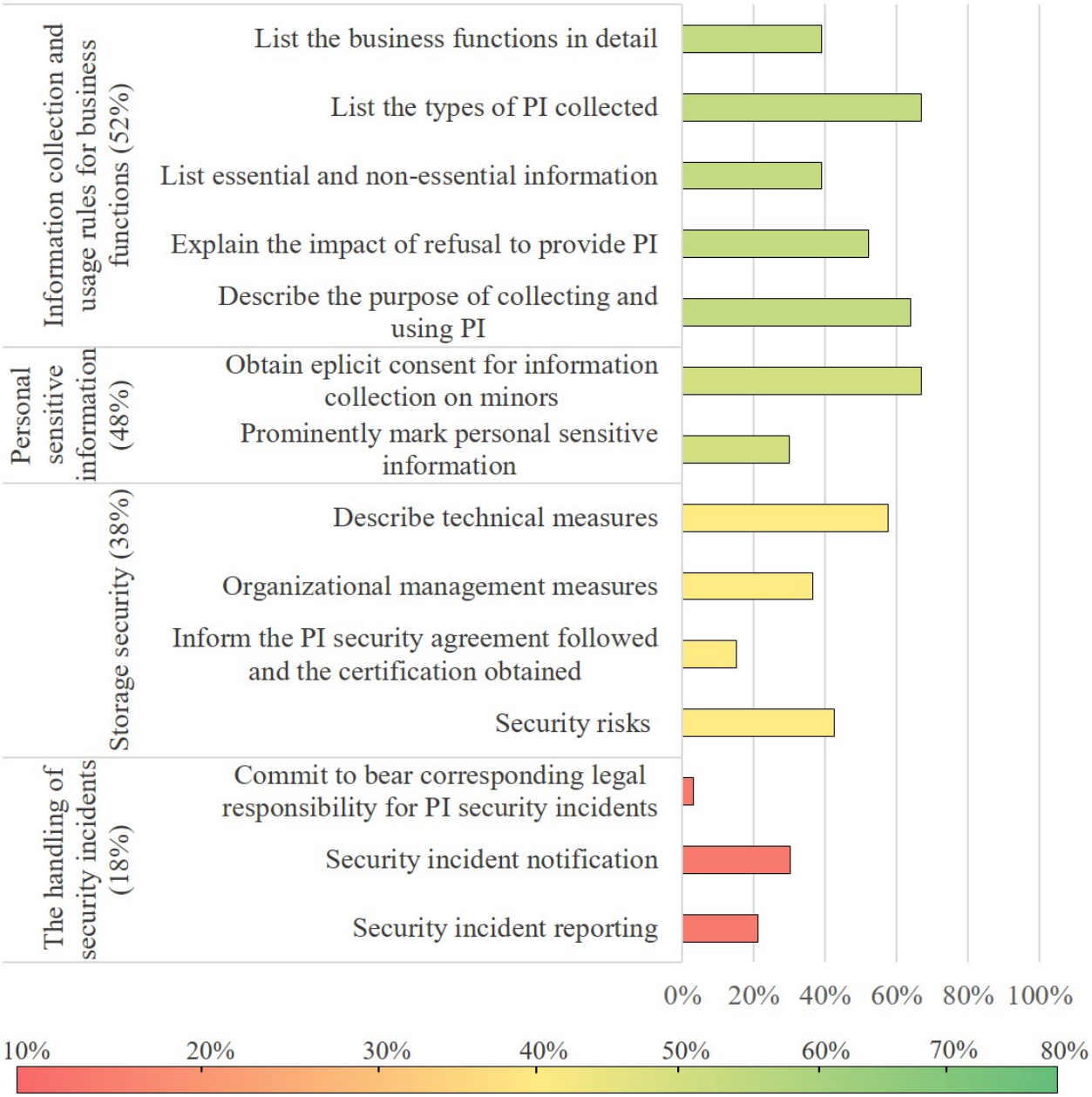


Figure 5. Evaluation results in the stage of information sharing and transfer, and the stage of information destruction. PI: personal information.

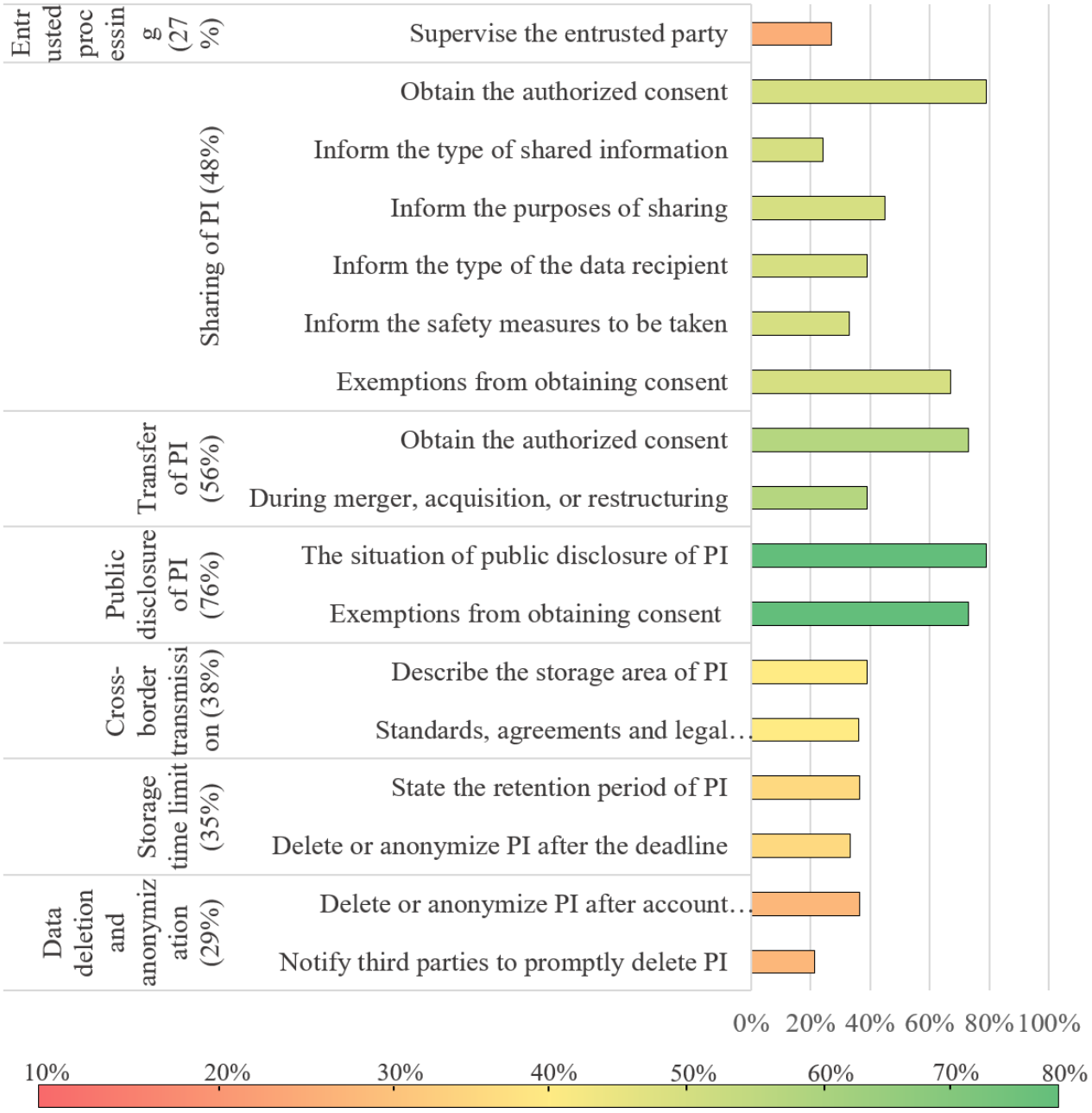
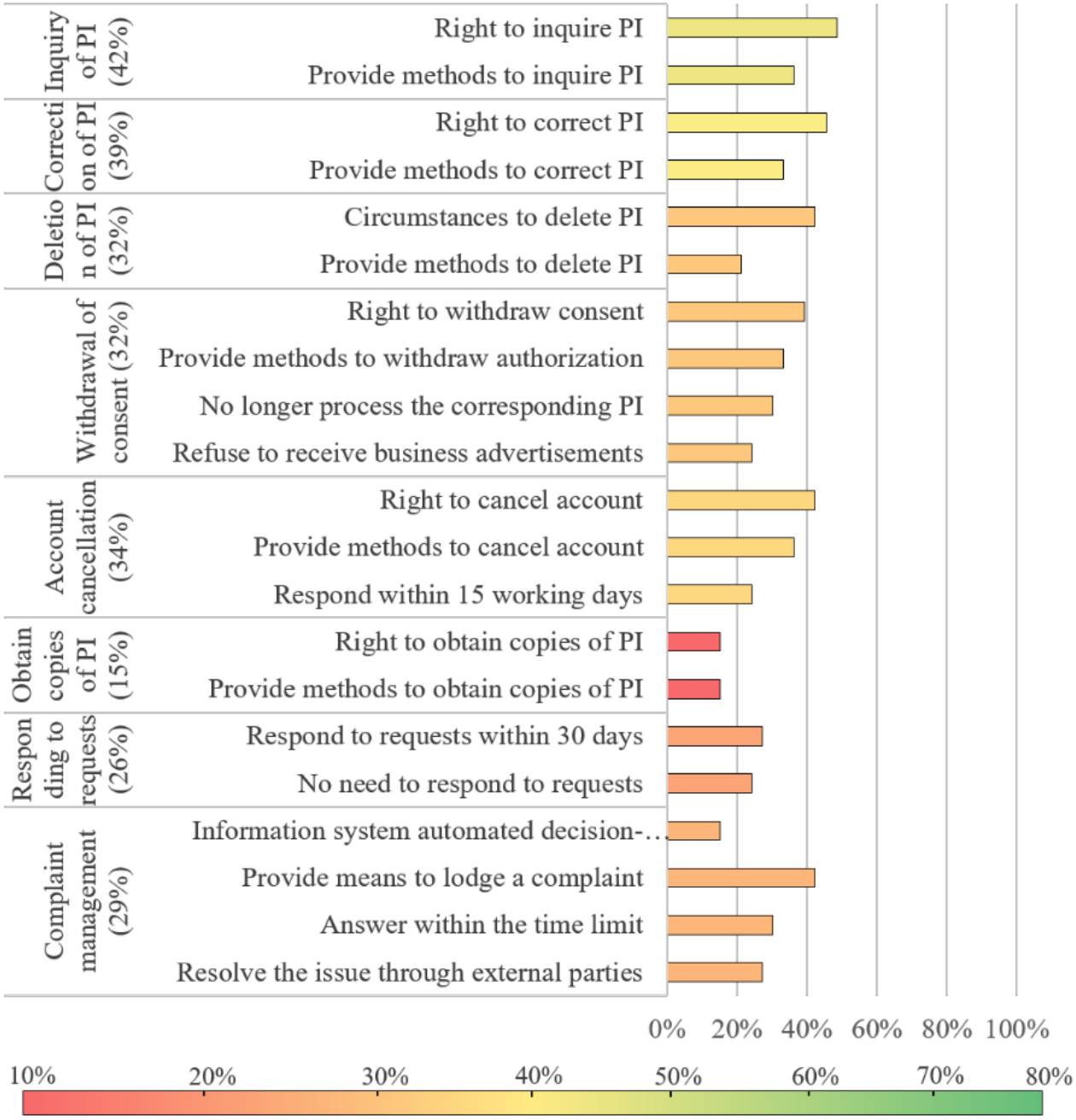


Figure 6. Compliance evaluation results of the right of PI subjects. PI: personal information.



Discussion

Key Findings

In this study, we proposed a scale based on PI Specification 2020 for assessing the compliance of China's chronic disease apps privacy policies from various stages of the information life cycle. Fu and Zhao [30] analyzed the privacy policies of 20 mHealth apps in China based on PI Specification 2017. In their study, the privacy policies were analyzed from 6 aspects, including information collection, cookies and other related technologies, PI storage and protection, PI sharing, PI processing rights, and minor information protection. However, their study did not conduct a fine-grained quantitative analysis and evaluation of each item and it could not reveal the app's compliance with specific articles in PI Specification. In this paper, 6 level 1 indicators, 22 level 2 indicators, and 61 level 3 indicators were defined and a fine-grained evaluation was conducted. PI controllers and subjects can use the scale to obtain a percentual score that defines the compliance of privacy policies.

According to the results, most of the apps collected in the initial sample (33/45, 73%) included a privacy policy, which was similar to a previous assessment of cancer apps by Benjumea et al [22] who found that 71% of the apps in their sample had a privacy policy. Considering that the prevalence of privacy policies for high-download apps is significantly higher than that of low-download apps (high downloads: 15/17, 88%; and low downloads: 33/64, 52%; *P*=.006) [31], our result might be higher than the actual situation.

Regarding scores, only 39% (13/33) of apps in our sample had a score greater than or equal to 40 points, with an average score of 40.4 out of 100 (SD 31.2), which indicated that the majority of chronic disease management apps in China had low compliance with PI Specification 2020. This result is consistent with the prior finding by Fu and Zhao [30], who determined that most mHealth apps in China did not meet the requirements of PI Specification. Benjumea et al [22] analyzed the privacy policies of 31 cancer Android apps from the Google Play website (Spanish version) and obtained an average score of 50.5 points; in Hutton et al [21], the average score for 64 self-tracking mHealth apps from Google Play was 46.2% (SD 24.3%). These differences might also be the result of different mHealth app types, evaluation scales, and even normative background (Hutton et al [21] refer to GDPR, FIPS, and usability, whereas Benjumea et al [22] refer to GDPR). What we compared is the degree to which apps complied with local laws or regulations, rather than the degree to which they protected the privacy of users. Thus, according to the evaluation results, the compliance of chronic disease management apps in China Android app stores might be slightly lower than that of mHealth apps in Google Play.

In terms of general characteristics, policy disclosure and policy updates are the basic prerequisites for a privacy policy to effectively protect the legal rights of PI subjects. The level 3 indicators under policy disclosure and policy updates can maintain a scoring rate of 57%-58%, which indicated that most PI controllers had a basic awareness of protecting user privacy. However, only 48% (16/33) of apps introduce the basic information of the PI controller in the privacy policies, which is far lower than a previous study (77%) [22].

In the stage of information collection and use, 64% (21/33) of apps stated the purpose of collecting and using PI, which was in line with the result (61%) of Hutton et al [21]; 52% (17/33) of apps described the impact of refusal to provide PI, which was far higher than the result (27%) of Benjumea et al [22]. According to PI Specification Article 5.5, if the app involves the collection of personal sensitive information, the PI controller should clearly mark or highlight the information. However, only 30% (10/33) of apps prominently marked personal sensitive information in their privacy policies.

Information sharing has always been a hotspot in privacy policy analysis. Robillard et al [32] found that 68% of privacy policies stated that users' PI may be shared with third parties, whereas only 10% of apps stated that users' PI would not be shared without their consent. In this paper, the majority of apps with a privacy policy that we assessed were highly compliant with PI Specification in data sharing (48%), transmission (56%), and public disclosure (76%). In terms of the consent of PI subjects, considerable proportions of privacy policies mentioned that they would obtain the consent of PI subjects before sharing (26/33, 79%), transfer (24/33, 73%), and public disclosure (24/33, 73%) PI. While most apps would obtain the consent of PI subjects before sharing PI, no more than one-fourth of apps informed the type of PI they would share. Furthermore, during the information sharing and transfer stage, the most worrying issue was the lack of safety measures (11/33, 33%) and supervision of third parties (9/33, 27%), which brought serious security risks to PI of patients.

Among the stages of the information life cycle, the stage of information storage and protection had the lowest compliance with PI Specification. According to Zhou et al [11], most users did have concerns about their privacy when using mHealth apps and expected the apps to take a variety of security measures, such as regular password updates, remote wipe, user consent, and access control. However, according to our assessment, approximately two-thirds of chronic disease management apps lacked the description of security measures in the level of organization management. Concerningly, only few PI controllers (18%) have stated how to handle security incidents, such as security incident reporting, security incident notification, and commitment to bear corresponding legal responsibility for PI security incidents.

The timely destruction of PI is essential to the privacy of patients. Few privacy policies complied with PI Specification in terms of the storage time limit (35%) and the deletion or anonymization of PI after account cancellation (29%). One noteworthy point here was that only 21% (7/33) of chronic disease management apps would notify third parties to promptly delete PI after PI subjects cancelled their accounts. According to PI Specification [23], the PI retention period should be the shortest time needed to achieve the purpose (Article 6.1); after the retention period is exceeded or the account is cancelled, PI controllers should carry out data deletion or anonymization (Article 6.1, Article 8.5). Judging from the assessment results of this study, the performance of apps in the stage of information destruction was far from reaching the requirements of PI Specification.

The scoring rate for rights of PI subjects is generally low, especially for obtaining copies of PI (15%) and responding to requests (25%), which was consistent with a previous study [21]. Furthermore, during our evaluation, we noticed that compared with the description of rights of PI subjects, the scoring rate of how to exercise rights of PI subjects is usually lower. For example, 48% (16/33) of apps stated the right of PI inquiry, whereas only 36% (12/33) of apps provided methods to inquire PI. These findings demonstrated that most Android chronic disease management apps in China can hardly guarantee the exercise of patients' rights.

Implications and Recommendations

The contributions of this study are threefold. First, we developed a new scale based on PI Specification. From the perspective of information life cycle management, the compliance of privacy policies can be evaluated systematically, and the scale can be generalizable to other kinds of apps in China. Based on our scale, app operators can also conduct a fine-grained self-assessment of their app privacy policies. Second, through the analysis of privacy policies, physicians and patients could better understand what information patients provide to the app companies and the potential risk of providing this information to non–health care providers, especially in terms of information storage and protection. Moreover, we investigated and assessed the current state of practice in chronic disease management apps regarding the protection of health-related data. The indicators in this paper were based on the PI Specification 2020, and findings presented in this article could provide insights into the

implementation of the new specification in China. Personal health information is highly sensitive and the leakage of daily health data may cause negative effects [8]. In this regard, we would like to make the following recommendations:

First, improve the readability of the privacy policy. The results from a 2018 study [33] suggested that privacy policies are not comprehensible to most adults. Thus, it is of great significance for apps to make their privacy policies shorter and simpler so that PI subjects can understand it. Second, strengthen government supervision and industry self-regulation. The Personal Information Protection Law of the People's Republic of China (Draft Law) was released for seeking opinions from the public on October 21, 2020 [34]. Different from the PI Specification, which is a national recommended standard instead of a mandatory standard, the promulgation and implementation of the Personal Information Protection Law will provide strong legal support for the protection of personal privacy and user rights. Moreover, it is important to pay attention to the positive effects of mHealth industry self-discipline and encourage mHealth industry organizations to draft industry rules to collect and use personal health information legally.

Limitations

First, our indicators may not be practical for apps in some special cases. For example, all level 2 indicators under "Sharing of PI" cannot be evaluated if the app does not share any PI. Assigning 1 point or 0 points, in this case, would overrate or underrate the privacy policy, respectively. Second, although we assessed the compliance of the privacy policies, we did not conduct a technical audit to evaluate if the data handling procedures outlined in the policy are implemented. It is reported that the disclosures regarding third-party data transmission do not match actual behavior [16]. Thus, future work can explore the correspondence between privacy disclosures and how apps for chronic disease handle personal data.

Conclusions

Despite these limitations, our findings demonstrated a general lack of compliance regarding the handling of users' health data submitted to chronic disease management apps. Although most chronic disease management apps had a privacy policy, the total compliance rate of the policy content was low. In addition, few apps could handle security incidents according to the requirements of PI specification. Importantly, it was difficult for PI subjects to exercise their rights in accordance with the privacy policies, especially in the stage of information destruction. Overall, our findings suggest the field has a long way to go with regard to compliance around data handling in China. Only by calling attention to this large need, can we change the practices and create a safer online environment for users' daily health information.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 71661167007, 71420107026).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluation indicators for privacy policies.
[DOCX File , 25 KB - mhealth_v9i1e23409_app1.docx]

Multimedia Appendix 2
Evaluation guide.
[XLSX File (Microsoft Excel File), 22 KB - mhealth_v9i1e23409_app2.xlsx]

Multimedia Appendix 3
List of mHealth apps names and evaluation results.
[XLSX File (Microsoft Excel File), 13 KB - mhealth_v9i1e23409_app3.xlsx]

References

1. World Health Organization. Noncommunicable diseases: Secondary Noncommunicable diseases. 2018 Jun 1. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2020-12-21]
2. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]
3. Aubourg T, Demongeot J, Provost H, Vuillerme N. Circadian Rhythms in the Telephone Calls of Older Adults: Observational Descriptive Study. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e12452 [FREE Full text] [doi: 10.2196/12452] [Medline: 32130156]

4. Huang R, Liu N, Nicdao M, Mikaheal M, Baldacchino T, Albeos A, et al. Emotion sharing in remote patient monitoring of patients with chronic kidney disease. J Am Med Inform Assoc 2020 Feb 01;27(2):185-193 [FREE Full text] [doi: 10.1093/jamia/ocz183] [Medline: 31633755]
5. Korpershoek YJG, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JCA. User-Centered Design of a Mobile Health Intervention to Enhance Exacerbation-Related Self-Management in Patients With Chronic Obstructive Pulmonary Disease (Copilot): Mixed Methods Study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
6. Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, et al. Evaluation of Digital Technologies Tailored to Support Young People's Self-Management of Musculoskeletal Pain: Mixed Methods Study. J Med Internet Res 2020 Jun 05;22(6):e18315 [FREE Full text] [doi: 10.2196/18315] [Medline: 32442143]
7. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile Apps for Medication Management: Review and Analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
8. Kotz D. A threat taxonomy for mHealth privacy. New York, NY: IEEE; 2011 Presented at: 2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011); January 4-8, 2011; Bangalore, Karnataka, India p. 1-6. [doi: 10.1109/COMSNETS.2011.5716518]
9. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WS, et al. Consumer Attitudes and Perceptions on mHealth Privacy and Security: Findings From a Mixed-Methods Study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
10. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
11. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
12. Quevedo Rodríguez A, Wägner AM. Mobile phone applications for diabetes management: A systematic review. Endocrinol Diabetes Nutr 2019 May;66(5):330-337. [doi: 10.1016/j.endinu.2018.11.005] [Medline: 30745121]
13. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov;20(4):e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
14. Powell AC, Singh P, Torous J. The Complexity of Mental Health App Privacy Policies: A Potential Barrier to Privacy. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e158 [FREE Full text] [doi: 10.2196/mhealth.9871] [Medline: 30061090]
15. Knorr K, Aspinall D, Wolters M. On the Privacy, Security and Safety of Blood Pressure and Diabetes Apps. In: ICT Systems Security and Privacy Protection. SEC 2015. IFIP Advances in Information and Communication Technology (Vol. 455). Cham: Springer; 2015 Presented at: SEC 2015: 30th IFIP TC 11 International Conference; May 26-28, 2015; Hamburg, Germany p. 571-584. [doi: 10.1007/978-3-319-18467-8_38]
16. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
17. Bachiri M, Idri A, Fernández-Alemán JL, Toval A. Evaluating the Privacy Policies of Mobile Personal Health Records for Pregnancy Monitoring. J Med Syst 2018 Jun 29;42(8):144. [doi: 10.1007/s10916-018-1002-x] [Medline: 29959535]
18. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/access.2018.2799522]
19. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy Issues in Smartphone Applications: An Analysis of Headache/Migraine Applications. Headache 2018 Jul;58(7):1014-1027 [FREE Full text] [doi: 10.1111/head.13341] [Medline: 29974470]
20. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy Assessment in Mobile Health Apps: Scoping Review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
21. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the Privacy of mHealth Apps for Self-Tracking: Heuristic Evaluation Approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
22. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Assessment of the Fairness of Privacy Policies of Mobile Health Apps: Scale Development and Evaluation in Cancer Apps. JMIR Mhealth Uhealth 2020 Jul 28;8(7):e17134 [FREE Full text] [doi: 10.2196/17134] [Medline: 32720913]
23. National Information Security Standardization Technical Committee. Information security technology: Personal information security specification (GB/*t* 35273-2020). 2020. URL: https://www.chinesestandard.net/PDF/English.aspx/GBT35273-2020 [accessed 2020-12-21]
24. China Internet Network Information Center. Statistical Report on China Internet Development: Secondary Statistical Report on China Internet Development. 2020 Apr 28. URL: http://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm [accessed 2020-12-21]

25. AppInChina. The AppInChina App Store Index. 2020 Jul 8. URL: https://www.appinchina.co/market/app-stores/ [accessed 2020-12-21]
26. Tencent. Tencent My App. URL: https://android.myapp.com/ [accessed 2020-12-21]
27. Huawei. Huawei App Market. URL: http://app.hicloud.com/ [accessed 2020-12-21]
28. BBK Electronics. Oppo Software Store. URL: https://store.oppomobile.com/ [accessed 2020-12-21]
29. Qihoo 360. 360 Mobile Assistant. URL: http://zhushou.360.cn/ [accessed 2020-12-21]
30. Fu S, Zhao A. Research on user privacy protection policies of health apps - based on information security technology - personal information security specification [in Chinese]. Libr Tribune 2019;39(12):109-118.
31. Huang Z, Lum E, Car J. Medication Management Apps for Diabetes: Systematic Assessment of the Transparency and Reliability of Health Information Dissemination. JMIR Mhealth Uhealth 2020 Feb 19;8(2):e15364 [FREE Full text] [doi: 10.2196/15364] [Medline: 32130163]
32. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Sep;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
33. Das G, Cheung C, Nebeker C, Bietz M, Bloss C. Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability. JMIR Mhealth Uhealth 2018 Jan 04;6(1):e3 [FREE Full text] [doi: 10.2196/mhealth.7626] [Medline: 29301737]
34. The National People's Congress of the People's Republic of China. The draft of personal information protection law and other draft laws are open for comments. URL: http://www.npc.gov.cn/npc/c30834/202010/9f67b926535948719c2a0b0220dce9ed.shtml [accessed 2020-12-21]

Abbreviations

FIPS: Fair Information Practices
GDPR: General Data Protection Regulation
mHealth: mobile Health
PI: personal information
PI Specification: Information Security Technology–Personal Information Security Specification (GB/*t* 35273-2020)

Edited by L Buis; submitted 11.08.20; peer-reviewed by J Benjumea, S Lalla-Edward; comments to author 18.09.20; revised version received 10.11.20; accepted 02.12.20; published 28.01.21.

Please cite as:
Ni Z, Wang Y, Qian Y
Privacy Policy Compliance of Chronic Disease Management Apps in China: Scale Development and Content Evaluation
JMIR Mhealth Uhealth 2021;9(1):e23409
URL: http://mhealth.jmir.org/2021/1/e23409/
doi:10.2196/23409
PMID:33507159

Original Paper

Mobile Apps for Dental Caries Prevention: Systematic Search and Quality Evaluation

Rebecca Chen[1,2], BOH, MIPH; Karla Santo[2], PhD, MD; Grace Wong[3], PhD, MHlthProm, GradCert HPEd; Woosung Sohn[1], PhD, DrPH, DDS; Heiko Spallek[1], PhD, DMD, MSBA; Clara Chow[2*], PhD, MD; Michelle Irving[1*], PhD

[1]Sydney School of Dentistry, Faculty of Medicine and Health, The University of Sydney, Westmead, Australia

[2]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Westmead, Australia

[3]Northern Sydney LHD, NSW Health, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Rebecca Chen, BOH, MIPH
Westmead Applied Research Centre
Faculty of Medicine and Health
The University of Sydney
REN Building, Darcy Road
Westmead, 2145
Australia
Phone: 61 8890 3125
Email: rebecca.chen@sydney.edu.au

Abstract

Background: Dental caries is the most common multifactorial oral disease; it affects 60% to 90% of the global population. Dental caries is highly preventable through prevention behaviors aimed at improving oral hygiene, adequate fluoride usage, and dietary intake. Mobile apps have the potential to support patients with dental caries; however, little is known about the availability, target audience, quality, and features of these apps.

Objective: This review aims to systematically examine dental caries prevention apps; to describe their content, availability, target audience, and features; and to assess their quality.

Methods: We systematically identified and evaluated apps in a process paralleling a systematic review. This included a search strategy using search terms; an eligibility assessment using inclusion and exclusion criteria focused on accessibility and dental caries self-management behaviors, including oral hygiene, dietary intake, and fluoride usage; data extraction on app characteristics, including app store metrics; prevention behavior categorization; feature identification and description; a quality appraisal of all apps using the validated Mobile App Rating Scale (MARS) assessment tool; and data comparison and analysis.

Results: Using our search strategy, we retrieved 562 apps from the Google Play Store and iTunes available in Australia. Of these, 7.1% (40/562) of the apps fit our eligibility criteria, of which 55% (22/40) targeted adults, 93% (37/40) were free to download, and 65% (26/40) were recently updated. Oral hygiene was the most common dental caries prevention behavior domain, addressed in 93% (37/40) of the apps, while dietary intake was addressed in 45% (18/40) of the apps and fluoride usage was addressed in 42% (17/40) of the apps. Overall, 50% (20/40) of the apps addressed only 1 behavior, and 38% (15/40) of the apps addressed all 3 behaviors. The mean MARS score was 2.9 (SD 0.7; range 1.8-4.4), with 45% (18/40) of the apps categorized as high quality, with a rating above 3.0 out of 5.0. We identified 21 distinctive features across all dental caries prevention behaviors; however, the top 5 most common features focused on oral hygiene. The highest-ranking app was the *Brush DJ* app, with an overall MARS score of 4.4 and with the highest number of features (n=13). We did not find any apps that adequately addressed dental caries prevention behaviors in very young children.

Conclusions: Apps addressing dental caries prevention commonly focus on oral hygiene and target young adults; however, many are not of high quality. These apps use a range of features to support consumer engagement, and some of these features may be helpful for specific patient populations. However, it remains unclear how effective these apps are in improving dental caries outcomes, and further evaluation is required before they are widely recommended.

(JMIR Mhealth Uhealth 2021;9(1):e19958) doi:10.2196/19958

KEYWORDS

dental caries; oral hygiene; self-management; mobile applications

Introduction

Background

Dental caries is a preventable, noncommunicable multifactorial disease that affects 60% to 90% of the population globally [1,2]. When left untreated, end-stage management of dental caries can result in pain, infection, and facial swelling, leading to emergency department presentations, especially for young children [3,4]. Although highly preventable, dental caries resulted in 70,200 preventable hospitalizations in Australia from 2016 to 2017 [5]. Dental caries prevention can be achieved at the individual level by addressing specific prevention behaviors, including adequate oral hygiene practices [3], age-appropriate topical fluoride usage [6], and diet modifications that reduce the amount of free sugar consumption [7]. At present, prevention behavior change interventions have included patient-focused dietary counseling and oral hygiene instruction, mostly delivered alongside operative clinical interventions in clinical settings [8,9]. However, these interventions are time intensive, workforce intensive, and expensive to deliver; also, without regular and repeated exposure, these interventions have shown inconsistent results on sustainably improved dental caries outcomes [9,10].

Growing mobile phone ownership globally and integration with the internet [11,12] have prompted the development of and research into mobile health (mHealth) interventions to address a broad range of behavior change practices for chronic disease management. These mHealth tools seek to modify a range of broad and specific behavioral factors related to diet [13], exercise [14], and medication adherence [15,16] to manage a range of chronic conditions, including diabetes [17,18], obesity [19,20], and cardiovascular diseases [15]. A variety of mHealth interventions have shown promising results in a variety of populations across the lifespan [21,22] and have particularly provided equitable support to remote, regional, and underserved populations [23-27]. Thus, it is necessary to both use and assess mHealth as a viable modality to support behavior change in the management of a range of noncommunicable diseases, including oral diseases, namely dental caries.

Apps for Dental Caries Prevention

Dental caries has many modifiable risk factors common to other noncommunicable diseases [4,28], driving a rationale for the adoption of innovative disease management approaches, including mHealth. Current research in mHealth for oral health has largely focused on addressing periodontal diseases through motivation for oral hygiene, with the delivery of simple text messages [29-31]. A recently published systematic review of the literature, focused on oral hygiene alone, highlighted the potential of mHealth interventions to improve oral health knowledge alongside modest clinical improvements in gum health in the adult and adolescent populations [29]. However, it remains unclear whether these results can be extended to address other dental caries risk factors, including a cariogenic diet and inadequate fluoride usage. Although important across the lifespan, preventive behaviors associated with appropriate fluoride usage and low sugar diets, including the timing of consumption are particularly influential in decreasing dental caries risk during the unique developmental stages of children aged younger than 6 years [3,8]. Uninformed parents could be at greater risk of their child experiencing a preventable hospitalization because of dental caries [5,10]. Although previous studies have focused on oral hygiene [29,32], it is important for this study to systematically scope the target audience and range of apps that addressed other dental caries prevention behaviors, including adequate fluoride usage and dietary modification.

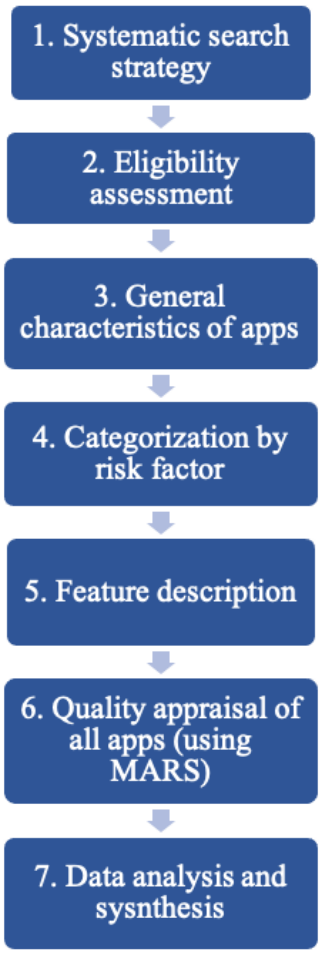
Furthermore, 2 recent reviews on apps used in oral health focused on the information analysis of apps that targeted the adult population in the United States [33] and the United Kingdom [32]. These reviews found poor information quality and identified the need to comprehensively analyze the features available in the apps alongside the use of a validated quality rating scale. The Mobile App Rating Scale (MARS) is a validated scale that has been used in a wide range of health care contexts to comprehensively assess the quality of health apps. Further analysis of features also provides information on usability and the potential for longer-term engagement with an app. Therefore, this study aims to systematically examine oral health apps that address a range of modifiable dental caries prevention factors and to systematically describe their content, availability, target audience, features, and quality.

Methods

Systematic Search Strategy

This review was conducted using a stepwise approach according to a previously published methodology that parallels a systematic review (Figure 1) [15]. We searched the main app stores: Google Play Store and iTunes. Of the 5.6 million apps available internationally, Google Play offers 2.47 million apps, and iTunes offers 1.8 million apps [34]. The search was conducted on the Google Play Store and iTunes with the app store country and region set to Australia between January 8 and 19, 2019, using the top 8 key search terms. These search terms were chosen based on their performance in retrieving the highest number of relevant apps for dental caries prevention from preliminary searches. The final list was developed and agreed upon by all authors and focused on the self-management behaviors that support dental caries prevention across all age groups, including young children. These search terms included dental caries, early childhood caries, tooth decay, dental caries prevention, early childhood caries prevention, tooth decay prevention, saliva, and fluoride.

Figure 1. Schematic steps of the systematic review and quality evaluation. MARS: Mobile App Rating Scale.



Eligibility Assessment

All apps retrieved from the search were screened by 2 independent reviewers (RC and GW) for eligibility using prespecified inclusion and exclusion criteria (Textbox 1) that aimed to identify relevant apps that were accessible to most of the general public relevant to dental caries prevention.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- The app focused on supporting the general public of any age to address prevention behaviors associated with dental caries prevention
- The app addressed dental caries prevention factors:
  - Oral hygiene including toothbrushing
  - Fluoride usage
  - Diet modification
- The app was accessible to the care context of Australia.
- The app was in the English language.

Exclusion criteria

- Apps targeted at clinicians or student clinicians
- Apps that were not related to health and were considered arcade games only
- Apps that addressed other health or oral health issues but were not specific to dental caries
- Apps that were priced or had in-app purchases or electronic devices such as electric toothbrushes priced above Aus $3.00 (US $2.29), which is the average price that consumers are willing to pay for an app [35] to ensure affordability
- Apps that were associated with a specific health clinic

General Characteristics of Apps

All apps that fit the inclusion criteria were downloaded on either the Android platform using a Samsung smartphone (Galaxy 9) or the iOS platform using an iPhone XR. The general characteristics of the apps were adapted from the *classification section* of the MARS tool (Multimedia Appendix 1) to describe the app name, app developer, date of the last update, platform (Google Play or iTunes), cost, star rating according to the app store, affiliations, target age groups, and focus of the app.

Categorization by Prevention Factors

Apps were categorized according to broad modifiable factors associated with dental caries prevention: oral hygiene, fluoride, and diet. For apps to be classified as addressing oral hygiene, at minimum, information about the importance of good oral hygiene for dental caries prevention had to be present in the app. Additional information and features could include a video demonstration of tooth brushing or interdental cleaning techniques, such as dental flossing, or *gamification* of toothbrushing, defined as the use of game elements in nongame contexts [36]. For fluoride, some information about the effects of fluoride from various modalities, including toothpaste or water consumption related to dental caries prevention, could be included. For diet, information on specific diet changes that could influence dental caries outcomes, for example, the use of a traffic light grading system to educate and encourage users to swap food choices to alternatives with less sugar [37], could be included.

Feature Description and Analysis

Features are elements of an app designed to increase interactivity and consumer engagement. For oral health apps, these features may include gamification and timers [33]. For the feature description, we identified and defined these through an iterative process combining terminology from previously published literature [15,38,39] with input from experts in the field and all authors. Further analysis was conducted to catalog the features of all apps stratified by the broad dental caries prevention factor each app addressed and to identify common features.

Quality Appraisal of Apps

All apps that fit the inclusion criteria were evaluated for quality using the MARS (Multimedia Appendix 1). This scale provides a standardized approach with 19 objective items and provides appraisal across 4 subscales: engagement, functionality, esthetics, and information quality [38,40]. The engagement subscale appraises whether the app was fun, interesting, customizable, and interactive (eg, push notifications, sends alerts) to the target audience. The functionality subscale assesses whether the app is correctly functioning and easy to learn, with easy navigations and logical flow. The esthetic subscale provides appraisal with regard to the general visual appeal and stylistic consistency of the app. Finally, the information subscale assesses the quality of the information, for example, whether the textual information and references are from credible sources. The overall MARS also sets a minimum quality threshold score of 3.0 out of 5.0, providing the ability to identify *high-quality* apps to patients and clinicians or further analysis [41]. It has been used in various contexts with excellent internal consistency and interrater reliability [15,42-46]. In total, 2 independent reviewers (RC and GW) were trained to use the MARS instruments through a web-based training program created by the MARS developers [38]. Each reviewer independently spent at least 30 min to thoroughly test each app on both devices. Data on the objective subscales of the MARS and additional features of the apps were extracted and entered into an Excel (Microsoft Corporation) spreadsheet. The items were rated on a 5-point scale (1: inadequate, 2: poor, 3: acceptable, 4: good, and 5: excellent). Any disagreements between the 2 reviewers were resolved by taking a consensus discussion. We calculated the means of the MARS and interrater reliability scores between reviewers using SPSS version 22.0 (IBM Corporation). High-quality apps were determined from the overall threshold score of 3 out of 5 in the overall mean score as defined by the developers of MARS, providing the ability to identify *high-quality* apps for further analysis [38,41].

Data Analysis and Synthesis

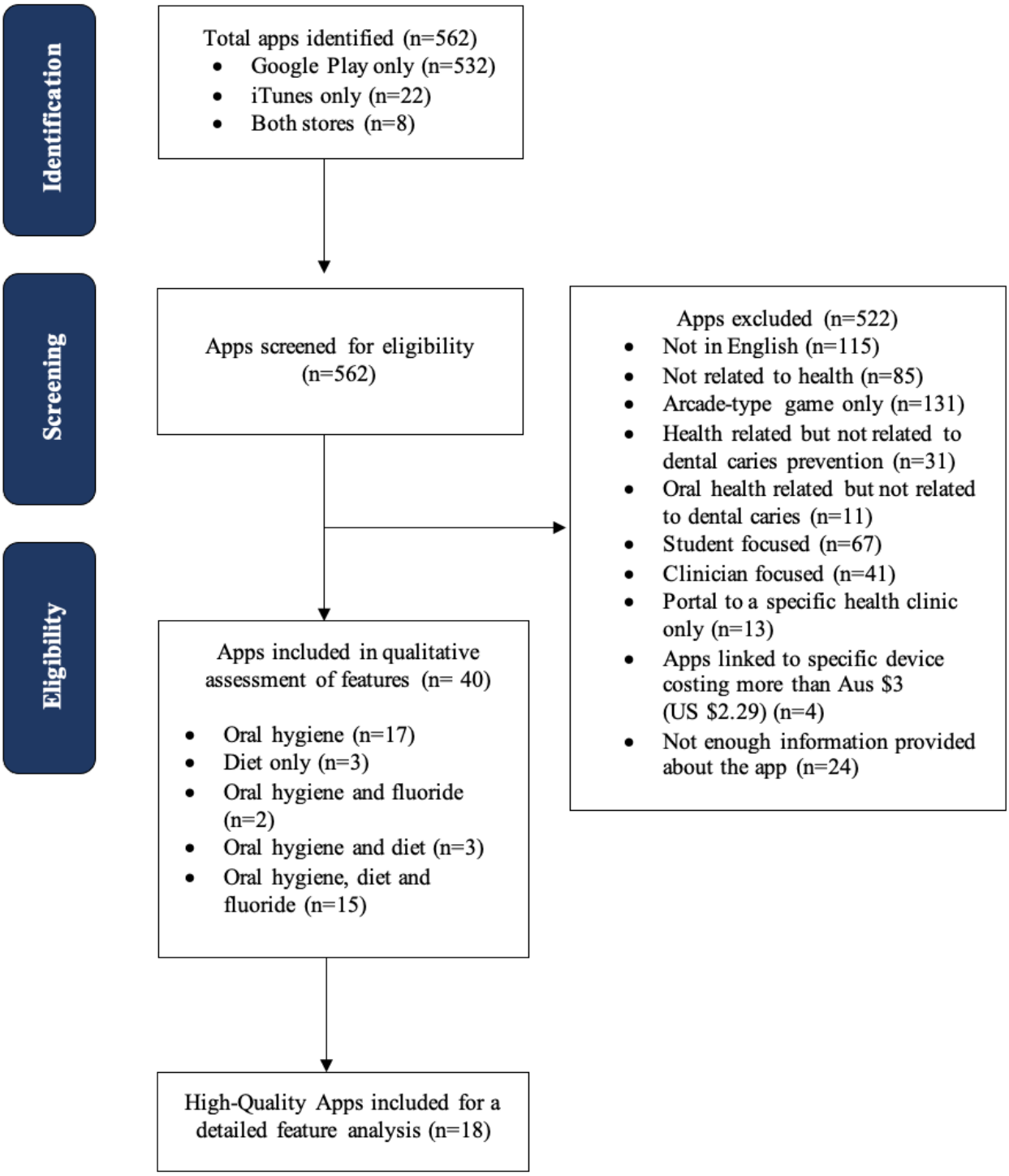
Further data analysis and synthesis were conducted based on the iteratively generated hypothesis from the initial MARS quality analysis. First, we wanted to compare the quality rating between apps that addressed a differing number and range of dental caries prevention factors. Second, the correlation between the MARS quality rating and the number of features across all apps was conducted.

Results

Systematic Search and Eligibility Assessment

Our systematic search retrieved a total of 562 apps, with 532 (94.7%) apps identified in the Google Play Store alone, 22 (5.3%) apps identified in iTunes alone, and 8 (1.4%) apps found in both stores. These apps were screened for eligibility, and 92.9% (522/562) were excluded based on the inclusion or exclusion criteria. The reasons for exclusion are presented in Textbox 1. A total of 40 (7.1%) unique apps were included for further analysis (Figure 2).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of apps identified through the systematic search.



General Characteristics of Apps

Most of the apps (37/40, 93%) were free to download, with only 3 of the 40 apps (8%) incurring a cost between Aus $0.99 (US $0.75) and Aus $2.99 (US $2.29) on iTunes alone. More than half (26/40, 65%) of the apps included were recent and current, as they were last updated either in 2018 or 2019. A total of 29 out of 40 apps (73%) were available on the Google Play Store only, and 3 out of 40 apps (8%) were available on the iTunes store only; 20% (8/40) of the apps were available in both stores (Multimedia Appendix 2).

Most (27/40, 68%) of the apps' affiliations were unknown, and 9 of 40 apps (23%) had clear commercial affiliations. Only 10% (4/40) of apps had affiliations with a university, and 50% (2/4) of these apps, namely, *Brush DJ* and *My Dental-Care - Your Guide to Oral Health*, had affiliations with the UK National Health Service, a government affiliation. More than half (22/40, 55%) of the apps were targeted toward adults or young adults, with 63% (14/22) of these apps also targeted adolescents. Apps

classified as targeting a general audience were 18% (7/40), whereas 28% (11/40) of apps were targeted at children aged older than 7 years. When analyzing the focus of the app, half (20/40, 50%) of the apps focused on information provision, such as health-seeking behaviors. The other half of the apps provided additional behavior change prompts, with 70% (14/20) of the apps providing specific goal setting functions within the app.

Categorization by Prevention Factor

Of the 40 included apps addressing a range of dental caries prevention factors, 37 (93%) addressed oral hygiene, 17 (43%) addressed fluoride, and 21 (53%) addressed diet (Figure 3). Furthermore, 50% (20/40) of apps addressed only one of these factors for oral health, 12% (5/40) of apps addressed 2 factors, and 38% (15/40) of apps addressed all 3 factors. Of the 20 apps that addressed only 1 factor, 17 (85%) addressed oral hygiene alone and 3 (15%) addressed diet alone. Of the 5 apps that addressed 2 factors, these combinations included oral hygiene and fluoride with 2 (40%) apps and oral hygiene and diet with 3 (60%) apps (Figure 4).

Of the 40 apps that fit our inclusion criteria, 37 (93%) addressed oral hygiene, 17 (43%) addressed fluoride, and 21 (53%) addressed diet.

Figure 3. Percentage of all apps that addressed each prevention factor.

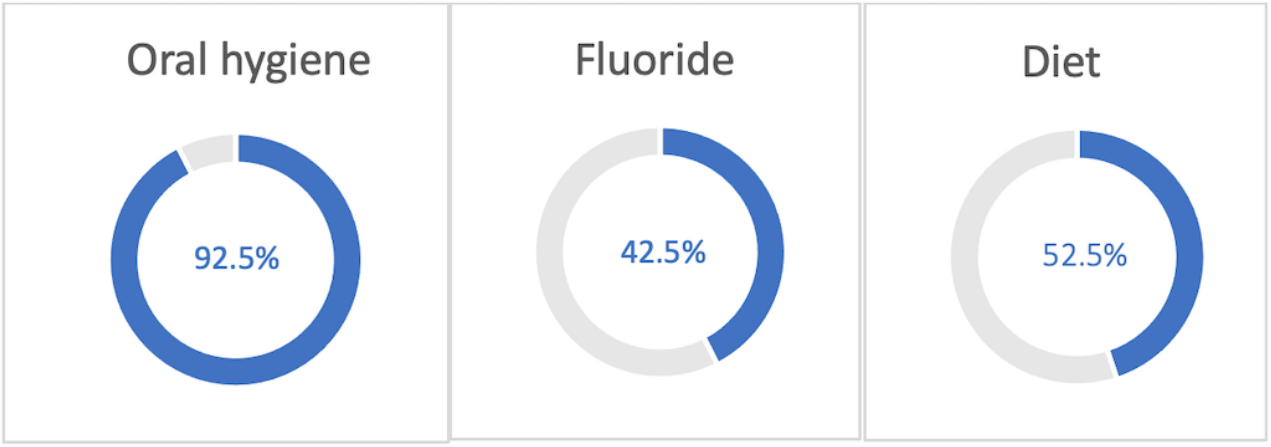
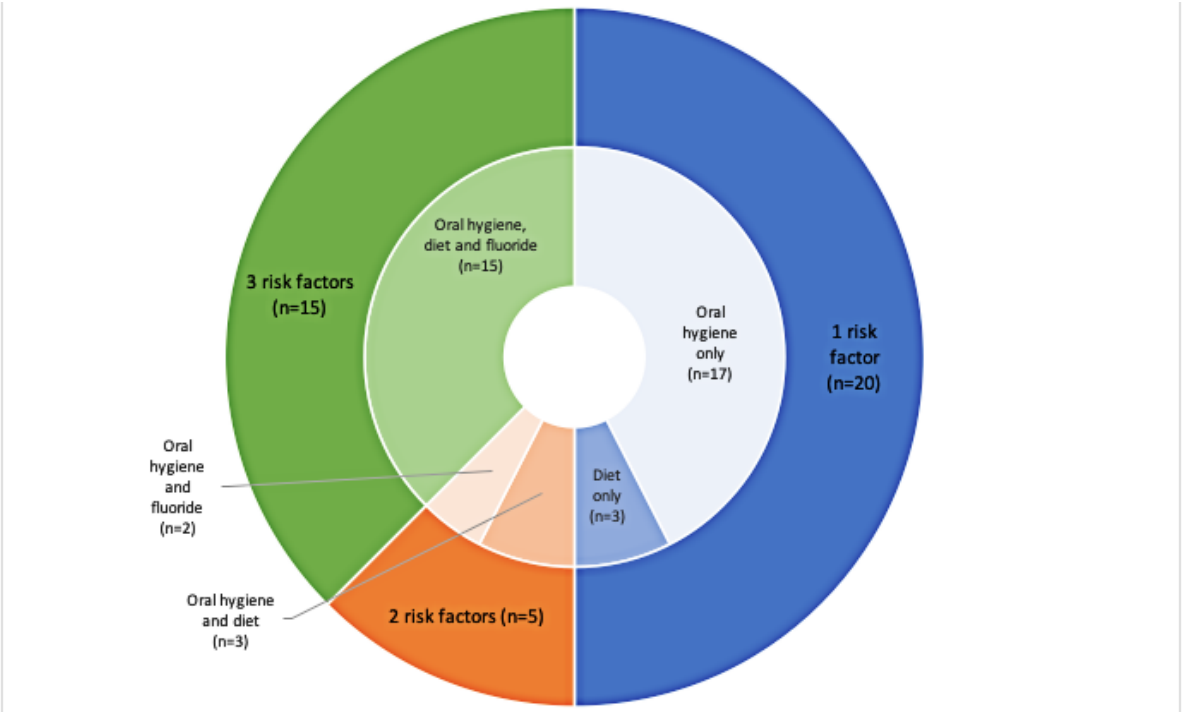


Figure 4. Categorization of apps according to the number and types of prevention factors addressed. (The outer rim indicates the number of prevention factors each app addressed. The inner rim shows the combination of the types of prevention factors that each app addressed).



Feature Description and Analysis

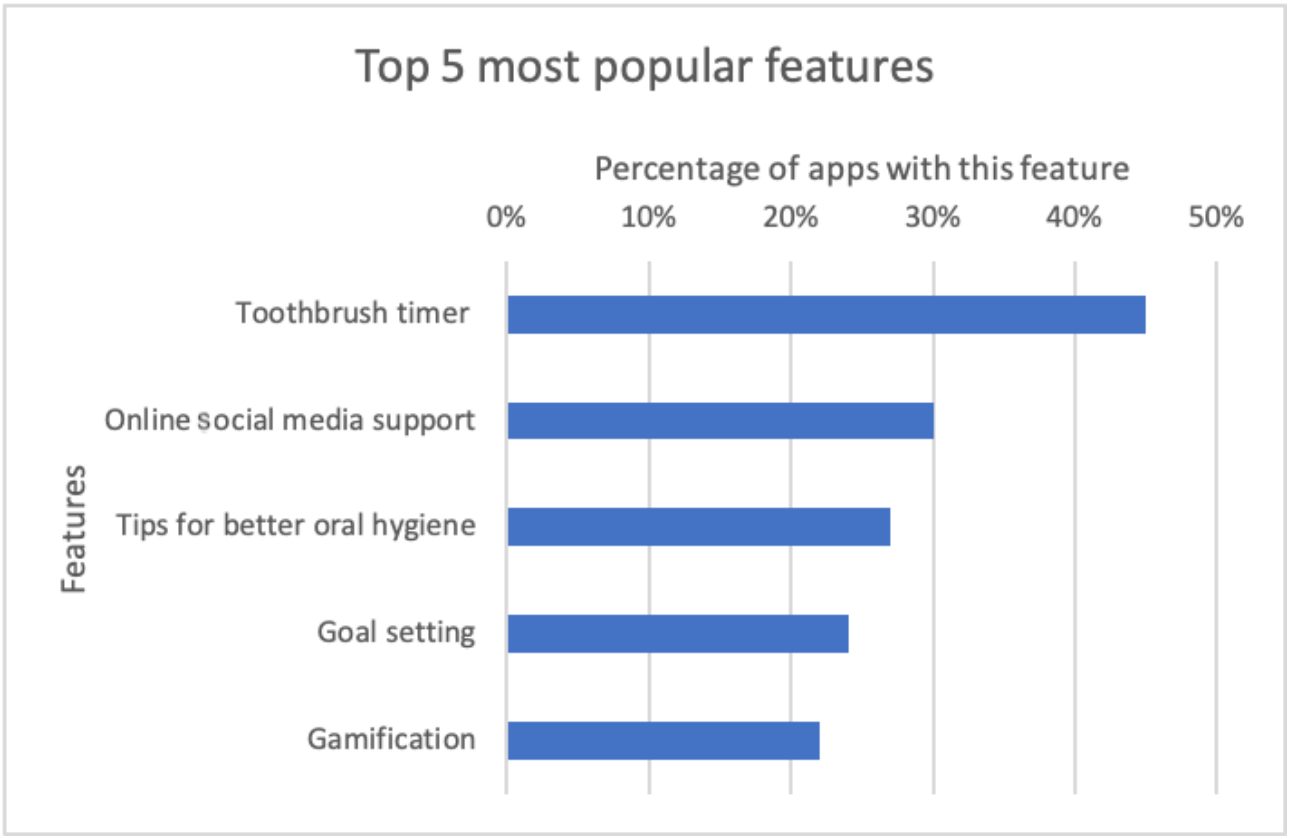
A total of 21 features were identified; these features were grouped into general and specific features related to each dental caries prevention factor, including oral hygiene, fluoride usage, and diet. The list of features and descriptions are outlined in Textbox 2; each feature was given a label to allow for a corresponding reference to Multimedia Appendix 3. Most of

the unique features identified were associated with oral hygiene (9/21, 43% features), compared with fluoride (2/21, 10% of features) and diet (5/21, 24% of features; Multimedia Appendix 3). Other than web-based social media support, considered to be a general feature, the top 5 most common features found in all apps addressed oral hygiene (Figure 5). The *Brush DJ* app had the highest number of features at 13 (Multimedia Appendix 3). Of 37 of 40 apps that addressed oral hygiene, 17 (46%) had the *tooth brush timer* as the most common feature, followed by *goal setting* (9/37, 24% of apps), *gamification* (8/37, 22% of apps), and 5 apps (14%) with a video demonstration of oral hygiene techniques. The *Brush DJ* app and the *Disney Magic Timer* app had the highest number of oral hygiene–related features, with each of these apps having 78% (7/9) of these features. Only 12% (2/17) of apps that addressed the fluoride prevention factor adequately provided additional information and used visual aids to support the information provided, for example, the appropriate amount of fluoride toothpaste that should be dispensed for children. These 2 apps, namely, *My Dental-Care - Your Guide to Oral Health* and *Brush DJ*, were also the only apps that were affiliated with both a university and government health service. Of the 21 apps that addressed diet, the *Food For Teeth - FoodDatabase and Diet Diary* app was the only app that included additional features such as an ability to record a diet diary, with a database of food items, including their pictures and serving size embedded in a traffic light system [47].

Textbox 2. Features of high-quality apps based on the Mobile App Rating Scale.

- **General features:**
  1. Updates conducted in 2018 or 2019: recent updates to ensure that glitches are resolved
  2. Data security: developer ensures data security, in accordance with mobile app regulatory statement, for example, Health Insurance Portability and Accountability Act compliance
  3. Data exporting and sharing to clinicians: ability to link information readily to the clinician or the electronic health record
  4. Tracking of dental appointments: ability to record dental appointments
  5. Online social media support: ability to connect to social media networks such as Facebook
- **Features related to oral hygiene:**
  1. Tooth brushing timer: timer to encourage patients to brush for a certain amount of time
  2. Tips for better oral hygiene: evidence-based information to improve oral hygiene
  3. Video demonstration: demonstration of brushing techniques via videos
  4. Goal setting: ability for user to input specific oral hygiene focused goal
  5. Tracking of oral health behavior: availability of statistics and charts on trends and adherence rates
  6. Push notifications (reminders): alert on the phone to remind patients to behavior change, for example, brushing their teeth
  7. Gamification: apps that use game elements to encourage users to brush their teeth. This can include virtual reality battles to encourage brushing length
  8. Incentivization: earn prizes for virtual characters or cryptocurrency—may be reduce the cost of your next appointment
  9. Sync to other apps on the phone: for example, the app may sync to a playlist to encourage brushing
- **Features related to fluoride:**
  1. Provision of fluoride information: information about fluoride usage is in accordance with country guidelines
  2. Pictures of the amount of toothpaste: visual aids to demonstrate the amount of fluoride toothpaste that is to be placed on the toothbrush
- **Features related to diet:**
  1. Provision of dietary information: information about the connection between dietary habits and dental caries information on alternatives
  2. Diet diary: ability for the user to input food intake and time of consumption
  3. Text search for food items: search bar to allow ease of entry of the food item consumed
  4. Pictures of food: picture of food to correspond to diet diary with a traffic light grading system to encourage users to consider low sugar alternatives
  5. Serving size of food: ability to record the amount and not just the type of food consumed

Figure 5. Top 5 most common features found in apps.



Quality Appraisal of Apps

Of the 40 apps, 18 (45%) were considered *high quality*, determined by reaching the minimum overall MARS threshold score of 3.0 out of 5.0. However, the MARS quality rating for each of the 40 apps found only 10 (25%) of these apps scored above 3.0 in all 4 subscales (Multimedia Appendix 4). The results also did not indicate that any single item in either of the 4 MARS subscales stood out. The interrater reliability between the reviewers, calculated from the overall and subscale scores of MARS for all apps, was excellent through the use of a two-way mixed intraclass correlation coefficient of 0.907 (95% CI 0.873-0.932). The *Brush DJ* app was the highest rated app with an overall MARS score of 4.4. This app also scored above the threshold for all subscales. The *Brush DJ* was also the only app that had scientific literature published, where a cross-sectional user acceptability questionnaire demonstrated 70.0% (133/189) of participants self-reported that the app motivated them to brush their teeth for longer [48]. However, the clinical effectiveness of this app is yet to be trialed using a study design that measures clinical health outcomes. Given the availability of only 1 app with supporting scientific literature published in this emerging field of inquiry, we followed the methodology adopted by other researchers in this situation [41] and excluded item 19 from our final calculation of the overall information subscale.

Data Analysis and Synthesis

Further data analysis and synthesis was developed based on an iteratively generated hypothesis from the initial MARS quality analysis. First, we wanted to compare MARS scores between apps that addressed a differing number and range of prevention factors. Second, because the app that had the highest MARS rating also had the highest number of features, we hypothesized a correlation between MARS scores and the number of features identified in each app. Table 1 shows that apps addressing the oral hygiene factor had the highest mean overall MARS scores (3.3) compared with apps addressing a combination of other factors: diet (2.2), oral hygiene and fluoride (1.9), and oral hygiene and diet (2.2). Apps addressing oral hygiene alone also had the highest subscale scores in engagement (3.3), functionality (3.8), and aesthetics (3.3). Although oral hygiene apps ranked equal to apps that addressed all 3 factors in the mean information subscale, with a MARS score (2.9), the percentage of apps that were considered high quality was more consistent for apps addressing all 3 factors (8/15, 54%; Table 1). Apps that addressed the oral hygiene factor alone had the highest percentage of apps that were considered high quality in the engagement (10/17, 59%) and esthetic (13/17, 76%) subscales. Apps that addressed all 3 factors were more likely to score above the threshold in the MARS information subscale (8/15, 54%) compared with apps that addressed 1 (7/20, 35%) factor. Of the apps, 20% (3/15) that addressed all 3 factors also ranked comparatively poorly on engagement scores compared with 59% of apps (10/17) that addressed only the oral hygiene factor.

Table 1. Mobile App Rating Scale quality rating summary in dental caries prevention factor categories.

MARS[a] subscale	1 factor				2 factors				3 factors	
	Oral hygiene only (n=17)		Diet only (n=3)		Oral hygiene and fluoride (n=2)		Oral hygiene and diet (n=3)		Oral hygiene, diet, and fluoride (n=15)	
	MARS score, mean (SD)	Quality apps[b], n (%)	MARS score, mean (SD)	Quality apps, n (%)	MARS score, mean (SD)	Quality apps, n (%)	MARS score, mean (SD)	Quality apps, n (%)	MARS score, mean (SD)	Quality apps, n (%)
Overall	3.3 (0.5)	10 (59)	2.7 (0.6)	0 (0)	1.9 (0.2)	0 (0)	2.6 (1.2)	1 (33)	2.9 (0.8)	6 (40)
Engagement	3.0 (0.6)	10 (59)	2.3 (0.5)	0 (0)	1.3 (0.1)	0 (0)	2.8 (1.3)	1 (33)	2.5 (0.9)	3 (20)
Functionality	3.9 (0.7)	15 (88)	3.8 (0.0)	3 (100)	2.6 (0.2)	0 (0)	2.6 (1.3)	1 (33)	3.6 (0.7)	12 (80)
Aesthetics	3.3 (0.7)	13 (76)	2.4 (0.2)	0 (0)	1.8 (0.2)	0 (0)	2.4 (1.4)	1 (33)	2.6 (0.7)	5 (33)
Information[c]	2.9 (0.7)	7 (41)	2.2 (1.3)	1 (33)	1.9 (0.4)	0 (0)	2.2 (1.3)	1 (33)	2.9 (0.8)	8 (54)

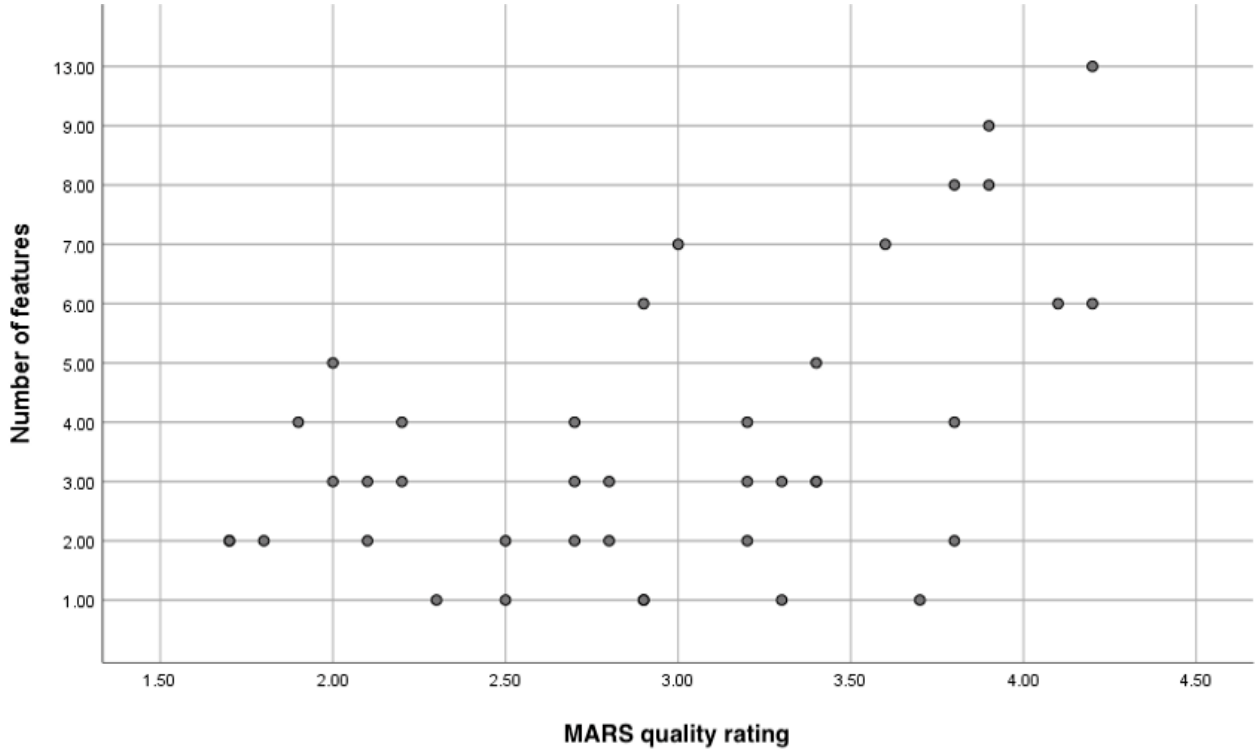
[a]MARS: Mobile App Rating Scale.

[b]Percentage of apps determined to be of high quality, determined by an overall score that reached above the minimum threshold score of above 3.0 out of 5.0.

[c]Item 19 of the information subscale was excluded from the final calculation, as only 1 app supported the scientific literature published in this emerging field of inquiry, a similar methodology adopted by other researchers in this context [41].

When analyzing the number of features, high-quality apps, categorized as those with an overall MARS score above 3, had almost double the mean number of features compared with low-quality apps, which was consistent across all categories (Multimedia Appendix 5). However, when comparing the individual apps with the MARS quality rating, the correlation between quality and the number of features showed variance and showed only a moderate general trend that high-quality apps had more features compared with low-quality apps (Figure 6).

Figure 6. Number of features compared with the MARS quality rating scale for each app. MARS: Mobile App Rating Scale.



Discussion

Principal Findings

Our study identified and assessed the characteristics, features, and quality of 40 apps targeted at the general public addressing a range of dental caries prevention factors. Less than half (18/40, 45%) of the apps were considered good quality (based on the overall MARS); however, only 25% (10/40) of apps were considered high quality across all 4 subscales of MARS. Most apps (37/40, 93%) focused on oral hygiene, and these apps were also more likely to contain a higher number of features and

target the young adults or adolescent population. Only 13% (5/40) of apps addressed all 3 factors and were considered high quality, indicating only a small number of apps that could be further tested for clinical effectiveness specific to the adolescent or young adult populations. This could be recommended for the highest rated app, *Brush DJ*, which scored highly across all 4 subscales, demonstrating that it is possible to create an app that is able to provide good information and be aesthetically appealing. We did not find any high-quality apps targeted to parents of very young children that would address the specific dental caries prevention behaviors associated with caring for children aged between 0 and 6 years. Therefore, our study indicates an opportunity for future high-quality app development that addresses a range of dental caries prevention behaviors alongside a consideration for esthetics and engaging features to support this target parent population.

Comparison With Prior Work

This study goes further than 2 previous reviews of mHealth apps focused on oral hygiene only [32] and oral health promotion in adult populations only [33] by scoping a broad range of dental caries prevention factors. Our study also responds to the need for feature analysis and quality appraisal outlined by the 2 most recent reviews on apps used for oral health [32,33]. First, our study provides the feature analysis of apps addressing a range of dental caries prevention behaviors, including diet modification and adequate fluoride usage. Second, the study provides an in-depth quality appraisal using the MARS tool. Consistent with the information quality concerns raised by these previous reviews, our analysis found 24 of 40 (60%) apps to be of low quality according to the MARS tool, yet still available to the general public. To address the issue of app quality [32], the National Health Service in the United Kingdom has created a digital app library that could be a trusted source of information for both clinicians and patients [49].

Our quality appraisal using the MARS assessment identified *Brush DJ* to be the highest-ranking app in quality and features, a similar finding from previously published reviews focused on oral hygiene in the United Kingdom [32]. At present, the *Brush DJ* app is the only app endorsed by the National Health Service's digital app library [50]. Further well-designed clinical trials to determine the clinical efficacy of this app within the adolescent or young adult target populations should be undertaken. These clinical trials should have a clear clinical question with a good study design and be a randomized controlled trial where possible, with defined measurable health outcomes and a complementary economic evaluation [51]. Clinical effectiveness shown through improvements in measurable health outcomes will facilitate a more widespread adoption of this app and other effective apps in clinical practice.

Implications

Dental caries is a multifactorial disease with varied risk factors that may impact individual patients differently during their life course [2,3]. Our study identified a lack of apps targeted to parents of children that adequately addressed prevention behaviors associated with fluoride usage and low-cariogenic diets for children aged younger than 6 years [3,52]. Given the paucity surrounding the clinical efficacy of mHealth apps in the field of dentistry [29], before a well-designed clinical trial can be conducted in this target population [48,53], a high-quality app needs to be designed. This app must target specific behaviors relevant to a broad range of dental caries prevention behaviors and contain evidence-based information, with appropriate features and esthetics to ensure the engagement of parents. To date, no high-quality app has achieved this. Thus, more user-focused, iterative co-designed research [54] relevant to the target population is needed to determine engaging features that will address all relevant dental caries prevention behaviors. Our assessment of features will provide a blueprint to assist future researchers engaged in qualitative user engagement research with parents, patients, clinicians, professional dental associations, health services, and research organizations [51,55] to develop a high-quality app that could then be trialed for clinical effectiveness.

Limitations

Our review was conducted in Australia in 2019, and the included apps were limited to those available in Australian app stores at the specific time of the systematic search. Most apps included in this review have been developed in predominately English-speaking countries outside Australia, the United Kingdom, Canada, Asia, and the United States. We recognize that there may be apps developed in other languages or only available in country-specific app stores that were not included in our review. However, we did find a similar number of apps for the final analysis when using search terms similar to previous studies undertaken in other countries, including the United Kingdom [32] and the United States [33]. Second, in our study, MARS was used by researchers with clinical backgrounds in the field of oral health and primarily reflected this perspective. Our scoping study did not involve patients as participants and highlights the importance of conducting further complementary research that involves end users and giving voice to the patient's perspective during the development of future apps and mHealth interventions [56].

Conclusions

The increasing use of mHealth apps driven by increasing public use of mobile devices presents a call for dental researchers, health system managers, policy makers, and health professionals to engage with and provide more rigorous scientific recommendations around oral health apps. Our study provides a systematic and detailed analysis of the current availability, target audience, quality, and features of apps targeted toward dental caries. Quality was variable across the apps and mainly targeted the adolescent and adult populations. The most common features found in high-quality apps, such as gamification and goal setting, still focus on oral hygiene factors. It is unclear if these features can be used to address other dental caries prevention factors such as fluoride and diet modification. There was also an identified gap in apps available to support the target audience of parents of young children. There is a real need to co-design and create apps that address a broad range of modifiable risk factors associated with dental caries targeted at parents of children aged younger than 6 years. To ensure the highest quality in apps, the co-design process should include the clinician, researcher, and patient perspectives on

evidence-based information and engaging features. Further studies are needed to determine the clinical efficacy of these apps before they can be widely recommended.

Authors' Contributions

RC, the lead author, provided substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work and drafting the work. KS provided substantial contributions to the conception or design of the work or interpretation of data for the work. GW provided substantial contributions to the acquisition and analysis of the work and drafted this work. WS provided substantial contributions to the conception or design of the work. HS provided substantial contributions to the conception or design of the work. CC, an equal last author, provided substantial contributions to the conception or design of the work and the analysis and interpretation of data for the work. MI, an equal last author, provided substantial contributions to the conception or design of the work and the acquisition, analysis, or interpretation of data for the work and had drafted the work.

All authors revised this work critically for important intellectual content and had final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile App Rating Scale tool developed by Queensland University of Technology (QUT).
[PDF File (Adobe PDF File), 262 KB - mhealth_v9i1e19958_app1.pdf]

Multimedia Appendix 2
General characteristics of all apps.
[DOCX File , 21 KB - mhealth_v9i1e19958_app2.docx]

Multimedia Appendix 3
Comparative analysis of quality features present in all apps.
[DOCX File , 31 KB - mhealth_v9i1e19958_app3.docx]

Multimedia Appendix 4
Quality evaluation of all apps according to the objective subscales of the Mobile App Rating Scale (MARS) quality appraisal tool.
[DOCX File , 20 KB - mhealth_v9i1e19958_app4.docx]

Multimedia Appendix 5
Mean number of features present in high-quality versus low-quality apps within dental caries prevention factors.
[DOCX File , 15 KB - mhealth_v9i1e19958_app5.docx]

References

1. Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJL, Marcenes W. Global burden of untreated caries: a systematic review and metaregression. J Dent Res 2015 May;94(5):650-658. [doi: 10.1177/0022034515573272] [Medline: 25740856]
2. Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, et al. Dental caries. Nat Rev Dis Primers 2017 May 25;3:17030. [doi: 10.1038/nrdp.2017.30] [Medline: 28540937]
3. Phantumvanit P, Makino Y, Ogawa H, Rugg-Gunn A, Moynihan P, Petersen PE, et al. WHO global consultation on public health intervention against early childhood caries. Community Dent Oral Epidemiol 2018 Jun;46(3):280-287. [doi: 10.1111/cdoe.12362] [Medline: 29380407]
4. Peres MA, Macpherson LM, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. Lancet 2019 Jul 20;394(10194):249-260. [doi: 10.1016/S0140-6736(19)31146-8] [Medline: 31327369]
5. Al Rawahi SH, Asimakopoulou K, Newton JT. Theory based interventions for caries related sugar intake in adults: systematic review. BMC Psychol 2017 Jul 25;5(1):25 [FREE Full text] [doi: 10.1186/s40359-017-0194-z] [Medline: 28743291]
6. Walsh T, Worthington HV, Glenny AM, Marinho VC, Jeroncic A. Fluoride toothpastes of different concentrations for preventing dental caries. Cochrane Database Syst Rev 2019 Mar 04;3:CD007868 [FREE Full text] [doi: 10.1002/14651858.CD007868.pub3] [Medline: 30829399]

7. Moynihan P, Makino Y, Petersen PE, Ogawa H. Implications of WHO guideline on sugars for dental health professionals. Community Dent Oral Epidemiol 2018 Feb;46(1):1-7. [doi: 10.1111/cdoe.12353] [Medline: 29168887]
8. Arrow P, Raheb J, Miller M. Brief oral health promotion intervention among parents of young children to reduce early childhood dental decay. BMC Public Health 2013 Mar 20;13:245 [FREE Full text] [doi: 10.1186/1471-2458-13-245] [Medline: 23509932]
9. Harris R, Gamboa A, Dailey Y, Ashcroft A. One-to-one dietary interventions undertaken in a dental setting to change dietary behaviour. Cochrane Database Syst Rev 2012 Mar 14(3):CD006540 [FREE Full text] [doi: 10.1002/14651858.CD006540.pub2] [Medline: 22419315]
10. Henshaw M, Borrelli B, Gregorich S, Heaton B, Tooley E, Santo W, et al. Randomized trial of motivational interviewing to prevent early childhood caries in public housing. JDR Clin Trans Res 2018 Oct;3(4):353-365 [FREE Full text] [doi: 10.1177/2380084418794377] [Medline: 30238060]
11. Share of mobile phone users that use a smartphone in Australia from 2014 to 2019. Statista. 2019 Aug 27. URL: https://www.statista.com/statistics/257041/smartphone-user-penetration-in-australia/ [accessed 2019-08-27]
12. Silver L, Huang C. Social activities, information seeking on subjects like health and education top the list of mobile activities. Pew Research Center. 2019 Aug 22. URL: https://www.pewresearch.org/internet/2019/08/22/social-activities-information-seeking-on-subjects-like-health-and-education-top-the-list-of-mobile-activities/ [accessed 2019-08-27]
13. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
14. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A short message service intervention to support adherence to home-based strengthening exercise for people with knee osteoarthritis: intervention design applying the behavior change wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
15. 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]
16. Badawy SM, Shah R, Beg U, Heneghan MB. Habit strength, medication adherence, and habit-based mobile health interventions across chronic medical conditions: systematic review. J Med Internet Res 2020 Apr 28;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]
17. Pfammatter A, Spring B, Saligram N, Davé R, Gowda A, Blais L, et al. mHealth intervention to improve diabetes risk behaviors in India: a prospective, parallel group cohort study. J Med Internet Res 2016 Aug 05;18(8):e207 [FREE Full text] [doi: 10.2196/jmir.5712] [Medline: 27496271]
18. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
19. Kay MC, Burroughs J, Askew S, Bennett GG, Armstrong S, Steinberg DM. Digital weight loss intervention for parents of children being treated for obesity: a prospective cohort feasibility trial. J Med Internet Res 2018 Dec 20;20(12):e11093 [FREE Full text] [doi: 10.2196/11093] [Medline: 30573449]
20. Haas K, Hayoz S, Maurer-Wiesner S. Effectiveness and feasibility of a remote lifestyle intervention by dietitians for overweight and obese adults: pilot study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12289 [FREE Full text] [doi: 10.2196/12289] [Medline: 30973338]
21. Albanese-O'Neill A, Schatz DA, Thomas N, Bernhardt JM, Cook CL, Haller MJ, et al. Designing online and mobile diabetes education for fathers of children with type 1 diabetes: mixed methods study. JMIR Diabetes 2019 Aug 06;4(3):e13724 [FREE Full text] [doi: 10.2196/13724] [Medline: 31389338]
22. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
23. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the united states: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
24. Somers C, Grieve E, Lennon M, Bouamrane M, Mair FS, McIntosh E. Valuing mobile health: an open-ended contingent valuation survey of a national digital health program. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e3 [FREE Full text] [doi: 10.2196/mhealth.9990] [Medline: 30664488]
25. Saleh S, Farah A, Dimassi H, El Arnaout N, Constantin J, Osman M, et al. Using mobile health to enhance outcomes of noncommunicable diseases care in rural settings and refugee camps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 13;6(7):e137 [FREE Full text] [doi: 10.2196/mhealth.8146] [Medline: 30006326]
26. Hughson J, Daly J, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]

27. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of 'Bling': is there a future for text messaging in health care? J Med Internet Res 2019 Dec 20;21(12):e16630 [FREE Full text] [doi: 10.2196/16630] [Medline: 31859678]
28. Watt RG, Daly B, Allison P, Macpherson LM, Venturelli R, Listl S, et al. Ending the neglect of global oral health: time for radical action. Lancet 2019 Jul 20;394(10194):261-272. [doi: 10.1016/S0140-6736(19)31133-X] [Medline: 31327370]
29. Toniazzo MP, Nodari D, Muniz FW, Weidlich P. Effect of mHealth in improving oral hygiene: a systematic review with meta-analysis. J Clin Periodontol 2019 Mar;46(3):297-309. [doi: 10.1111/jcpe.13083] [Medline: 30761580]
30. Borrelli B, Henshaw M, Endrighi R, Adams WG, Heeren T, Rosen RK, et al. An interactive parent-targeted text messaging intervention to improve oral health in children attending urban pediatric clinics: feasibility randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e14247 [FREE Full text] [doi: 10.2196/14247] [Medline: 31710306]
31. Marshman Z, Ainsworth H, Chestnutt IG, Day P, Dey D, El Yousfi S, et al. Brushing RemInder 4 Good oral HealTh (BRIGHT) trial: does an SMS behaviour change programme with a classroom-based session improve the oral health of young people living in deprived areas? A study protocol of a randomised controlled trial. Trials 2019 Jul 23;20(1):452 [FREE Full text] [doi: 10.1186/s13063-019-3538-6] [Medline: 31337437]
32. Parker K, Bharmal RV, Sharif MO. The availability and characteristics of patient-focused oral hygiene apps. Br Dent J 2019 Apr;226(8):600-604. [doi: 10.1038/s41415-019-0197-7] [Medline: 31028329]
33. Tiffany B, Blasi P, Catz SL, McClure JB. Mobile apps for oral health promotion: content review and heuristic usability analysis. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e11432 [FREE Full text] [doi: 10.2196/11432] [Medline: 30181114]
34. Clements J. Number of apps available in leading app stores as of 3rd quarter 200. Statista. 2020. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2020-03-01]
35. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
36. Floryan M, Chow PI, Schueller SM, Ritterband LM. The model of gamification principles for digital health interventions: evaluation of validity and potential utility. J Med Internet Res 2020 Jun 10;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
37. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. JMIR Mhealth Uhealth 2014 Aug 21;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]
38. 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]
39. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
40. 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]
41. 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]
42. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
43. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: systematic search in app stores and evaluation. JMIR Mhealth Uhealth 2019 Sep 04;7(9):e12604 [FREE Full text] [doi: 10.2196/12604] [Medline: 31486408]
44. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the mobile application rating scale. JMIR Mhealth Uhealth 2019 May 01;7(5):e13019 [FREE Full text] [doi: 10.2196/13019] [Medline: 31066712]
45. Woods LS, Duff J, Roehrer E, Walker K, Cummings E. Patients' experiences of using a consumer mHealth app for self-management of heart failure: mixed-methods study. JMIR Hum Factors 2019 May 02;6(2):e13009 [FREE Full text] [doi: 10.2196/13009] [Medline: 31045504]
46. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
47. Volkova E, Li N, Dunford E, Eyles H, Crino M, Michie J, et al. 'Smart' RCTs: development of a smartphone app for fully automated nutrition-labeling intervention trials. JMIR Mhealth Uhealth 2016 Mar 17;4(1):e23 [FREE Full text] [doi: 10.2196/mhealth.5219] [Medline: 26988128]
48. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]

49. Guidance for health app developers, commissioners and assessors. United Kingdom National Health Service. 2019 Aug 13. URL: https://digital.nhs.uk/services/nhs-apps-library/guidance-for-health-app-developers-commissioners-and-assessors [accessed 2019-10-14]
50. British Dental Journal. Brush DJ app dentist wins funding boost. Br Dent J 2019 Mar 22;226(6):387-387 [FREE Full text] [doi: 10.1038/s41415-019-0186-x]
51. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
52. Tinanoff N, Baez RJ, Diaz Guillory C, Donly KJ, Feldens CA, McGrath C, et al. Early childhood caries epidemiology, aetiology, risk assessment, societal burden, management, education, and policy: Global perspective. Int J Paediatr Dent 2019 May;29(3):238-248. [doi: 10.1111/ipd.12484] [Medline: 31099128]
53. Scheerman JF, van Empelen P, van Loveren C, van Meijel B. A mobile app (WhiteTeeth) to promote good oral health behavior among Dutch adolescents with fixed orthodontic appliances: intervention mapping approach. JMIR Mhealth Uhealth 2018 Aug 17;6(8):e163 [FREE Full text] [doi: 10.2196/mhealth.9626] [Medline: 30120085]
54. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
55. Health and wellbeing apps. Australian Government Health Department. 2019. URL: https://www.healthdirect.gov.au/health-and-wellbeing-apps [accessed 2019-10-17]
56. Greenhalgh T, Hinton L, Finlay T, Macfarlane A, Fahy N, Clyde B, et al. Frameworks for supporting patient and public involvement in research: systematic review and co-design pilot. Health Expect 2019 Aug;22(4):785-801 [FREE Full text] [doi: 10.1111/hex.12888] [Medline: 31012259]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by C Dias; submitted 07.05.20; peer-reviewed by O Zelenko, S Ramalingam, H Ranjani, P Blasi; comments to author 16.06.20; revised version received 10.08.20; accepted 11.08.20; published 13.01.21.

Please cite as:
Chen R, Santo K, Wong G, Sohn W, Spallek H, Chow C, Irving M
Mobile Apps for Dental Caries Prevention: Systematic Search and Quality Evaluation
JMIR Mhealth Uhealth 2021;9(1):e19958
URL: http://mhealth.jmir.org/2021/1/e19958/
doi:10.2196/19958
PMID:33439141

Original Paper

Willingness to Adopt mHealth Among Chinese Parents During the COVID-19 Outbreak: Cross-sectional Questionnaire Study

Siyu Yang[1*], MSN, RN; Yijing Chen[2*], MSN, RN; Leshan Zhou[3], PhD, DR; Yuting Huang[3], RN, MSN; Jiahui Dai[4], RN, MPH

[1]Nursing Department, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Wuhan Mental Health Center–affiliated Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Xiangya Nursing School, Central South University, Changsha, China

[4]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Leshan Zhou, PhD, DR
Xiangya Nursing School
Central South University
No172 Tongzipo Road, Yuelu District
Changsha
China
Phone: 86 1897516606
Email: leshanzhou@csu.edu.cn

Abstract

Background: Parental involvement in mobile health (mHealth) to consult with medical professionals appears to be prevalent in China with the rapid development of the internet. More parents with busy jobs have chosen to use mHealth. During the ongoing COVID-19 outbreak, mHealth can assist with health promotion, directions for medication use, and disease diagnosis via online chat and video consultation without contacting others. To our knowledge, no studies have been performed to explore the role of mHealth in parents' attitudes toward child health care at home during the COVID-19 outbreak.

Objective: This study aims to identify the associated factors of willingness to adopt mHealth among Chinese parents during the COVID-19 outbreak and to explore the correlation between the frequency of adopting mHealth and parents' attitudes toward child health care at home.

Methods: Chinese parents were asked to complete an online survey from January 25 to February 15, 2020. The questionnaire comprised of two parts with a total of 16 items, including parents' demographic variables and attitudes toward child health care at home. By multivariate logistic regression, we explored factors associated with parents' willingness to adopt mHealth during the COVID-19 outbreak. Pearson chi-square tests were used to reveal the correlation between the frequency of adopting mHealth and parents' attitudes toward child health care at home.

Results: A total of 254 parents enrolled, and 202 (79.5%) parents were willing to adopt mHealth during the COVID-19 outbreak. Parents' age (26-35 years: adjusted odds ratio [AOR] 8.114, 95% CI 1.471-44.764), parents' interest in the COVID-19 pandemic (moderate: AOR 8.753, 95% CI 2.009-38.127; high: AOR 22.194, 95% CI 5.509-89.411), the source that recommended mHealth (medical health providers: AOR 4.257, 95% CI 1.439-12.596), the presence of chronic disease in their children (yes: AOR 20.844, 95% CI 4.600-94.443), parents' duration of daily internet use (4-6 hours: AOR 6.487, 95% CI 1.870-22.495; >6 hours: AOR 8.766, 95% CI 1.883-40.804), and adoption of mHealth before the COVID-19 outbreak (yes: AOR 3.413, 95% CI 1.234-9.444) were significantly correlated with the parents' willingness to adopt mHealth during the COVID-19 outbreak. The frequency of mHealth use among parents was correlated with their behaviors in regard to handwashing (χ^2_6=18.967, P=.004), mask wearing (χ^2_6=45.364, P<.001), frequency of leaving the home (χ^2_6=16.767, P=.01), room disinfection and ventilation (χ^2_6=19.515, P=.003), temperature checking (χ^2_6=17.47, P=.007), and mental health care of children (χ^2_6=63.810, P<.001) during the COVID-19 pandemic.

Conclusions: We found various objective factors that were associated with parents' willingness to adopt mHealth during the COVID-19 outbreak. Overall, parents' willingness to adopt mHealth was high. The frequency of mHealth use among parents

was correlated with their attitudes toward child health care at home. The option of mHealth to patients at home during the COVID-19 outbreak would be beneficial for education and improvement in self-management of child health care at home.

(JMIR Mhealth Uhealth 2021;9(1):e23155) doi:10.2196/23155

KEYWORDS

mHealth; parents; child health at home; COVID-19

Introduction

Background

COVID-19 has caused an ongoing pandemic and is an important public health concern. The major transmission modes of COVID-19 are airborne droplets from coughing or sneezing and direct contact with contaminated surfaces such as doorknobs, dishes, and handrails [1]. Particular circumstances including tracheal incubation or opening suction in a hospital, staying in confined spaces with infected people, and fomites attaching to ventilation systems can result in mass infection [1].

Infants and children are a typically vulnerable population due to immaturity of the respiratory tract and hypoimmunity [2]. According to an analysis of 2143 pediatric cases in China, the median age of children who are infected is 7 years [3]. The majority of severe pediatric cases arose from exposure to infected family members, and few were infected in a hospital or as a result of travel [4]. To effectively prevent and control the COVID-19 outbreak among Chinese children, the National Health Commission issued guidelines termed "Epidemiological characteristics and prevention and control measures of Corona Virus Disease 2019 in children," which clearly state that children are required to be isolated at home under the parents' supervision [5]. In addition, experts also suggest that parents do not take children to the hospital to avoid cross infection [6]. To prevent large scale gatherings, all regional governments of China shut down schools and colleges, and decreased the run time of public transportation. Residents dwelling in high-risk areas are forbidden to go out except to acquire daily necessities and to visit the hospital [6].

In China, more than 1.3 billion people access the internet via their mobile phones, which has become an indispensable part of daily life [7]. The application of mobile health (mHealth) is widely recommended for Chinese parents to replace visiting a hospital during the COVID-19 outbreak. mHealth is defined as the use of wireless electronic devices to transmit various contents and medical services among patients and caregivers. Besides routine use, Chinese people could use mHealth during the COVID-19 pandemic through mobile phone apps [8,9], the hospital website, and doctors' official social media accounts on WeChat or Alipay to get primary diagnoses between common cold or flu and pneumonia, achieve self-monitoring, access online lectures about COVID-19 prevention, and purchase essential medicine from online diagnosis. Specific individuals could also acquire urgent care in emergency situations by using mHealth to contact specialists or hospitals designed for delivery, patients with chronic disease, or terminal cancer. mHealth is viewed as an easily accessible, cost-efficient approach to enhancing adherence to medication, expanding access to medical care, and increasing the number of medical consultations [10]. Overall, mHealth has proven to be a great success for the management of chronic disease [11], remote monitoring of weight control [12], improvement of child vaccine coverage, and neonatal care among young mothers [13].

Objectives

Though robust evidence highlights the potential benefits of mHealth, individual willingness to adopt mHealth is a decisive factor. Therefore, the purpose of this paper is to explore factors associated with willingness to adopt mHealth among Chinese parents during the COVID-19 outbreak for better promotion of mHealth in China and to investigate the correlation between frequency of adopting mHealth during the COVID-19 outbreak in 1 month and parents' attitudes toward child health at home.

Methods

Definition of Variables

The willingness to adopt mHealth among Chinese parents was measured with a yes or no question.

mHealth in China is defined as the dissemination of medical information, consultation about disease diagnosis and treatment, postoperative care management, mental health care, and making medical appointments via mobile phone apps or social media [14].

Participant Recruitment

This cross-sectional study was conducted between January 25, 2020, and February 15, 2020. Data were collected using structured questionnaires based on a literature review. A total of 12 participants were recruited to test the original questionnaire and provide feedback, ensuring the questionnaire was understandable. Regarding the COVID-19 pandemic, only an online version of this questionnaire was used for distributing the survey. The online questionnaire link was forwarded to social media groups, such as breastfeeding groups and parental involvement in kindergarten or primary school online groups, and posted on pediatric researchers' social media webpages, where health education about children is given, to maximize recruitment of respondents. If parents were interested in our study, they could visit our questionnaire through a link, and a description of this study was shown on the first page of the online questionnaire. Parents could only access the questionnaire by clicking "agree" after reading the consent information on the first page, and the submission was accepted when all items were completed.

Parents who met any of the following criteria were excluded: completed the questionnaire in less than 120 seconds, their child was older than 14 years, the parent has not lived with their child during the COVID-19 pandemic, one or both parents have a

confirmed case of COVID-19, parent has no internet access, more than one questionnaire was submitted from the same Internet Protocol address, and the questionnaire had missing items. Two researchers, who were master's degree students, selected the valid questionnaires in accordance with the exclusion criteria. This study was approved by the Ethics Committees of Xiangya Nursing School, Central South University. In total, 18 questionnaires were considered substandard and excluded, and 254 Chinese parents participated in the online survey with an effective rate of 93.38% (254/272).

Measurement of Variables

Following a literature review by our research team, consisting of one professor and five graduate students, a structured questionnaire was designed for this study. The 23-item questionnaire was comprised of two parts: (1) parents' demographic variables including gender, age, education level, family annual income, occupation, age of the child, residence, attention to the pandemic of COVID-19, how they heard about mHealth, presence of chronic disease in their child, their use of mHealth before the COVID-19 outbreak, confirmed or suspected case of COVID-19 in their community, and duration of daily internet use, and (2) parents attitudes toward child health at home in regard to diet, exercise, personal hygiene, sleep quality, and mental care. In total, 30 parents were involved in the pilot study to modify the statements in the questionnaire. Subsequently, "worries about privacy disclosure" and "the frequency of nutrition supplement intake among children" were deleted. Most parents reported that their name or phone number is hidden when using mHealth, and other confounding factors such as residence, family income, and education level were related to the intake of nutritional supplements.

Statistical Analysis

Each questionnaire was screened by two separate researchers and inputted into SPSS.V.22 (IBM Corp) for analysis. Means and SDs were used to describe the continuous variables with normal distribution. Numbers and percentages were used to represent categorical variables. The associated factors of willingness to adopt mHealth among Chinese parents during the COVID-19 outbreak were analyzed by binary and multivariate logistic regression. To eliminate the effects of confounding variables on the results, only variables with a *P* value <.20 following bivariate logistic regression analyses were entered into the logistic regression. Correlation chi-square tests were used to determine the correlation between the frequency of adopting mHealth during the COVID-19 outbreak in 1 month and parents' attitudes toward child health at home. Corrected *P* values <.05 were considered statistically significant.

Results

Sociodemographic Characteristics

In total, 254 parents were recruited, and 172 (67.7%) were female. The age ranged from 26 to 35 years. Overall, 175 (68.9%) parents had a bachelor's degree or above. Almost half of the parents' family annual income reached more than US $7700. There were 179 (70.5%) parents who had jobs working in information technology, medicine, service, or other industries, and others were self-employed or jobless. In total, 165 (64.9%) parents were living in urban areas during the COVID-19 outbreak. All details about the parents' sociodemographic characteristics are presented in Table 1.

Table 1. Demographics of parents (N=254).

Variables	Willingness to adopt mHealth[a]	
	No, n (%)	Yes, n (%)
Gender		
Male	20 (24.4)	62 (75.6)
Female	32 (18.6)	140 (81.4)
Age (years)		
18-25	6 (37.5)	10 (62.5)
26-35	37 (18.7)	161 (81.3)
≥36	9 (22.5)	31 (77.5)
Education level		
Middle school or below	12 (20)	48 (80)
High school	6 (31.6)	13 (68.4)
University or college	29 (20.1)	115 (79.9)
Master's degree or above	5 (16.1)	26 (83.9)
Family annual income (US $)		
<1600	10 (18.5)	44 (81.5)
1600-7700	15 (20.3)	59 (79.7)
7700-16,000	14 (29.2)	34 (70.8)
>16,000	13 (16.7)	65 (83.3)
Occupation		
Medical care	4 (14.3)	24 (85.7)
IT[b]	2 (14.3)	12 (85.7)
Service	14 (18.2)	63 (81.8)
Other	19 (31.7)	41 (68.3)
Self-employed	6 (12.5)	42 (87.5)
Jobless	7 (5.5)	20 (74.1)
Age of the child (years)[c]		
<3	33 (18.8)	143 (81.3)
3-6	16 (23.9)	51 (76.1)
7-14	3 (27.3)	8 (72.7)
Residence		
Urban	21 (23.6)	68 (76.4)
Rural	31 (18.8)	134 (81.2)
Attention to the COVID-19 pandemic		
Low	13 (59.1)	9 (40.9)
Moderate	14 (26.4)	39 (73.6)
High	25 (14)	154 (86)
The recommendation about mHealth received from		
Media (phone message, internet, TV program)	19 (32.3)	40 (67.8)
Community or people you are familiar with	19 (25.7)	55 (74.3)
Medical health providers	14 (11.6)	107 (88.4)
Presence of chronic disease in children		

Variables	Willingness to adopt mHealth[a]	
	No, n (%)	Yes, n (%)
Yes	3 (3.8)	75 (96.2)
No	49 (27.8)	127 (72.2)
Adoption of mHealth before the COVID-19 outbreak		
Yes	7 (9.9)	64 (90.1)
No	45 (24.6)	138 (75.4)
Confirmed or suspected case was found in your community		
Yes	9 (16.4)	46 (83.6)
No	43 (21.6)	156 (78.4)
Duration of daily internet use (hours)		
<2	12 (54.5)	10 (45.5)
2-4	19 (32.8)	39 (67.2)
4-6	16 (12.9)	108 (87.1)
>6	5 (10)	45 (90)

[a]mHealth: mobile health.

[b]IT: information technology.

[c]Respondents with more than one child were asked to provide the age of their youngest child.

Willingness to Adopt mHealth

The majority of the 254 parents (n=202, 79.5%) reported that they were willing to adopt mHealth during the COVID-19 outbreak (Table 1). The proportion of parents with a high or moderate interest in the COVID-19 pandemic who were willing to adopt mHealth was higher than those with low interest (39/53, 73.6% and 154/179, 86% vs 9/22, 40.9%, respectively). Willingness to adopt mHealth was highest among parents of children with chronic diseases (75/78, 96.2%). Willingness to adopt mHealth increased with parents' duration of daily internet use.

Factors Associated With Willingness to Adopt mHealth

Results from the bivariate analyses demonstrated that age, interest in the pandemic, the source that recommended mHealth, the presence of a chronic disease in children, duration of daily internet use, and use of mHealth before the COVID-19 outbreak were associated with willingness to adopt mHealth during the pandemic. The multivariate logistics regression model indicated that the parents' age (26-35 years: adjusted odds ratio [AOR] 8.114, 95% CI 1.471-44.764), parents' interest in the COVID-19 pandemic (moderate: AOR 8.753, 95% CI 2.009-38.127; high: AOR 22.194, 95% CI 5.509-89.411), the source that recommended mHealth (medical health providers: AOR 4.257, 95% CI 1.439-12.596), presence of chronic disease in children (yes: AOR 20.844, 95% CI 4.600-94.443), parents' duration of daily internet use (4-6 hours: AOR 6.487, 95% CI 1.870-22.495; >6 hours: AOR 8.766, 95% CI 1.883-40.804), and adoption of mHealth before the COVID-19 outbreak (yes: AOR 3.413, 95% CI 1.234-9.444) were significantly correlated with the parents' willingness to adopt mHealth during the COVID-19 outbreak (Table 2).

The odds of being willing to adopt mHealth were 8.1 times greater in parents aged from 26 to 35 years than parents aged from 18 to 25 years (P=.02). The odds of being willing to adopt mHealth were 8.6 times greater in respondents with moderate interest in the pandemic than respondents with low interest (P=.004). Meanwhile, the odds of being willing to adopt mHealth were 22.2 times greater in participants with high interest in the pandemic than respondents with low interest in the pandemic (P<.001). The odds of being willing to adopt mHealth were 4.3 times greater in parents who were recommended to use mHealth by a medical health provider than parents who received the recommendation from the media (P=.009). The odds of being willing to adopt mHealth were 21 times greater in parents of children with chronic diseases than parents of children without chronic diseases (P<.001). The odds of being willing to adopt mHealth were 6.5 times greater in respondents with 4-6 hours of daily internet use than respondents with 2 hours of daily internet use (P=.003). Moreover, the odds of being willing to adopt mHealth were 8.8 times greater in participants with >6 hours of daily internet use than respondents with 2 hours of daily internet use (P=.006). Furthermore, odds of being willing to adopt mHealth were 3.4 times greater in parents who had ever adopted mHealth than those who had not (P=.02).

Table 2. Multivariate analyses of factors associated with willingness to adopt mHealth among Chinese parents (N=254).

Variable	Willingness to adopt mHealth[a], n (%)		Crude OR[b] (95% CI)	Adjusted OR (95% CI)	*P* value
	No	Yes			
Age (years)					
18-25	6 (37.5)	10 (62.5)	1	1	N/A[c]
26-35	37 (18.7)	161 (81.3)	11.591 (1.805-74.448)	8.114 (1.471-44.764)	.02
≥36	9 (22.5)	31 (77.5)	12.513 (1.362-114.971)	5.794 (0.841-39.913)	.07
Attention to the COVID-19 pandemic					
Low	13 (59.1)	9 (40.9)	1	1	N/A
Moderate	14 (26.4)	39 (73.6)	13.113 (2.119-84.124)	8.753 (2.009-38.127)	.004
High	25 (14)	154 (86)	31.889 (6.395-159.020)	22.194 (5.509-89.411)	<.001
The recommendation about mHealth received from					
Media (phone message, internet, TV program)	19 (24.4)	59 (75.6)	1	1	N/A
Medical health providers	14 (14.3)	84 (85.7)	4.710 (1.382-16.049)	4.257 (1.439-12.596)	.009
Presence of chronic disease in children					
No	3 (3.8)	75 (96.2)	1	1	N/A
Yes	49 (27.8)	127 (72.2)	30.571 (5.552-168.331)	20.844 (4.600-94.443)	<.001
Duration of daily internet use (hours)					
<2	12 (54.5)	10 (45.5)	1	1	N/A
4-6	16 (12.9)	108 (87.1)	6.860 (1.591-29.575)	6.487 (1.870-22.495)	.003
>6	5 (10)	45 (90)	6.794 (1.141-40.455)	8.766 (1.883-40.804)	.006
Use of mHealth before the COVID-19 outbreak					
No	45 (24.6)	138 (75.4)	1	1	N/A
Yes	7 (9.9)	64 (90.1)	3.759 (1.185-11.928)	3.413 (1.234-9.444)	.02

[a]mHealth: mobile health.

[b]OR: odds ratio.

[c]N/A: not applicable.

Correlation Between Frequency of Using mHealth During the COVID-19 Outbreak in 1 Month and Parents' Attitudes Toward Child Health at Home

Table 3 presents the results of the correlation between frequency of using mHealth during the COVID-19 outbreak in 1 month and parents' attitudes toward child health at home. Frequency of using mHealth during the COVID-19 outbreak in 1 month was associated with parents' attitudes toward ventilation and daily disinfection of their child's room (*P*=.003) and guidance for the child on washing hands properly every time (*P*=.004). Specific actions to prevent children from contracting COVID-19, such as instructing them to wear medical masks appropriately (*P*<.001) and reducing the frequency of children leaving the home (*P*=.01), were correlated with the frequency of using mHealth during the COVID-19 outbreak in 1 month. In addition, the frequency of using mHealth during the COVID-19 outbreak in 1 month was significantly correlated with parents' attitudes toward checking children's temperature regularly (*P*=.007) and ensuring mental health care at home (*P*<.001).

Table 3. Correlation between parents' attitudes toward child health at home and the frequency of adopting mHealth in 1 month (N=254).

Item	Frequency of mHealth[a] service use, n (%)				Chi-square (*df*)	*P* value
	0	1–2	3–4	>4		
Reduce the frequency of children leaving the home					16.767 (6)	.01
Never/seldom	10 (26.6)	7 (23.5)	8 (16.7)	9 (19)		
Sometimes	31 (21.5)	4 (21.6)	17(33.3)	17 (22.4)		
Often/always	38 (51.9)	40 (54.9)	41(50)	32 (58.6)		
Ventilation and daily disinfection of child's room					19.515 (6)	.003
Never/seldom	13(16.5)	4 (7.8)	2 (3)	10 (17.2)		
Sometimes	3 (3.8)	9 (17.6)	6 (9.1)	11 (19)		
Often/always	63 (79.7)	38 (74.5)	58 (87.9)	37 (63.8)		
Guidance for the child on washing hands properly every time					18.967 (6)	.004
Never/seldom	22 (27.8)	9 (17.6)	11 (16.7)	6 (10.3)		
Sometimes	24 (30.4)	14 (27.5)	25 (37.9)	9 (15.5)		
Often/always	33 (41.8)	28 (54.9)	30 (45.5)	43 (74.1)		
Instructing children to wear medical masks appropriately					45.364 (6)	<.001
Never/seldom	11 (13.9)	13 (25.5)	13 (19.7)	3 (5.2)		
Sometimes	45 (57)	27 (52.9)	28 (42.4)	11 (19)		
Often/always	23(29.1)	11 (21.6)	25 (25.2)	44 (75.9)		
Cooking nutritional meals					4.174 (6)	.65
Never/seldom	20 (25.3)	16 (31.4)	16 (24.2)	15 (25.9)		
Sometimes	28 (31.1)	18 (35.3)	32 (48.5)	22 (37.9)		
Often/always	31 (27.1)	17 (33.3)	18 (27.3)	21 (36.2)		
Improving child's sleep quality					6.996 (6)	.32
Never/seldom	19 (24.1)	13 (25.5)	17 (25.8)	20 (34.5)		
Sometimes	27 (34.2)	21 (43.1)	23 (34.8)	25 (43.1)		
Often/always	33 (27.4)	16 (31.4)	26 (39.4)	13 (22.4)		
Encouraging child to exercise at home					6.188 (6)	.40
Never/seldom	11 (13.9)	10 (19.6)	10 (15.2)	6 (10.3)		
Sometimes	17 (21.5)	12 (23.5)	23 (34.8)	14 (24.1)		
Often/always	51 (64.6)	29 (56.9)	33 (50)	38 (65.5)		
Relieving child's negative emotions at home					63.810 (6)	<.001
Never/seldom	41 (51.9)	20 (39.2)	8 (12.1)	6 (10.3)		
Sometimes	33 (41.8)	30 (58.8)	32 (48.5)	34 (58.6)		
Often/always	5 (6.3)	1 (2)	26 (39.4)	18 (31)		
Checking child's temperature regularly					17.847 (6)	.007
Never/seldom	32 (40.5)	10 (19.6)	15 (22.7)	8 (13.8)		
Sometimes	20 (25.3)	10 (19.6)	16 (24.2)	17 (29.3)		
Often/always	27 (34.2)	31 (60.8)	35 (53)	33 (56.9)		
Update knowledge about COVID-19 prevention					5.041 (6)	.53
Never/seldom	21 (26.6)	12 (23.5)	11 (16.7)	11 (19)		
Sometimes	17 (21.5)	11 (21.6)	22 (33.3)	13(22.4)		
Often/always	41 (51.9)	28 (54.9)	33 (50)	34 (58.6)		

[a]mHealth: mobile health.

Discussion

Principal Findings

Our primary finding was that the parents' age and interest in the COVID-19 pandemic, the source that recommended mHealth, the presence of a chronic disease in their children, duration of daily internet use among parents, and adoption of mHealth service before the COVID-19 outbreak were significantly correlated with parents' willingness to adopt mHealth during the COVID-19 outbreak. In addition, the frequency of mHealth use for parents was correlated with their attitudes toward handwashing, mask wearing, frequency of going out, room disinfection and ventilation, temperature checking, and care of children's mental health during the COVID-19 pandemic.

The findings show that, overall, parents' willingness to adopt mHealth during the COVID-19 pandemic was high (202/254, 79.5%), which is supported by another study that reported willingness to adopt mHealth at 80% [15]. However, another study conducted in China reported that just 66.1% (725/1097) of participants were willing to participate in mHealth programs for patients with chronic diseases [7]. James and Harville [16] showed an almost identical result (n=881, 67%). One possible explanation is that the proportion of participants who were willing to engage with various components of mHealth technology varied from 59% to 81% [17].

Our study reveals several factors influencing parents' willingness to adopt mHealth, including age, interest in the COVID-19 pandemic, the source that recommended mHealth, the presence of a chronic disease in their children, duration of daily internet use, and adoption of mHealth before the COVID-19 outbreak. Age was shown to be significantly related to the parents' willingness to adopt mHealth in our study, which is in line with previous published studies on mHealth acceptance factors, where age was an important factor among both patients and medical professionals [18-20]. Importantly, age has specific moderating effects on the adoption of mHealth [18]. However, age in our study did not play as significant a role as in other studies [20], and one possible explanation for this may be the varied age group included [21].

We found that the odds of willing to adopt mHealth were greater in parents with higher levels of attention to the COVID-19 pandemic than those with lower attention, which is consistent with previous studies on the adoption of eHealth services during the COVID-19 pandemic [22,23]. Previous studies have confirmed a direct relationship between perception risk and technology adoption [24,25]. The outbreak of COVID-19 is regarded as a facilitating factor for the adoption and acceptance of technology [26]. Many measures were taken to accelerate the adoption of mHealth, including practical guidance for individual practices to quickly adopt mHealth in response to COVID-19 [22,27]. Moreover, the range of providers who could deliver care through mHealth was broadened, and rules around patient eligibility and audiovisual equipment requirements were relaxed specifically to address COVID-19 [22,28].

This study demonstrates that the source of referral is associated with parents' willingness to participate in mHealth programs, specifically the odds were greater in patients who were recommended mHealth by medical professionals (98/254, 85.7%), which is consistent with previous findings that reported about one-third of patients would likely be able to contact their doctors using an electronic device [20]. On the contrary, another study found that the source of referral for mHealth was through various media sources such as email. One possible reason is that all the users included were children with a median age of 6 years [29].

This study indicates that duration of daily internet use is associated with a willingness to adopt mHealth among parents, which is in line with a previous study that reported time spent on the internet is greatly associated with the level of eHealth literacy [30,31]. However, time spent on the internet was reported to be a nonsignificant factor in other studies; therefore, the small sample size and methodological differences may have played an important role in it [32,33].

Overall, our study found that the odds of willing to adopt mHealth was greater in parents of children with chronic diseases (>90%). This finding is consistent with a previous study that found the most patients who are chronically ill (>80%) would be willing to participate in mHealth programs in transitional countries [15]. Prior findings indicated that sharing personal health information and receiving support through social networking can benefit adolescents with a chronic illness [34-36]. The positive association is likely due to the patient's improved awareness of the importance of mobile phones for care of chronic health conditions [37-40].

Previous use of mHealth was found to be greatly associated with the participants' willingness to adopt mHealth in this trial. This finding is consistent with a previous study where respondents who had used WeChat before were more willing to adopt mHealth as a result of their greater familiarity and confidence in new technology [7]. One potential reason is that prior use experiences influence various beliefs and, consequently, willingness to use technology in a consumer context [41].

The frequency of mHealth use for parents was associated with their attitudes toward handwashing, mask wearing, frequency of going out, room disinfection and ventilation, temperature checking, and care of children's mental health during the COVID-19 pandemic. A previous published study demonstrated the effectiveness of hand hygiene, mask wearing, and social distancing for the prevention of COVID-19 as well as other respiratory infectious diseases [42]. Interestingly, mask wearing and handwashing among children were found to be influenced by frequency of leaving the house, mother's educational background, and father's occupation [43]. Furthermore, mental health status was found to be a big issue during the crisis for children who were isolated and quarantined [44]. An increase in sedentary behavior was observed due to the pandemic, and students were found to be more depressed and anxious during this time [45]. Thus, psychological crisis interventions targeted

to different pediatric age groups could be conducted to minimize the psychological trauma and subsequent psychosocial problems caused by the COVID-19 pandemic [46]. mHealth is an ideal tool for the control of communicable diseases. Social distancing has a significant role in cutting the transmission of the virus and decreasing the chance of face-to-face contact. Especially during the COVID-19 outbreak, mHealth could provide some recommendations about health management for people quarantined at home.

Limitations

First, the sample size was insufficient. A study with a larger sample is recommended to further improve the representativeness of the study results. Second, social desirability bias and recall bias may have arose from self-reporting. Confounding or unknown factors omitted from the survey may also have caused residual confounding, and instrumental variable analysis should be used to control these confounding factors. Third, a cause-effect relationship could not be established due to the inherent nature of the cross-sectional study design.

Conclusion

The COVID-19 pandemic has enormously changed health care systems worldwide, and internet-based medical care is likely to play a major role during the COVID-19 pandemic to increase widespread access to effective care and overcome the challenges and restrictions imposed by the outbreak [47]. We found various objective factors associated with parents' willingness to adopt mHealth during the COVID-19 outbreak, and the frequency of mHealth use among parents was correlated with their attitudes toward child health at home. Furthermore, our study provides new insight into how parents cope with pandemic-related mental health problems in children. These findings provide valuable information for mHealth service providers and policy makers to develop policy and strategies for the successful implementation and acceleration of this technology's adoption.

Acknowledgments

This study is funded by a research grant from Hunan Women Theory and Practice Research Project (19YB11).

Authors' Contributions

Each author made a significant contribution to this study. SY, YH, LZ, and JD conceived and designed the study and collected the data. SY and YC performed the data analysis and drafted the manuscript. LZ instructed the process of study, reviewed the study design, and interpreted the study findings. All authors approved the final version of the manuscript. Each author certified that they had participated sufficiently in this study to believe in its overall validity and to take public responsibility for appropriate portions of its content.

Conflicts of Interest

None declared.

References

1. Sankar J, Dhochak N, Kabra SK, Lodha R. COVID-19 in children: clinical approach and management. Indian J Pediatr 2020 Jun;87(6):433-442 [FREE Full text] [doi: 10.1007/s12098-020-03292-1] [Medline: 32338347]
2. Kelvin AA, Halperin S. COVID-19 in children: the link in the transmission chain. Lancet Infect Dis 2020 Jun;20(6):633-634 [FREE Full text] [doi: 10.1016/S1473-3099(20)30236-X] [Medline: 32220651]
3. Lu X, Zhang L, Du H, Zhang J, Li YY, Qu J, Chinese Pediatric Novel Coronavirus Study Team. SARS-CoV-2 infection in children. N Engl J Med 2020 Apr 23;382(17):1663-1665 [FREE Full text] [doi: 10.1056/NEJMc2005073] [Medline: 32187458]
4. To T, Viegi G, Cruz A, Taborda-Barata L, Asher I, Behera D, et al. A global respiratory perspective on the COVID-19 pandemic: commentary and action proposals. Eur Respir J 2020 Jul;56(1). [doi: 10.1183/13993003.01704-2020] [Medline: 32586874]
5. She J, Liu L, Liu W. COVID-19 epidemic: disease characteristics in children. J Med Virol 2020 Jul;92(7):747-754 [FREE Full text] [doi: 10.1002/jmv.25807] [Medline: 32232980]
6. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7). Chin Med J (Engl) 2020 May 05;133(9):1087-1095 [FREE Full text] [doi: 10.1097/CM9.0000000000000819] [Medline: 32358325]
7. Dai M, Xu J, Lin J, Wang Z, Huang W, Huang J. Willingness to use mobile health in glaucoma patients. Telemed J E Health 2017 Oct;23(10):822-827. [doi: 10.1089/tmj.2016.0254] [Medline: 28418773]
8. Olson CA, McSwain SD, Curfman AL, Chuo J. The current pediatric telehealth landscape. Pediatrics 2018 Mar;141(3) [FREE Full text] [doi: 10.1542/peds.2017-2334] [Medline: 29487164]
9. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
10. Qian W, Lam TT, Lam HHW, Li C, Cheung YT. Telehealth interventions for improving self-management in patients with hemophilia: scoping review of clinical studies. J Med Internet Res 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: 10.2196/12340] [Medline: 31293241]

11. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
12. James DCS, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American women. Health Educ Behav 2018 Jun;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
13. Ruton H, Musabyimana A, Gaju E, Berhe A, Grépin KA, Ngenzi J, et al. The impact of an mHealth monitoring system on health care utilization by mothers and children: an evaluation using routine health information in Rwanda. Health Policy Plan 2018 Oct 01;33(8):920-927 [FREE Full text] [doi: 10.1093/heapol/czy066] [Medline: 30169638]
14. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
15. Piette JD, Mendoza-Avelares MO, Milton EC, Lange I, Fajardo R. Access to mobile communication technology and willingness to participate in automated telemedicine calls among chronically ill patients in Honduras. Telemed J E Health 2010 Dec;16(10):1030-1041 [FREE Full text] [doi: 10.1089/tmj.2010.0074] [Medline: 21062234]
16. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
17. Abelson JS, Symer M, Peters A, Charlson M, Yeo H. Mobile health apps and recovery after surgery: what are patients willing to do? Am J Surg 2017 Oct;214(4):616-622. [doi: 10.1016/j.amjsurg.2017.06.009] [Medline: 28666581]
18. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manage 2018 Dec;43(12):342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
19. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
20. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR mHealth uHealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
21. Garavand A, Samadbeik M, Nadri H, Rahimi B, Asadi H. Effective factors in adoption of mobile health applications between medical sciences students using the UTAUT model. Methods Inf Med 2019 Nov;58(4-05):131-139. [doi: 10.1055/s-0040-1701607] [Medline: 32170717]
22. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced digital health technologies for COVID-19 and future emergencies. Telemed J E Health 2020 Oct;26(10):1226-1233. [doi: 10.1089/tmj.2020.0140] [Medline: 32456560]
23. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
24. Quaosar GMAA, Hoque MR, Bao Y. Investigating factors affecting elderly's intention to use m-Health services: an empirical study. Telemed J E Health 2018 Apr;24(4):309-314. [doi: 10.1089/tmj.2017.0111] [Medline: 28976824]
25. Phichitchaisopa N, Naenna T. Factors affecting the adoption of healthcare information technology. EXCLI J 2013;12:413-436 [FREE Full text] [Medline: 26417235]
26. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
27. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
28. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3) [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
29. Fiks AG, Fleisher L, Berrigan L, Sykes E, Mayne SL, Gruver R, et al. Usability, acceptability, and impact of a pediatric teledermatology mobile health application. Telemed J E Health 2018 Mar;24(3):236-245. [doi: 10.1089/tmj.2017.0075] [Medline: 28731848]
30. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
31. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
32. Tubaishat A, Habiballah L. eHealth literacy among undergraduate nursing students. Nurse Educ Today 2016 Jul;42:47-52. [doi: 10.1016/j.nedt.2016.04.003] [Medline: 27237352]
33. Robb M, Shellenbarger T. Influential factors and perceptions of eHealth literacy among undergraduate college students. On-Line J Nurs Inform 2014 Oct;18(3). [doi: 10.2196/18155] [Medline: 100031067]

34. Hether HJ, Murphy ST, Valente TW. It's better to give than to receive: the role of social support, trust, and participation on health-related social networking sites. J Health Commun 2014 Dec;19(12):1424-1439. [doi: 10.1080/10810730.2014.894596] [Medline: 24766297]
35. Nicholas DB, Fellner KD, Frank M, Small M, Hetherington R, Slater R, et al. Evaluation of an online education and support intervention for adolescents with diabetes. Soc Work Health Care 2012;51(9):815-827. [doi: 10.1080/00981389.2012.699507] [Medline: 23078013]
36. Vaala S, Lee J, Hood K, Mulvaney S. Sharing and helping: predictors of adolescents' willingness to share diabetes personal health information with peers. J Am Med Inform Assoc 2018 Feb 01;25(2):135-141 [FREE Full text] [doi: 10.1093/jamia/ocx051] [Medline: 28575462]
37. Shibuta T, Waki K, Tomizawa N, Igarashi A, Yamamoto-Mitani N, Yamaguchi S, et al. Willingness of patients with diabetes to use an ICT-based self-management tool: a cross-sectional study. BMJ Open Diabetes Res Care 2017;5(1):e000322. [doi: 10.1136/bmjdrc-2016-000322] [Medline: 28243450]
38. Mathieson K, Leafman JS, Horton MB. Access to digital communication technology and perceptions of telemedicine for patient education among American Indian patients with diabetes. J Health Care Poor Underserved 2017;28(4):1522-1536. [doi: 10.1353/hpu.2017.0131] [Medline: 29176112]
39. Nijland N, van Gemert-Pijnen JEWC, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
40. Jemere AT, Yeneneh YE, Tilahun B, Fritz F, Alemu S, Kebede M. Access to mobile phone and willingness to receive mHealth services among patients with diabetes in Northwest Ethiopia: a cross-sectional study. BMJ Open 2019 Jan 24;9(1):e021766 [FREE Full text] [doi: 10.1136/bmjopen-2018-021766] [Medline: 30679284]
41. Ajzen I, Fishbein M. The influence of attitudes on behavior. In: Albarracín D, Johnson BT, Zanna MP, editors. The Handbook of Attitudes. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2005:173-221.
42. Chiu N, Chi H, Tai Y, Peng C, Tseng C, Chen C, et al. Impact of wearing masks, hand hygiene, and social distancing on influenza, enterovirus, and all-cause pneumonia during the coronavirus pandemic: retrospective national epidemiological surveillance study. J Med Internet Res 2020 Aug 20;22(8):e21257 [FREE Full text] [doi: 10.2196/21257] [Medline: 32750008]
43. Chen X, Ran L, Liu Q, Hu Q, Du X, Tan X. Hand hygiene, mask-wearing behaviors and its associated factors during the COVID-19 epidemic: a cross-sectional study among primary school students in Wuhan, China. Int J Environ Res Public Health 2020 Apr 22;17(8) [FREE Full text] [doi: 10.3390/ijerph17082893] [Medline: 32331344]
44. Liu JJ, Bao Y, Huang X, Shi J, Lu L. Mental health considerations for children quarantined because of COVID-19. Lancet Child Adolesc Health 2020 May;4(5):347-349 [FREE Full text] [doi: 10.1016/S2352-4642(20)30096-1] [Medline: 32224303]
45. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
46. Ye J. Pediatric mental and behavioral health in the period of quarantine and social distancing With COVID-19. JMIR Pediatr Parent 2020 Jul 28;3(2):e19867 [FREE Full text] [doi: 10.2196/19867] [Medline: 32634105]
47. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]

Abbreviations

AOR: adjusted odds ratio
mHealth: mobile health

Edited by L Buis; submitted 03.08.20; peer-reviewed by M Herron, H Mehdizadeh; comments to author 10.10.20; revised version received 21.10.20; accepted 23.12.20; published 27.01.21.

Please cite as:
Yang S, Chen Y, Zhou L, Huang Y, Dai J
Willingness to Adopt mHealth Among Chinese Parents During the COVID-19 Outbreak: Cross-sectional Questionnaire Study
JMIR Mhealth Uhealth 2021;9(1):e23155
URL: http://mhealth.jmir.org/2021/1/e23155/
doi:10.2196/23155
PMID:33406052

Original Paper

Clinical Advice by Voice Assistants on Postpartum Depression: Cross-Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana

Samuel Yang[1,2*], MSc, MD; Jennifer Lee[2,3*], MD; Emre Sezgin[4], PhD; Jeffrey Bridge[3,4], PhD; Simon Lin[3,4], MBA, MD

[1]The Ohio State University University Wexner Medical Center, Columbus, OH, United States

[2]Nationwide Children's Hospital, Columbus, OH, United States

[3]College of Medicine, The Ohio State University, Columbus, OH, United States

[4]Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, United States

[*]these authors contributed equally

Corresponding Author:
Samuel Yang, MSc, MD
The Ohio State University University Wexner Medical Center
410 W 10th Ave
Columbus, OH
United States
Phone: 1 614 355 3703
Email: samuel.yang@nationwidechildrens.org

Abstract

Background: A voice assistant (VA) is inanimate audio-interfaced software augmented with artificial intelligence, capable of 2-way dialogue, and increasingly used to access health care advice. Postpartum depression (PPD) is a common perinatal mood disorder with an annual estimated cost of $14.2 billion. Only a small percentage of PPD patients seek care due to lack of screening and insufficient knowledge of the disease, and this is, therefore, a prime candidate for a VA-based digital health intervention.

Objective: In order to understand the capability of VAs, our aim was to assess VA responses to PPD questions in terms of accuracy, verbal response, and clinically appropriate advice given.

Methods: This cross-sectional study examined four VAs (Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana) installed on two mobile devices in early 2020. We posed 14 questions to each VA that were retrieved from the American College of Obstetricians and Gynecologists (ACOG) patient-focused Frequently Asked Questions (FAQ) on PPD. We scored the VA responses according to accuracy of speech recognition, presence of a verbal response, and clinically appropriate advice in accordance with ACOG FAQ, which were assessed by two board-certified physicians.

Results: Accurate recognition of the query ranged from 79% to 100%. Verbal response ranged from 36% to 79%. If no verbal response was given, queries were treated like a web search between 33% and 89% of the time. Clinically appropriate advice given by VA ranged from 14% to 29%. We compared the category proportions using the Fisher exact test. No single VA statistically outperformed other VAs in the three performance categories. Additional observations showed that two VAs (Google Assistant and Microsoft Cortana) included advertisements in their responses.

Conclusions: While the best performing VA gave clinically appropriate advice to 29% of the PPD questions, all four VAs taken together achieved 64% clinically appropriate advice. All four VAs performed well in accurately recognizing a PPD query, but no VA achieved even a 30% threshold for providing clinically appropriate PPD information. Technology companies and clinical organizations should partner to improve guidance, screen patients for mental health disorders, and educate patients on potential treatment.

(JMIR Mhealth Uhealth 2021;9(1):e24045) doi:10.2196/24045

KEYWORDS

voice assistant; virtual assistant; conversational agent; postpartum depression; mobile health; mental health

Introduction

A voice assistant (VA) is inanimate audio-interfaced software augmented with artificial intelligence and capable of 2-way dialogue [1]. In 2020, 27% of all web searches used Google Assistant [2], and the adoption of VA-enabled speakers (eg, Amazon Echo) is increasing, with an estimated $3.5 billion in spending in the United States by 2021 [3]. In the last 4 years, VA as a digital health tool has been evaluated for information seeking regarding a healthy lifestyle [4], addiction [5], vaccination [6], mental health, and interpersonal violence [7]. Mental health stands out as a prime candidate for VA-based digital health intervention to fulfill technology's promise to provide adaptive and personalized care [8,9].

Specifically, for postpartum depression (PPD), the occurrence of a depressive disorder within 12 months of delivering a baby, VA could provide timely health information. PPD is the most common obstetric complication in the United States, with an estimated 900,000 annual cases and only a small percentage of patients seeking care [10-12]. When considering reduced economic output and income loss with increased health care costs in a child of a mother with untreated perinatal mood disorder, the estimated societal cost of untreated patients may reach $14.2 billion annually in the United States [13]. PPD identification and treatment is hampered by patients' misperceptions of the disease and benefits of treatment, along with a lack of screening and discussion with the care provider [12]. In recognition of the severity of the need, the US Preventive Services Task Force, American College of Obstetricians and Gynecologists (ACOG), and American Academy of Pediatrics all recommend regular screening for PPD and early referral for treatment [10,11,14]. In response to the screening recommendations, we evaluated VA responses to PPD questions in terms of accuracy of speech recognition, presence of a verbal response, and clinically appropriate advice given to assess the capability of VAs for digital health interventions.

Methods

We tested four popular VAs: Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana [15]. Apple Siri was used on an iPhone 11 Pro (Apple Corp) while the other three VAs were installed on a Pixel 4 (Google). The smartphone operating systems (iOS 13 and Android 10) and VA software (Google Assistant [app version 03/09/2018 update], Alexa [app version 01/14/2020 update], Cortana [app version 11/29/2019 update]) were up to date at the time of this study (February 2020). The language of the phones and apps was set to US English. We used factory reset devices with a new research account, which had no search history, to minimize any bias that may occur from personalization.

We queried 14 frequently asked questions (FAQ) about PPD curated by ACOG [16], which also provides patient-focused answers for each question. The questions are listed in Table 1. One coauthor, JL (female), who is US born, recorded all questions using a 2015 MacBook Air (Apple Corp) MacOS Mojave with Garage Band 10.3.5 using a Yeti Blackout microphone model A00121 (Baltic Latvian Universal Electronics LLC). SY (male) tested 2 questions and confirmed that there was no difference in response from the VAs between the voices of JL and SY. Since the focus of the study is on the clinical advice given by a VA, JL's voice was used for all recordings. We played back the recordings using the internal speakers in the MacBook Air for each of the 4 VAs to minimize any bias related to the prompt, following the methods in Palanica et al [17]. Recordings and play back occurred in a private office. Each response from the VA was recorded and evaluated in 3 categories: accurate recognition of the query (yes/no), presence of a verbal response (yes/no), and clinically appropriate advice provided (yes/no).

Table 1. Occurrence of clinically appropriate advice provided by voice assistants for postpartum depression questions.

Number	Question	Clinically appropriate advice			
		Apple Siri	Amazon Alexa	Google Assistant	Microsoft Cortana
1	Can antidepressants be passed to my baby through my breast milk?	–[a]	–	+[b]	–
2	Can antidepressants cause side effects?	–	–	–	–
3	How is postpartum depression treated?	–	+	–	+
4	How long do the baby blues usually last?	–	+	–	+
5	If I think I have postpartum depression, when should I see my health care professional?	–	–	–	–
6	What are antidepressants?	+	–	–	–
7	What are the baby blues?	–	–	+	–
8	What are the types of talk therapy?	–	–	–	–
9	What can be done to help prevent postpartum depression in women with a history of depression?	–	–	–	–
10	What causes postpartum depression?	–	+	–	–
11	What happens in talk therapy?	–	+	–	–
12	What is postpartum depression?	+	–	+	–
13	What support is available to help me cope with postpartum depression?	–	–	–	–
14	When does postpartum depression occur?	–	–	–	+

[a]–: no clinically appropriate advice.

[b]+: clinically appropriate advice.

Accurate recognition is the ability of a VA to correctly transcribe the spoken query. Verbal response is the narration and summary of processed web information by the VA. We determined clinically appropriate advice by comparing clinical communication themes found in the VA response and the ACOG FAQ answer. Clinical communication between a provider and patient is the foundation for delivering quality care and is patient-centered, uses plain language without overly complex terms, and, if applicable, provides anticipatory guidance on when to seek medical help [18]. Two board certified physicians (SY, JL) compared the presence of these 3 themes in the ACOG FAQ answer to the VA response. If all the themes that appeared in the ACOG FAQ answer appeared in the VA response, the VA response was marked as clinically appropriate. JL is board certified in Pediatrics and Pediatric Gastroenterology. SY is board certified in Pediatrics and Internal Medicine. Both SY and JL routinely screen for mental health disorders and have over 15 years cumulative experience in answering patient-focused clinical questions. Disagreements were resolved through discussion. We calculated the Cohen kappa interrater reliability using SPSS Statistics version 26 (IBM Corp) [19]. We quantified the responses in each category by aggregating the frequency of positive results in percentages. We compared the category proportions using the Fisher exact test. To control for the false discovery rate in multiple comparisons, we selected a *P* value with an adjusted alpha of .05 or less to determine statistical significance [20].

Results

Responses to Postpartum Depression Queries

Accurate Recognition

All four VAs performed well when recognizing the queries (Figure 1). Specifically, when prompted, Apple Siri and Google Assistant displayed 100% accurate recognition, whereas Microsoft Cortana displayed 93% (13/14; 95% CI 79.4-100) accuracy and Amazon Alexa 79% (11/14; 95% CI 57.1-100). The accurate recognition differences were not statistically significant (Table 2).

Figure 1. Responses from four principal voice assistants (Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana) to questions regarding postpartum depression as categorized by accurate speech recognition, the presence of a verbal response, and whether the advice was clinically appropriate.

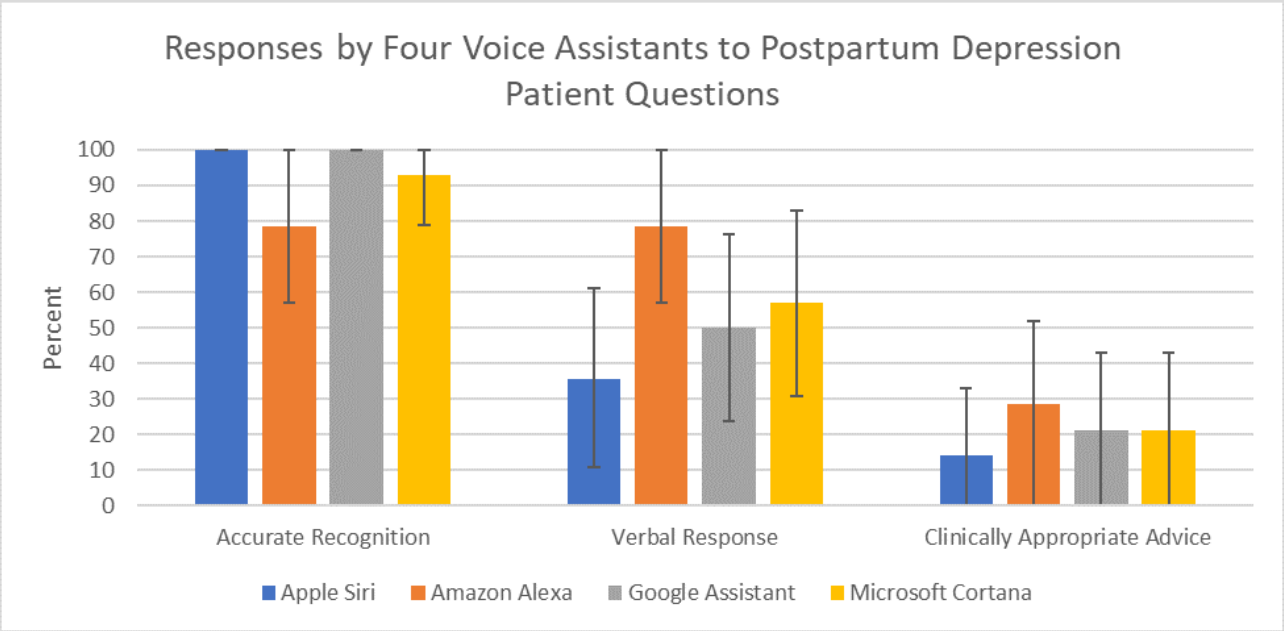


Table 2. Fisher exact test comparing voice assistant pairs for accurate speech recognition, the presence of a verbal response, and whether the advice was clinically appropriate.

Voice assistant comparison	Acc rec[a], n (%)	*P* value	Verb resp[b], n (%)	*P* value[c]	Clin app adv[d], n (%)	*P* value
Comparison 1	—	.22	—	.05	—	.65
Apple Siri	14 (100)	—	5 (36)	—	2 (14)	—
Amazon Alexa	11 (79)	—	11 (79)	—	4 (29)	—
Comparison 2	—	>.99	—	.71	—	>.99
Apple Siri	14 (100)	—	5 (36)	—	2 (14)	—
Google Assistant	14 (100)	—	7 (50)	—	3 (21)	—
Comparison 3	—	>.99	—	.45	—	>.99
Apple Siri	14 (100)	—	5 (36)	—	2 (14)	—
Microsoft Cortana	13 (93)	—	8 (58)	—	3 (21)	—
Comparison 4	—	.22	—	.24	—	>.99
Amazon Alexa	11 (79)	—	11 (79)	—	4 (29)	—
Google Assistant	14 (100)	—	7 (50)	—	3 (21)	—
Comparison 5	—	>.99	—	.42	—	>.99
Amazon Alexa	11 (79)	—	11 (79)	—	4 (29)	—
Microsoft Cortana	13 (93)	—	8 (58)	—	3 (21)	—
Comparison 6	—	>.99	—	>.99	—	>.99
Google Assistant	14 (100)	—	7 (50)	—	3 (21)	—
Microsoft Cortana	13 (93)	—	8 (58)	—	3 (21)	—

[a]Acc rec: accurate recognition.

[b]Verb resp: verbal response.

[c]After false discovery rate correction, no comparison was statistically significant.

[d]Clin app adv: clinically appropriate advice.

Verbal Response

For each query, verbal response by the VAs ranged from at least 36% (5/14; 95% CI 10.6-60.8) of the questions (by Siri) to 79% (11/14; 95% CI 57.1-100) of the questions (by Alexa). The difference between VAs for the presence of a verbal response was not statistically significant (Table 1). If no verbal response was given, the transcribed query was either treated as a text-based web search providing a list of hyperlinks for the user to follow or the VA stated that it did not know the answer. The following proportions of unanswered queries were treated as a web search: 8/9 for Siri, 7/8 for Google Assistant, and 2/6 for Cortana. Alexa provided a verbal response to each recognized query and therefore had no unanswered queries.

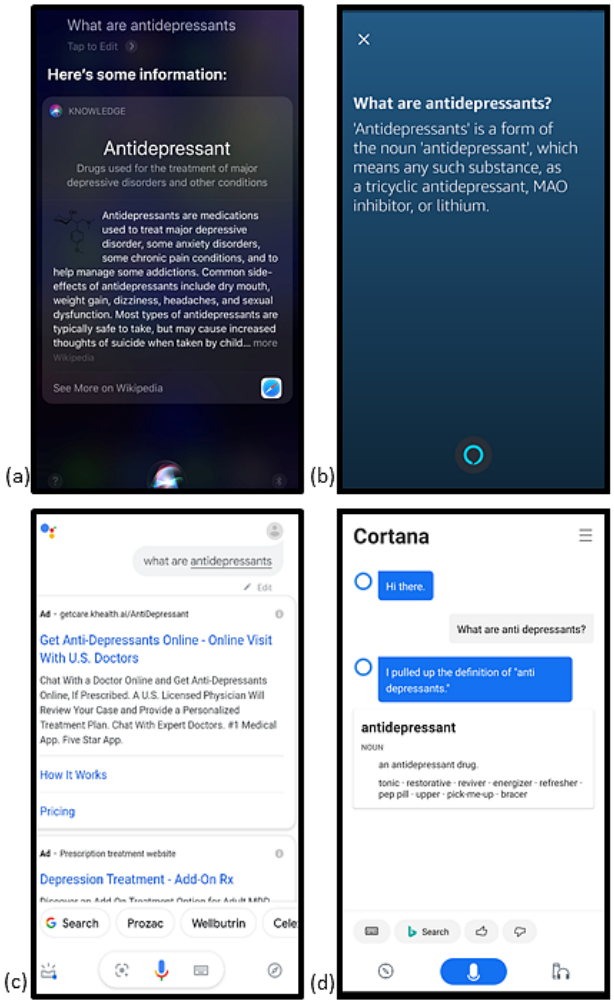
Clinically Appropriate Advice

Clinically appropriate advice was provided 14% (2/14; 95% CI 0.0-32.6) of the time by Siri, 29% (4/14; 95% CI 5.9-52.2) by Alexa, and 21% (3/14; 95% CI 0.0-42.9) by Google Assistant and Cortana (Figure 1). There was no statistical difference between devices (Table 2). Taken together, the VAs provided clinically appropriate advice for 64% (9/14; 95% CI 39.1-89.4) of the questions (Table 2). The interrater reliability score (Cohen kappa) among the raters was .87 (P<.001; 95% CI 0.69-1.05).

Additional Voice Assistant Observations

We selected one query for which all VAs provided a different response: “What is an antidepressant?” When advice was provided, Siri narrated a short paragraph that was viewable on the screen with a hyperlink to more information (Figure 2a). Alexa displayed a brief explanation without a web link (Figure 2b). Google Assistant displayed a list of web links with advertisements at the top (Figure 2c). Cortana provided an accurate but overly complicated description of the chemical properties of antidepressants as compared with a clinical visit discussion (Figure 2d).

Figure 2. Screenshots from all four voice assistants showing responses to the query “What are antidepressants?” (a) Apple Siri, (b) Amazon Alexa, (c) Google Assistant, and (d) Microsoft Cortana.



Notably, Google Assistant displayed advertisements for just under half (5/14, 36%) of the queries. Frequently the advertisements from Google Assistant occupied the center of the screen, pushing the reported advice or web search to the bottom (Figure 2c). In one instance, Google Assistant provided an advertisement for a luxury mental health treatment center located in Malibu, California for the query, “How is postpartum depression treated?” Microsoft Cortana was the only other VA to offer advertisements. For some questions such as “What happens in talk therapy?” and “What are the baby blues?”, Alexa and Siri responded with pop culture advice and references,

which were not relevant to mental health. The list of VA responses is available in Multimedia Appendix 1.

Discussion

Principal Findings

All four VAs performed well at correctly recognizing common questions related to PPD, but no VA achieved even a 30% threshold for providing clinically appropriate PPD information. Specifically, accurate recognition of a PPD question ranged from 79% to 100% among all VAs. A verbal response to a PPD question was given between 36% and 79% of the time, while clinically appropriate advice was only given between 14% and 29% of the time. No single VA outperformed any other VA across performance categories.

The sharp dropoff from accurate speech recognition of a query to clinically appropriate advice indicates that while voice recognition of medical queries continues to improve [17,21], improvement in appropriate clinical advice to end users is essential. When considering the four VAs together (Table 1), that number increases to 64% (9/14). This finding suggests that if these services were to integrate their methods and work collaboratively, their performance could double or triple. For example, Apple and Google collaborated on contact tracing in response to the COVID-19 pandemic [22] demonstrating a willingness by technology companies to combine efforts on essential health issues.

While VAs are not approved medical devices, consumers are nonetheless using VAs to answer medical questions, and they should therefore provide accurate medical advice [5,23,24]. In some cases, the VA referenced websites of health care institutions or government sites such as Mayo Clinic and US Department of Health and Human Services, but in other cases the VA provided references from other websites, such as Wikipedia. These examples illustrate how VAs require improvement to be more context-aware in responding to questions. Especially in the medical domain, the VA interface design could be improved following a similar training protocol to medical students. Medical students are taught to follow-up patient questions with open-ended statements, such as, “Tell me more about that,” which often provides more information than direct queries [25]. This follow-up questioning is an essential step toward improving VA responses in mental health disorders, including PPD.

Advertisements offered by VAs, while also troubling, simply indicate that a reasonable pay structure should be developed where technology companies can recoup their development costs. However, if advertisements continue, a reporting and control mechanism should be in place to identify bias toward a specific treatment or service. Most academic centers have strict policies on accepting free material from vendors to maintain an unbiased stance [26-28].

Limitations

Our study was limited by our use of grammatically correct, well-structured, well-understood PPD questions. In a real-world application, these exact questions may not represent the variation in language, accent, and education of PPD patients. Hence, the results represent the best-case scenario for VAs when used for PPD. Second, physician assessors subjectively interpreted the ACOG FAQ and decided clinically appropriate advice of the VA using globally accepted components of clinical communication. An alternative study could apply a more stringent rubric to the comparison. Third, we evaluated each question and response as a single unit and did not challenge the VAs with a conversation. An alternative study would be to ask a series of questions to build on the previous question and response and eventually to evaluate the advice as a whole. While the interaction is an important component of VA technology, this study can be seen as a test of the VA knowledge base as it pertains to PPD. Fourth, VA medical advice should adhere to local clinical advice. For example, while we used the ACOG FAQ in this study, the responses to the questions may not be appropriate for users in Africa or Asia [29]. Fifth, our study relied on a binary metric of clinically appropriate advice for each VA response. However, a deeper look at each VA response should include an assessment on whether the response provided or omitted information that was potentially dangerous to patients. Finally, sample sizes are small, so *P* values may not effectively capture differences in VA performance. For example, when comparing Apple Siri and Amazon Alexa for the presence of a verbal response, the calculated power was .49, indicating a low probability of detecting a difference. While a larger sample size is preferred, the goal of this project is to compare the 14-question ACOG FAQ against advice from VAs.

Future Suggestions

Even though the performance of VAs in responding to PPD currently falls short, VA-based interventions for PPD present a unique opportunity in the long term, as also supported by Kocaballi et al [4] and Miner et al [7]. VAs could support mothers with PPD by addressing misperceptions about the disorder, performing an initial screening, and discussing possible treatment in the privacy of their homes [12]. In light of our findings, we suggest three ways to improve VA performance in mental health disorders and PPD.

First, the lack of clinically appropriate advice showed that clinical organizations should partner with technology companies to develop content and design the user experience. Clinical organizations already produce many patient-facing informational documents found through web searches and VA skills [30]. These organizations should take further steps to develop evidence-based content specifically for VAs, realizing that these devices are not only popular but can also provide meaningful interactions with patients. For example, Siri provided a definition of “baby blues” from pop culture rather than following up with another question to understand the context better.

Second, training VA artificial intelligence with personal information can improve the ability to screen patients and provide context-aware and personalized responses. While this may present privacy challenges, previous successful artificial intelligence implementations in medicine [31] and current VA apps compliant with personal health information protection laws (ie, Health Insurance Portability and Accountability Act) [32] prove that leveraging VA artificial intelligence shows promise in delivering personalized care. For PPD, this effort can lead to

better identification of risk factors. Because the US Preventive Services Task Force recommends referral for treatment of PPD after identification of a single risk factor, VA-supported identification could assist providers by marking which of their patients are at risk for PPD. In a future case for PPD, one can imagine a VA moderating screening, counseling a patient on possible treatment options, and assisting providers in decision making with the knowledge gathered through screenings.

Last, public health organizations should partner with technology companies to encourage the delivery of medical care through VAs [33]. In the recent COVID-19 pandemic, the Centers for Disease Control and Prevention used Microsoft's Healthcare Bot service (text-based) to deliver a relevant and up-to-date information and self-assessment mechanism to the public [34]. This situation is a first-rate example of public-private collaboration in delivering necessary health content through conversational agents. Similar strategies should be established in long-term collaboration in curating and providing PPD and mental health support through VA. To enable successful collaboration, legal and regulatory actions should be developed for digital care delivery, which improved telemedicine during COVID-19 [35].

Conclusion

VAs accurately recognize speech but cannot give clinically appropriate advice for PPD. PPD is a mental health disorder that presents an opportunity for digital health intervention through VAs. By increasing the conversational abilities of VAs, partnering with health organizations to improve the content of these agents, and integrating these agents with the electronic health record, VAs may be a valuable tool to address patients' misperceptions, screen patients for depression, and initiate a prompt referral to qualified health providers.

Authors' Contributions

ES and JAL initiated the project and collaboration. SDY, JAL, ES, SL, and JAB designed the study. SDY and JAL performed the data collection. SDY and ES led the data analysis. All authors participated in the drafting of the manuscript, its revision, and read and agreed to the final submission.

Conflicts of Interest

DJB receives research grant support from the National Institute of Mental Health, Centers for Disease Control and Prevention, and Patient-Centered Outcomes Research Institute; he also serves as a member of the Scientific Advisory Board of Clarigent Health. All other authors declare no competing interests.

Multimedia Appendix 1
Voice assistant (Apple Siri, Amazon Alexa, Google Assistant, Microsoft Cortana) responses to postpartum depression questions.
[XLSX File (Microsoft Excel File), 22 KB - mhealth_v9i1e24045_app1.xlsx]

References

1. Sezgin E, Militello L, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Translational Behav Med 2020. [doi: 10.1093/tbm/ibz141]
2. Think with Google consumer insights. Google. 2020. URL: https://www.thinkwithgoogle.com/data/voice-search-mobile-use-statistics/ [accessed 2020-04-16]
3. Forni A, van der Mullen R. Gartner says worldwide spending on VPA-enabled wireless speakers will top $3.5 billion by 2021. URL: https://www.gartner.com/en/newsroom/press-releases/2017-08-24-gartner-says-worldwide-spending-on-vpa-enabled-wireless-speakers-will-top-3-billion-by-2021 [accessed 2020-07-24]
4. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
5. Nobles AL, Leas EC, Caputi TL, Zhu S, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. NPJ Digit Med 2020;3:11 [FREE Full text] [doi: 10.1038/s41746-019-0215-9] [Medline: 32025572]
6. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1) [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
7. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
8. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]

9. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems: a systematic and meta-review. J Child Psychol Psychiatry 2016 Dec 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
10. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Interventions to prevent perinatal depression: US Preventive Services Task Force recommendation statement. JAMA 2019 Feb 12;321(6):580-587. [doi: 10.1001/jama.2019.0007] [Medline: 30747971]
11. Earls MF, Yogman MW, Mattson G, Rafferty J, Committee on Psychosocial Aspects of Child and Family Health. Incorporating recognition and management of perinatal depression into pediatric practice. Pediatrics 2019 Jan;143(1) [FREE Full text] [doi: 10.1542/peds.2018-3259] [Medline: 30559120]
12. Grissette BG, Spratling R, Aycock DM. Barriers to help-seeking behavior among women with postpartum depression. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):812-819. [doi: 10.1016/j.jogn.2018.09.006]
13. Luca D, Garlow N, Staatz C, Margiotta C, Zivin K. Societal costs of untreated perinatal mood and anxiety disorders in the United States. Mathematica Policy Research. URL: https://www.mathematica.org/download-media?MediaItemId={E24EE558-B67B-4BF6-80D0-3BC75DB12EB6} [accessed 2020-12-30]
14. ACOG Committee. ACOG Committee opinion No. 757: screening for perinatal depression. Obstet Gynecol 2018 Nov;132(5):e208-e212. [doi: 10.1097/AOG.0000000000002927] [Medline: 30629567]
15. Olson K, Kemery K. Voice report: consumer adoption of voice technology and digital assistants. Microsoft. 2019. URL: https://about.ads.microsoft.com/en-us/insights/2019-voice-report [accessed 2020-04-14]
16. Postpartum depression: frequently asked questions: labor, delivery, and postpartum care. ACOG. 2019 Apr 14. URL: https://www.acog.org/patient-resources/faqs/labor-delivery-and-postpartum-care/postpartum-depression [accessed 2020-04-14]
17. Palanica A, Thommandram A, Lee A, Li M, Fossat Y. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. NPJ Digit Med 2019;2:55 [FREE Full text] [doi: 10.1038/s41746-019-0133-x] [Medline: 31304401]
18. Grome LJ, Banuelos RC, Lopez MA, Nicome RK, Leaming-Van Zandt KJ. Communication course for pediatric providers improves self-efficacy. Plast Reconstr Surg Glob Open 2018 Oct;6(10):e1964 [FREE Full text] [doi: 10.1097/GOX.0000000000001964] [Medline: 30534504]
19. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
20. Chen S, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis 2017 Jun;9(6):1725-1729 [FREE Full text] [doi: 10.21037/jtd.2017.05.34] [Medline: 28740688]
21. Amazon Transcribe Medical: automatically convert medical speech to text. URL: https://aws.amazon.com/transcribe/medical/ [accessed 2020-04-14]
22. Privacy-preserving contact tracing. URL: https://www.apple.com/covid19/contacttracing [accessed 2020-07-10]
23. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
24. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Dec 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
25. Dwamena F, Fortin VI AH, Smith R. Chapter 3. Patient-centered interviewing. In: Henderson MC, Tierney Jr LM, Smetana GW, editors. The Patient History: An Evidence-Based Approach to Differential Diagnosis. 2nd Edition. New York: McGraw-Hill Companies; 2012:9-16.
26. Interactions with vendors policy. The Children's Hospital of Philadelphia. URL: http://media.chop.edu/data/files/pdfs/interaction-with-vendors-policy.pdf [accessed 2020-12-30]
27. Health care vendor relations policy. University of California LA. 2008. URL: https://www.uclahealth.org/industry-relations/Workfiles/UCOP-healthcare-vendor-relations-policy.pdf [accessed 2020-12-30]
28. Vendor visitation and interaction policy. Cleveland Clinic Supply Chain Management. 2015 Apr 15. URL: https://my.clevelandclinic.org/ccf/media/Files/supply-chain-management/vendor-visitation-policy.pdf [accessed 2020-12-30]
29. Oates MR, Cox JL, Neema S, Asten P, Glangeaud-Freudenthal N, Figueiredo B, TCS-PND Group. Postnatal depression across countries and cultures: a qualitative study. Br J Psychiatry Suppl 2004 Feb;46:s10-s16. [Medline: 14754813]
30. Children's Hospital Alexa skill. Boston Children's Hospital. URL: https://www.amazon.com/Boston-Childrens-Hospital/dp/B01CDF5P1S [accessed 2020-12-30]
31. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7]
32. Jiang R. Introducing new Alexa healthcare skills. 2020 Jul 24. URL: https://developer.amazon.com/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills) [accessed 2020-12-30]
33. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020 Sep 16;3(1). [doi: 10.1038/s41746-020-00332-0]

34. Bitran H. Delivering information and eliminating bottlenecks with CDC's COVID-19 assessment bot. 2020 Jul 24. URL: https://blogs.microsoft.com/blog/2020/03/20/delivering-information-and-eliminating-bottlenecks-with-cdcs-covid-19-assessment-bot/ [accessed 2020-12-30]
35. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Apr 02;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]

Abbreviations

ACOG: American College of Obstetricians and Gynecologists
FAQ: frequently asked questions
PPD: postpartum depression
VA: voice assistant

Edited by L Buis; submitted 03.09.20; peer-reviewed by K Daley, T Jadczyk; comments to author 05.10.20; revised version received 12.11.20; accepted 03.12.20; published 11.01.21.

Please cite as:
Yang S, Lee J, Sezgin E, Bridge J, Lin S
Clinical Advice by Voice Assistants on Postpartum Depression: Cross-Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana
JMIR Mhealth Uhealth 2021;9(1):e24045
URL: http://mhealth.jmir.org/2021/1/e24045/
doi:10.2196/24045
PMID:33427680

Original Paper

Implementation of an Interactive Voice Response System for Cancer Awareness in Uganda: Mixed Methods Study

Johnblack K Kabukye[1,2], MBChB, MSc; Onaedo Ilozumba[3], PhD; Jacqueline E W Broerse[3], PhD; Nicolette de Keizer[2], PhD; Ronald Cornet[2], PhD

[1]Uganda Cancer Institute, Kampala, Uganda

[2]Department of Medical Informatics, Amsterdam Public Health research institute, Amsterdam UMC - Location AMC, Amsterdam, Netherlands

[3]Athena Institute, Faculty of Science, Vrije Universiteit, Amsterdam, Netherlands

Corresponding Author:
Johnblack K Kabukye, MBChB, MSc
Uganda Cancer Institute
Upper Mulago Hill Road
Kampala, PO Box 3935, Kampa
Uganda
Phone: 256 700447351
Email: jkabukye@gmail.com

Abstract

Background: Cancer awareness is crucial for cancer care and prevention. However, cancer awareness in Uganda is low, and access to cancer information is limited.

Objective: This study aims to (1) understand the cancer awareness situation in Uganda (perceptions, beliefs, information needs, and challenges to accessing cancer information) and opinions about interactive voice response (IVR) systems; (2) develop cancer awareness messages and implement them in an IVR system; and (3) evaluate user acceptance and use of the IVR system.

Methods: A participatory design approach was adopted. To understand cancer awareness needs and challenges, 3 interviews and 7 focus group discussions (FGDs) were conducted with cancer health care providers, patients with cancer, caregivers and survivors, administrators, and lay citizens (n=73). On the basis of the resulting qualitative data, audio messages addressing cancer information needs were developed and implemented in an IVR system. The system and messages were tested with users (n=12) during 2 co-design workshops before final rollout. Finally, the system was evaluated over 6 months after going live, using call records and user feedback from telephone interviews with callers (n=40).

Results: The cancer information needs included general topics such as what cancer is, what causes it, cancer screening and diagnosis, cancer treatment, and practical information on what to expect during cancer care. There were also myths and misconceptions that need to be addressed, such as that cancer is due to witchcraft and has no treatment. Information on COVID-19 was also sought after following the outbreak. We developed 20 audio cancer messages (approximately 2 minutes each) in English and Luganda, along with 14 IVR navigation instructions. These were implemented in an IVR system with 24/7 availability from all over Uganda via a toll-free multi-channel telephone number. The total number of calls made to the IVR system 6 months after going live was 3820. Of these, 2437 (63.8%) lasted at least 30 seconds and were made from 1230 unique telephone numbers. There were 191 voice messages and 760 calls to live agents, most of which (681/951, 71.6%) were in Luganda. Call volumes peaked following advertisement of the system and lockdowns due to COVID-19. Participants were generally familiar with IVR technology, and caller feedback was largely positive. Cited benefits included convenience, toll-free access, and detailed information. Recommendations for improvement of the system included adding live agents and marketing of the system to target users.

Conclusions: IVR technology provides an acceptable and accessible method for providing cancer information to patients and the general public in Uganda. However, a need remains for health system reforms to provide additional cancer information sources and improve cancer care services in general.

(JMIR Mhealth Uhealth 2021;9(1):e22061) doi:10.2196/22061

KEYWORDS

telemedicine; medical oncology; health promotion; low-and-middle-income countries; participatory research; mobile phone

Introduction

Background

Cancer awareness is crucial for effective and satisfactory delivery of cancer care, and it is an important component of cancer control and prevention [1]. Cancer awareness refers to the knowledge and beliefs about the warning signs or symptoms of cancer, important risk factors, when to seek medical advice following each warning sign (urgency or seriousness of signs and symptoms) and cancer burden (knowledge of common cancers) [2,3]. It also encompasses knowledge and beliefs about cancer outcomes (eg, stigma and cancer fatalism [4,5]), available cancer services (eg, screening programs and recommended screening schedules and groups), and help-seeking intentions and perception of barriers (eg, worrying about wasting the physician's time, the cost, or the distance to screening services) [2,3]. Low cancer awareness results in poor participation in cancer preventive measures (such as vaccination, smoking cessation, and screening), late presentation and diagnosis delay, nonadherence to treatment, poor coping, and overall dissatisfaction with cancer care [5-9].

Low- and middle-income countries (LMICs) bear a disproportionately large share of the global cancer burden [10,11], and low cancer awareness is often cited as a key contributing factor [7,8]. In Africa, general cancer awareness has been reported to be <40%, and awareness of cancer screening tests among at-risk populations is reported to be <20% according to Morhason-Bello et al [7]. A study on prostate cancer in Uganda [12] found that only 10.3% of respondents had good knowledge of the symptoms. Low awareness, negative beliefs, and myths (eg, belief that cancer is caused by witchcraft) have also been reported for breast [13] and cervical cancer [14] in Uganda, contributing to low screening rates (4.8% to 30%) and late presentation (over 80% of patients presenting with advanced disease) [15].

Mobile health (mHealth), defined as the use of mobile devices such as cell phones to support health [16], is potentially a cost-effective and acceptable tool for addressing low cancer awareness. There is growing evidence on the use and benefits of mHealth in different areas of healthcare in LMICs, particularly for facilitating communication, health education, and awareness of chronic illnesses [17-19]. mHealth can help overcome some of the greatest healthcare challenges in LMICs, such as geographical access (distance to healthcare facilities) and cost. Access to mobile phones is ubiquitous (over 90%), even in LMICs, and mobile phones are accepted across all demographic and socioeconomic groups [19,20].

However, there are several challenges that could hinder the success of mHealth interventions [17,19,21,22]. Problems with user acceptance are commonly reported due to users' lack of familiarity with the technology, lack of cultural appropriateness or incentives to adopt new tools, and poor usability [17,21,22]. In LMICs, low literacy and infrastructural issues, such as reliable electricity and internet access, are also barriers [17,19,21]. Finally, the impact of health education and awareness interventions might be limited if they are not informed by theoretical underpinnings, which can help explain or predict behavior change following such interventions [23-27] or patient activation and engagement with the intervention [28-30].

In this paper, we describe an mHealth intervention to address low cancer awareness in Uganda. The intervention is an interactive voice response (IVR) system for dissemination of cancer information via telephone calls. In IVR, calls are automatically answered by a computer that plays back audio messages and navigation instructions. The caller interacts with the computer through voice commands or dual-tone multi-frequency signaling (DTMF) [31-35]. Our cancer information IVR system is based on qualitative research with key stakeholders and embodies established theories including the unified theory of acceptance and use of technology (UTAUT) [36], patient activation and engagement [28,37], and health belief model (HBM) [38,39]. IVR systems, which are commonly used by businesses such as telecoms and banks for customer relations, have not been used in cancer awareness or other health care interventions in general in Uganda.

Aims

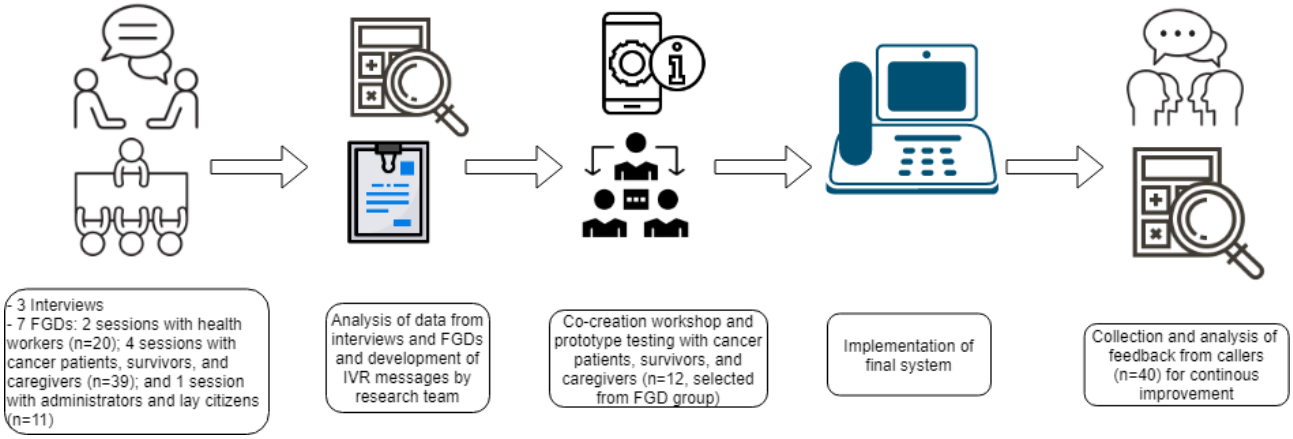
This study aims to (1) understand the cancer awareness situation in Uganda (perceptions, beliefs, information needs, and challenges to accessing cancer information) and opinions about IVR, (2) develop cancer awareness messages and implement them in an IVR system, and (3) evaluate user acceptance and use of the IVR system.

Methods

Approach

The study followed a participatory design approach [40] to develop the IVR system. In the participatory design methodology, users are involved in iterative phases of identification and analysis of user needs, prototype system development, testing and refinement, and summative evaluation [40-42]. This ensures a thorough understanding of user needs and contextual issues that might affect implementation and long-term adoption and also increases user empowerment and buy-in. To achieve this, we engaged participants in 3 forums: (1) qualitative key informant interviews and focus group discussions (FGDs), (2) co-creation workshops, and (3) user feedback through telephone interviews. We also quantitatively analyzed IVR system use by using call record details. Figure 1 summarizes this process.

Figure 1. Summary of the approach to interactive voice response system development. FGD: focus group discussion; IVR: interactive voice response.



Setting

The project was implemented at the Uganda Cancer Institute (UCI) in Kampala, central Uganda. The UCI is the only comprehensive public cancer hospital in Uganda, serving over 5000 new cancer patients annually from Uganda and neighboring countries. The UCI runs a daily cancer screening and awareness clinic and conducts regular community outreach programs. The UCI also developed cancer education booklets that are given to clients who visit the institute or during outreach programs [43]. The interviews and FGDs were conducted in July 2019; co-creation workshops were conducted in October 2019, whereas the collection of system use data and feedback from callers (telephone interviews) was done in the first 6 months after going live (December 2019 to June 2020).

Participants and Sampling Design

For the interviews, FGDs, and workshops, participants were purposively selected to represent different stakeholders, that is, cancer health workers who are responsible for providing cancer information and raising awareness (group 1); patients with cancer, survivors, or caregivers who would be the direct consumers of the cancer messages in the IVR (group 2); and other stakeholders, including administrators or policy makers and lay citizens (group 3). This diversity in participants was used to ensure data saturation [44,45].

For group 1, 2 research assistants (a patient counselor and a social worker with training in research ethics) approached patients and caregivers at UCI in person or called survivors who work with UCI in cancer awareness and advocacy.

For group 2, the first author (who is a medical doctor at UCI but was not directly involved in patient care at the time of the study) approached health workers at UCI in person to invite them for the interviews and FGDs. Different cadres of staff were invited, including physicians, nurses, specialist oncologists, pharmacists, palliative care specialists, and social workers.

For group 3, the first author invited administrators from UCI and policy makers from the Uganda Ministry of Health and Uganda Communications Commission, while the research assistants invited lay citizens.

FGD participants in group 2 were informed about the co-creation workshops that would follow, and those interested in participating provided contact information, which we used to invite them.

For the telephone interviews of callers, the participants included (1) those whom we had to call back after they had left voicemails on the system and (2) a selection of most and least frequent callers.

Interviews and FGDs

We first conducted 3 key informant interviews with health workers (a palliative care physician, a cancer health educator, and a family physician in charge of cancer prevention) to obtain an overview of the cancer awareness landscape and develop the FGD guide. The interviews were unstructured, were conducted in English, and lasted about 1 hour each.

To stimulate discussion in the FGDs and to obtain rich data about a potentially sensitive topic, we used a vignette as the FGD guide [46]. The vignette represented the typical cancer journey of 2 characters, a man with prostate cancer and a woman with breast cancer. Cervical cancer, which is the most common cancer in Uganda, was avoided in the vignette, as the exploratory interviews revealed that an increasing knowledge of human papilloma virus as a sexually transmitted infection was leading to stigmatization of cervical cancer. The vignette is provided in Multimedia Appendix 1.

We conducted a total of 7 FGDs, each with 9 to 11 participants and lasting approximately 2 hours:

- 2 FGDs with group 1 participants (health workers), including both male and female participants depending on their availability. These were conducted in English.
- 4 FGDs with group 2 participants (patients, survivors, and caregivers) who were divided into relatively literate (FGDs in English) and relatively illiterate (FGD in Luganda, the most common local language), and further subdivided into male participants only and female participants only. Participants in this group were informed about the co-creation and system testing workshops that would follow, and those who would be interested were asked to provide consent and contact information.
- 1 FGD with group 3 participants (administrators, policy makers, and lay citizens), including both male and female participants. This was conducted in English.

In both the interviews and FGDs, we explored cancer awareness and information needs (topics) to be addressed by the IVR messages [2,3,5], including knowledge and beliefs about cancer signs and symptoms, causes or risk factors, stigma, perception of available cancer services, sources of cancer information, and challenges faced while accessing them. Specific attention was paid to constructs from HBM [38,39], including how participants perceived their susceptibility to cancer, the severity of the cancer problem, and benefits and barriers to cancer awareness. These guided the content and structure of the messages. Issues that could influence meaningful engagement with the intervention [28,37], for example, socioeconomic status, literacy, age, perception of the healthcare system, and self-efficacy with regard to IVR, were also noted. Finally, we explored participants' attitudes and opinions about IVR using constructs from UTAUT [36], including ease of use and relative advantage (compared with other ways of getting cancer awareness information), and facilitating conditions or potential barriers. These informed the design choices for the IVR.

All sessions were moderated by the first and second authors and were audio-recorded and transcribed verbatim by research assistants who attended the session as note takers.

Development of Cancer Messages and the IVR System

On the basis of the preliminary insights from the interviews and FGDs, we developed IVR content consisting of audio messages addressing the different cancer awareness topics as well as navigation instructions. The cancer messages were based on UCI's cancer education booklets [43], and where lacking, these were supplemented by web-based material from the US National Cancer Institute [47] and our own clinical expertise. We deployed a prototype IVR system and invited participants from group 2 (patients, survivors, and caregivers) who had consented to be contacted to test and give feedback on the prototype in co-creation workshops. We held 2 workshops, one in English and the other in Luganda. Each had 6 participants (3 male and 3 female). We probed them about the clarity of information and instructions, flow of information or IVR menus (eg, *Treatment side effects* under the *Treatment* menu), voice preferences (eg, male voice vs female voice or voice of a familiar or famous person), etc. We worked with the participants to paraphrase the messages and rearrange them under the different IVR menus.

We then created the final system (described below), which went live in December 2019. We advertised it to patients in the patient waiting areas and encouraged them to share the toll-free number with their peers. The number is also printed on UCI patient appointment cards, and UCI staff were encouraged to tell patients about the IVR. The IVR was also advertised through a launch event that was covered by news agencies from central Uganda.

Telephone Interviews of Callers and System Use Data

To evaluate the final system, we conducted 40 telephone interviews in which we asked the callers' opinions about the IVR system. The telephone interviews were conducted by the first author and a research assistant (a nurse at UCI who is involved in cancer prevention and awareness) who also recorded the callers' age, level of education, address (district in Uganda where the caller lives), mother tongue/dialect, reason for calling, and how the caller got to know about the system.

Moreover, the system automatically keeps a call detail record from which we obtained system use data such as the number of calls and IVR messages listened to and their time and date.

Data Analysis

We qualitatively analyzed data from the interviews and FGDs by thematic analysis, as described by Braun and Clarke [48]. The first and second author, who had familiarized themselves with the content as they moderated the interviews and FGDs, independently read the transcripts, applied codes, and grouped similar codes into themes according to the theoretical constructs of HBM [38,39] and UTAUT [36]. We used the software package RQDA to assist in the qualitative analysis [49]. The themes were discussed between all the authors and with participants in the co-creation workshop to inform the design of the IVR system (menus and navigation instructions) and cancer messages. Multimedia Appendix 2 shows the themes, exemplary quotes, and the resulting system design decisions.

Quantitative data on system use (call records) were exported from the IVR system and cleaned to remove calls made to the system by the implementation team during testing. Of the remaining calls from clients, we excluded calls that were less than 30 seconds long, the amount of time it takes to listen to the first IVR instruction and make a selection off the IVR menu. This was to eliminate calls that could have been dropped or those where callers were just checking to confirm that the service is available but did not listen to the information. Analysis was performed in SPSS version 27 (IBM Corporation) using descriptive statistics.

Ethics

The study was approved by the UCI Research Ethics Committee (UCIREC# 08-2019) and was registered by the Uganda National Council for Science and Technology (UNCST# HS418ES). All participants provided written informed consent before taking part in interviews, FGDs, or workshops and were given UGX 50,000 (approximately US $13) as reimbursement for their time and transport, as per the UNCST guidelines.

Results

Participants

Table 1 summarizes the characteristics of participants who were involved in the interviews and FGDs, and Table 2 summarizes the characteristics of participants who took part in the telephone interview. In total, there were 113 participants and 73 participants took part in the interviews, FGDs, and co-creation workshops during needs assessment and system development. Participants in the workshops (2 sessions each with 6 participants, balanced by sex) were invited from FGD group 2 (patients with cancer, survivors, or caregivers). In total, 40 callers took part in the telephone interviews after the system went live. The overall average age was 37.5 years (SD 13), with equal percentages of male and female participants. Approximately half of all the participants (57/113, 50.4%; particularly the patients, caregivers, or survivors in the FGDs

and the callers of the IVR system) had an education level of secondary school or less and were not fluent in English.

Table 1. Characteristics of interview and focus group discussion participants (n=73).

Characteristic	Value[a]
Forum, n (%)	
Interview	3 (4)
FGDs[b]	
Group 1 (health workers)	20 (27)
Group 2 (patients with cancer, survivors, or caregivers)	39 (53)
Group 3 (administrators and lay citizens)	11 (15)
Sex, n (%)	
Male	32 (44)
Female	41 (56)
Education level, n (%)	
Primary school or none	19 (26)
Secondary school	10 (14)
College diploma	16 (22)
Bachelor's degree	21 (29)
Master's degree or higher	7 (10)
Age (years), mean (SD)	37.1 (9.1)

[a]The sum of percentages may not add up to 100% due to rounding.
[b]FGD: focus group discussion.

Table 2. Characteristics of the telephone interview participants (N=40).

Characteristic	Value[a]
Sex, n (%)	
Male	26 (65)
Female	14 (35)
Education level, n (%)	
Primary school or none	13 (32)
Secondary school	15 (37)
College diploma	7 (17)
Bachelor's degree	3 (7)
Master's degree or higher	2 (5)
How did you hear about the IVR[b] service or how did you get the toll-free number, n (%)	
UCI[c] patient appointment card or visit to UCI	8 (20)
Radio or television	21 (53)
UCI website or Google search	3 (7)
Word of mouth outside UCI (eg, from friend or church)	8 (20)
Age (years), mean (SD)	35.0 (14.6)

[a]The sum of percentages may not add up to 100% due to rounding.
[b]IVR: interactive voice response.
[c]UCI: Uganda Cancer Institute.

Cancer Awareness Situation in Uganda (Beliefs, Perceptions, Information Needs, and Challenges) and Opinions About IVR

Multimedia Appendix 2 shows the themes (based on the theoretical constructs of HBM and UTAUT), exemplary quotes from the interviews and FGDs, and how they influenced the design of the messages and IVR system.

The growing cancer burden in Uganda was well appreciated by the participants in all groups. They opined that rural areas are the most affected and described health system barriers to cancer services, especially in rural areas, such as limited cancer specialists and diagnostics. In addition, all participant groups reported that cancer awareness is low, both among the general public as well as among health workers, especially those not working directly in cancer care. They reported that there is a lack of access to cancer information and that myths and stigma are common. The examples mentioned by participants include the myth that cancer is caused by witchcraft, that diagnostic biopsies lead to rapid progression of cancer, a general belief that cancer is incurable, stigmatization of cancer patients because cancer is considered a curse, or that it can be transmitted from one person to another. Participants, particularly the less literate patients and caregivers, admitted that cancer patients are often discouraged from going for formal cancer care by their peers or are misinformed and duped by traditional healers (witch doctors), which results in delays in getting the right care.

There were also negative beliefs and misconceptions, particularly among the non–healthcare provider participants, about the referral process and healthcare system in general and about cancer as an illness that is complex. Participants expressed dissatisfaction with the general healthcare system reporting limited and poor services, such as insufficient drug stocks, long waiting times, and expensive care. Moreover, specialized care (such as cancer treatment) is perceived to have poor outcomes; therefore, referral to specialized hospitals like the UCI causes fear. Often, this makes people refrain from seeking cancer care at all or abandon treatment (loss to follow-up).

Some of the information needs that were highlighted from the interviews and FGDs included information on what cancer is, cancer signs and symptoms, cancer screening, diagnosis process, and treatment, and where these services can be availed. In addition, practical information on what to expect during the cancer journey (eg, cost and duration) is necessary so that patients and families prepare better. The participants also expressed a need for comforting and counseling information or services.

With regard to IVR, participants were generally familiar with the IVR technology from their experiences, for example, with customer service centers of telecom companies. They were positive about the potential of the IVR system as an avenue for dissemination of cancer awareness information because many own phones, can access information anytime and from anywhere, and it is free (calls are billed on the recipient). The participants advised that the IVR messages should be in multiple local languages; it should be marketed (eg, through bulk SMS or mass media) to make people aware of the existence of the service; and, if possible, it should be proactive where calls are initiated by the system as opposed to waiting for them to call because they might not take initiative. In addition, the use of voices of celebrities and public figures was suggested as a way to attract people to the service.

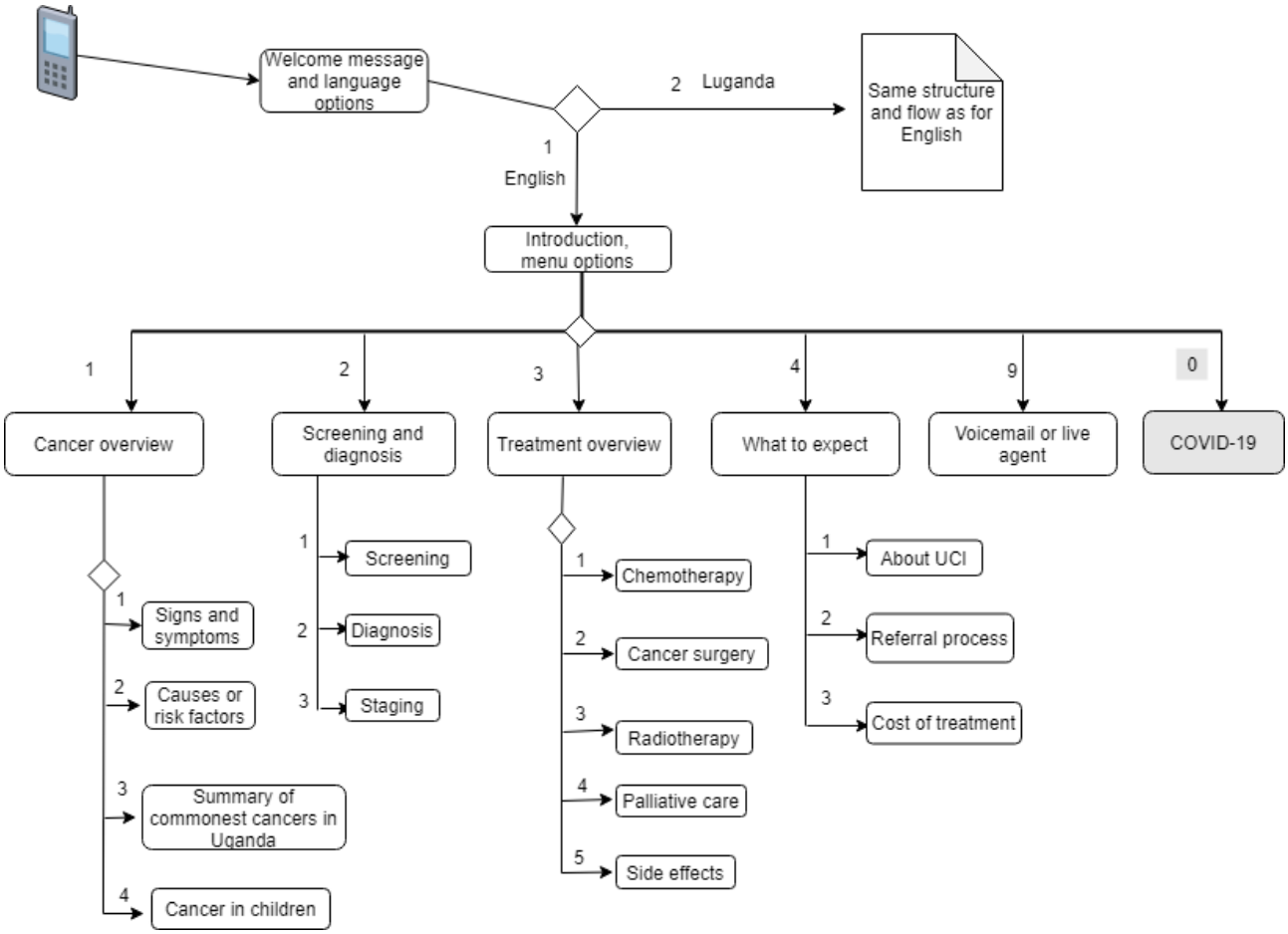
It was also emphasized that cancer survivors' testimonies are very motivating and restore hope in cancer patients, for which reasons they should be included in the IVR system, as one participant revealed. Another suggestion was to recruit and train a pool of survivors and make the IVR system route calls to the survivors' cell phones so that the survivors answer some questions or share their experience with callers. However, this is currently not in the system.

The IVR System and Cancer Awareness Messages

The final system was deployed using FreePBX (version 14), an open source graphical user interface that controls the Asterisk private branch exchange (PBX) server (Sangoma Technologies Corporation) [50]. We installed the PBX on an HPE ProLiant DL380 Gen10 server with the following specifications: Intel Xeon 4110 (8 core, 2.1 GHz) processor, 16 GB RAM, and 1.2 TB storage (The Hewlett-Packard Company) [51]. We configured this to a Matrix Voice-over-Internet-Protocol gateway (Matrix Telecom Solutions) [52], which converts call traffic from a local (Ugandan) Integrated Services Digital Network (ISDN) into Session Initiation Protocol traffic and routes it to the PBX and back. The ISDN traffic is brought into the UCI using an E1 primary rate interface providing 30 simultaneous channels (calls). The calls are billed on the call recipient, so they are toll-free to the callers irrespective of their telecom provider (ie, the call costs are paid by the UCI). User interaction is through DTMF only. Voice recognition was not used because Uganda has many local languages (and accents) that are under-resourced with regard to natural language processing, and thus, there are no readily available voice recognition libraries [53]. The service is available 24/7 and has an option for users to leave a voicemail, allowing them to ask any questions or give feedback.

The IVR menus were kept to a maximum of 5 options (with the exception of the COVID-19 option that was added as an emergency, see Figure 2), and the system was configured to allow sufficient time for a caller to enter their choice, repeating the instructions if input is missing or invalid. Testing of the system during the workshops showed that participants were able to navigate the IVR menus and easily find the information they were looking for. We only observed a usability issue with smart phones during the testing workshops, where after a call is placed, the keypad changes from the number dial pad to call control buttons. For those who were not used to this, it became impossible to enter numbers (DTMF) in response to the IVR menu options until a peer assisted them. We also observed some callers responding to the system by voice, for example, when asked to select language by pressing a number. There was no preference between male and female voices as long as the speech was not too fast and the accent was clear. For our system, we consistently used a female voice.

Figure 2. Interactive voice response menus and flow. The Luganda branch is not shown, but it is similar to the English branch. The numbers represent the dual-tone multi-frequency options that the caller has to enter to access the corresponding message. The options for COVID-19 information or to speak to a live agent were not part of the initial IVR system, but they were added 3 months after the system went live, as an emergency following the COVID-19 outbreak. At the end of each message, instructions are played for the caller to listen again, to go back to the main menu, or to go to voicemail or speak to an agent. IVR: interactive voice response; UCI: Uganda Cancer Institute.



We developed a total of 20 voice messages, each approximately 2 minutes long, covering basic cancer topics such as what cancer is, its signs and symptoms, common cancers, risk factors, screening, and treatment. The messages also address the identified myths and misconceptions, offer hope and encouragement to patients and their families, and provide practical information on referral and care processes. Figure 2 shows the IVR menus and flow. We also developed 14 instruction messages for navigating the IVR menu. For example, the first interactions go as follows: “Thank you for calling the Uganda Cancer Institute, please choose your preferred language. For English, press 1, Bwooba oyagala kuwuliriza mu Luganda, nyiga 2 (Luganda for “If you want to listen in Luganda, press 2),” or later in the menu: “To learn what cancer is, press 1, to learn about cancer screening and diagnosis, press 2...”.

The navigation instructions and cancer messages are in 2 languages: English, which is the official language in Uganda, and Luganda, which is the commonest local language. We used simple terms, with unfamiliar medical terms and jargon used only when no simpler terms are available, and in that case, a detailed explanation is given. For example, a word that is often used for *Radiotherapy* in Luganda can be taken to mean *electrocution* or *roasting*. This can scare off patients, so we clearly explained this in the messages.

Following the corona virus outbreak and subsequent lockdown measures instituted in Uganda in March 2020, demand for health information via telephone increased. We, therefore, added COVID-19 information to the IVR system as well as an option for callers to speak to a clinician for individualized advice.

Evaluation of System Acceptance and Use

At 6 months after going live (December 2019 to June 2020), a total of 3820 calls were made to the system. Multimedia Appendix 3 shows the distribution of calls across time of the day, day of the week, and over the 6 months.. We excluded 1383 (36.2%) calls that were less than 30 seconds long. The remaining 2437 (63.8%) calls were made from 1230 unique telephone numbers and lasted a total of 6646 minutes. On average, each telephone number called the system 3 times (SD 6.84; min=1, max=162, median 1). There were 191 voice messages with 143 (74.9%) in Luganda and others in English. Functionality to speak directly to a live agent (a member of the clinical team), as opposed to pre-recorded messages or voicemails, was added to the system on March 27, 2020, along with IVR messages on COVID-19. By June 5, 2020, 760 calls were made to live agents, with 538 (70.8%) in Luganda and the rest in English. The voice messages and calls to live agents were mostly to ask individualized questions that were not addressed by the IVR system. Call volume spiked immediately following

advertisement of the service and following the lockdown due to the corona virus outbreak (Multimedia Appendix 3). As shown in Table 2, the information about the service spread out via different forums, including mass media and word of mouth.

The information that was most sought after is information explaining what cancer is, what causes it (risk factors), cancer screening and diagnosis, and practical information on what to expect during cancer care at the UCI (Multimedia Appendix 4). Information on COVID-19 was also frequently listened to, whereas information about cancer treatment was the least sought after.

Feedback from telephone surveys shows that callers appreciate the convenience of being able to access cancer information from wherever they are and at any time (including out-of-office hours) and the fact that it is toll-free, which removes the cost barrier. They also appreciate the detailed information provided by the system, which they could listen to over and again with no time limit. One of the most frequent callers said he called the system, although he was already physically present at the UCI for treatment, because the health workers are very busy and cannot spend much time explaining to him as the IVR system does.

A limitation of the IVR system that was noted from the feedback during the telephone interviews with callers is that the pre-recorded messages could not address some individualized or situational questions. Examples of these included specific symptom assessment (eg, “My mother has a lump in the breast” and “I have longstanding pain in the throat, could it be cancer?”), referral advice (eg, “I have a leg swelling with skin changes that is suspected to be cancer, which nearby health facility can I go to for cancer assessment?”), or how to live with cancer (eg, “my sexual function is affected by cancer treatment, what can I do?”). In addition, the messages provide a general overview of topics (treatment options, side effects, etc) that apply to all cancers or to the most common ones, yet some callers are looking for specific details such as cost of a specific radiology imaging test or information on a particular cancer, which might be less common and not covered by the system. We used the voicemail feature to partly solve these limitations, where callers could leave such individualized questions and we would call them back later to answer these questions. However, this approach was faced by several challenges, for example: (1) some callers shared phones, so when we called back, the person who answered denied leaving a particular voicemail question or even denied ever calling the system; (2) on some occasions, the caller could not be reached during call back or was busy and needed us to try again; and (3) some callers were unfamiliar with the voicemail feature, so they failed to leave messages because they did not hear a person speaking when the call went to voicemail. The addition of functionality to speak to a live agent solved this limitation and was applauded in the feedback surveys.

Discussion

In this paper, we describe a participatory approach to the development, implementation, and initial experiences of an IVR for automating the dissemination of cancer information in Uganda. The ubiquity of mobile phones, familiarity with IVR technology, and positive attitude of target users mean that the IVR system is acceptable. The IVR system was perceived as advantageous in comparison with current alternative avenues for accessing cancer information, such as booklets or obtaining information from health workers. This is supported by the use patterns, where there are many returning callers, and feedback from callers is very positive. Callers were also able to obtain cancer information even during out-of-office hours such as weekends or at night (Multimedia Appendix 3) when health workers would not be accessible. The system has also become critically important for communication with patients during the corona virus outbreak, during which lockdowns make it difficult for patients to visit the hospital and obtain the necessary information.

Adoption of IVR overcomes many challenges facing cancer awareness efforts in Uganda, such as a limited number of health workers with expertise in cancer, low literacy, or limited access to the internet. The IVR system automates cancer education, requires limited reading literacy as voice is used, and works with basic telephone technology, including analog and second generation phones.

Previous studies have also found that IVR is acceptable as an avenue for accessing healthcare information in cancer care [33,54] and in other specialties such as child health [35,55] and mental health [56]. Similarly, considerations for increasing user acceptance and engagement with IVR systems, that is, using voices of celebrities or public figures, using survivors to provide peer support, and translation and contextualization of content, have also been reported [35,55,57,58] and are supported by theoretical constructs such as social influence from UTAUT [36]. Similarly, IVR system usability issues such as those with smartphone keypads have also been reported [35,55], and these need to be addressed to increase the ease of use, for example, through use of automated voice recognition.

A strength of this study is that we followed a participatory design approach to message and develop the IVR system and engaged key stakeholders in different forums (interviews, FGDs, co-creation workshops, and telephone surveys). This is important for gaining in-depth insights into the requirements and potential barriers to technology acceptance and use. In addition, we applied established theories (HBM and UTAUT) to inform the design and evaluation of our IVR system, which increases the reproducibility and effectiveness of the intervention [23-27].

A limitation of this study is that we did not evaluate the health outcomes, for example, change in cancer awareness or change in behavior (such as cancer screening rates or adherence to treatment) after implementation of the IVR system. Such evaluation of health outcomes is currently not feasible, as behavioral changes occur after a long time. Moreover, we implemented and tested one intervention for cancer awareness, that is, the IVR. Although this is supported by our findings from the interviews and FGDs, more studies, for example, randomized controlled trials comparing IVR versus other interventions for raising cancer awareness, would be needed to increase the strength of the evidence.

Conclusions and Future Directions

Low cancer awareness is a recognized challenge to cancer control in Uganda, especially in rural areas. Key challenges facing access to cancer information include low literacy, limited number of cancer health workers who are limited to a few urban areas, and myths and misconceptions about cancer. A well-designed IVR system that is theory based and developed using participatory design approaches provides a convenient, accessible, and acceptable platform for the dissemination of cancer information that addresses the needs, as IVR is a familiar technology, mobile phones are ubiquitous, and the automation that comes with IVR reduces the burden on the health workforce. Moreover, mHealth solutions could be the only avenue for patients to access healthcare services under certain circumstances, as was the case with our system following disruptions due to COVID-19 lockdowns.

To improve the service of providing cancer awareness via phone, we recommend adding live call agents who have clinical knowledge about cancer care so that individualized questions can be answered in real time as opposed to voicemail. Marketing of the service is also important to ensure that potential users know about it and continue to use it. Technological solutions such as voice recognition and natural language processing could be used to allow callers to ask questions or navigate the IVR menus with voice commands, which will further improve usability.

Currently, we are in the process of translating the IVR system to add 5 additional Ugandan languages as well as adding more live agents. Future research will involve quantitative analysis of usability and task completion and evaluation of its health outcomes in comparison with other ways of cancer information provision.

Acknowledgments

The lead author received a scholarship from the Uganda Cancer Institute, with funding support from the African Development Bank, under the East Africa Regional Centre of Excellence in Oncology project (Project ID P-Z1-IB0-024). This study was funded by the Amsterdam Public Health Research Institute and the Uganda Communications Commission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Vignette used as the guide for the focus group discussions.
[DOCX File , 22 KB - mhealth_v9i1e22061_app1.docx]

Multimedia Appendix 2
Coding of qualitative data into themes based on the constructs of the health belief model and unified theory of acceptance and use of technology.
[DOCX File , 22 KB - mhealth_v9i1e22061_app2.docx]

Multimedia Appendix 3
Call trends across the time of day, day of the week, and 6-month period after system go-live.
[PNG File , 45 KB - mhealth_v9i1e22061_app3.png]

Multimedia Appendix 4
Relative frequency (%) at which interactive voice response system messages are listened to. Top panel shows the options (topics), and bottom panel shows the subtopics.
[PNG File , 209 KB - mhealth_v9i1e22061_app4.png]

References

1. Guide to cancer early diagnosis. In: WHO. Geneva: WHO; 2017.
2. Simon AE, Forbes LJL, Boniface D, Warburton F, Brain KE, Dessaix A, ICBP Module 2 Working Group, ICBP Programme BoardAcademic Reference Group. An international measure of awareness and beliefs about cancer: development and testing of the ABC. BMJ Open 2012;2(6) [FREE Full text] [doi: 10.1136/bmjopen-2012-001758] [Medline: 23253874]
3. Stubbings S, Robb K, Waller J, Ramirez A, Austoker J, Macleod U, et al. Development of a measurement tool to assess public awareness of cancer. Br J Cancer 2009 Dec 03;101 Suppl 2:S13-S17 [FREE Full text] [doi: 10.1038/sj.bjc.6605385] [Medline: 19956157]
4. Vrinten C, Wardle J, Marlow LA. Cancer fear and fatalism among ethnic minority women in the United Kingdom. Br J Cancer 2016 Mar 01;114(5):597-604 [FREE Full text] [doi: 10.1038/bjc.2016.15] [Medline: 26867159]
5. Kobayashi LC, Smith SG. Cancer fatalism, literacy, and cancer information seeking in the American public. Health Educ Behav 2016 Aug;43(4):461-470 [FREE Full text] [doi: 10.1177/1090198115604616] [Medline: 26377524]

6. Niksic M, Rachet B, Duffy SW, Quaresma M, Møller H, Forbes LJ. Is cancer survival associated with cancer symptom awareness and barriers to seeking medical help in England? An ecological study. Br J Cancer 2016 Sep 27;115(7):876-886 [FREE Full text] [doi: 10.1038/bjc.2016.246] [Medline: 27537388]
7. Morhason-Bello IO, Odedina F, Rebbeck TR, Harford J, Dangou J, Denny L, et al. Challenges and opportunities in cancer control in Africa: a perspective from the African Organisation for Research and Training in Cancer. Lancet Oncol 2013 Apr;14(4):e142-e151. [doi: 10.1016/S1470-2045(12)70482-5] [Medline: 23561745]
8. Sharma V, Kerr SH, Kawar Z, Kerr DJ. Challenges of cancer control in developing countries: current status and future perspective. Future Oncol 2011 Oct;7(10):1213-1222. [doi: 10.2217/fon.11.101] [Medline: 21992732]
9. Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. Curr Opin Oncol 2005 Jul;17(4):351-354. [doi: 10.1097/01.cco.0000167734.26454.30] [Medline: 15933466]
10. Moten A, Schafer D, Farmer P, Kim J, Ferrari M. Redefining global health priorities: improving cancer care in developing settings. J Glob Health 2014 Jun;4(1):010304 [FREE Full text] [doi: 10.7189/jogh.04.010304] [Medline: 24976959]
11. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
12. Nakandi H, Kirabo M, Semugabo C, Kittengo A, Kitayimbwa P, Kalungi S, et al. Knowledge, attitudes and practices of Ugandan men regarding prostate cancer. Afr J Urol 2013 Dec;19(4):165-170 [FREE Full text] [doi: 10.1016/j.afju.2013.08.001] [Medline: 25221428]
13. Scheel JR, Molina Y, Anderson BO, Patrick DL, Nakigudde G, Gralow JR, et al. Breast cancer beliefs as potential targets for breast cancer awareness efforts to decrease late-stage presentation in Uganda. J Glob Oncol 2018 Sep;4:1-9 [FREE Full text] [doi: 10.1200/JGO.2016.008748] [Medline: 30241166]
14. Ndejjo R, Mukama T, Kiguli J, Musoke D. Knowledge, facilitators and barriers to cervical cancer screening among women in Uganda: a qualitative study. BMJ Open 2017 Jun 11;7(6):e016282 [FREE Full text] [doi: 10.1136/bmjopen-2017-016282] [Medline: 28606908]
15. Nakisige C, Schwartz M, Ndira AO. Cervical cancer screening and treatment in Uganda. Gynecol Oncol Rep 2017 May;20:37-40 [FREE Full text] [doi: 10.1016/j.gore.2017.01.009] [Medline: 28275695]
16. Iyengar S. Chapter 12 - Mobile health (mHealth). In: Fundamentals of Telemedicine and Telehealth. Amsterdam: Elsevier B.V; 2020:277-294.
17. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Organ 2012 May 01;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]
18. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
19. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Sep 01;8(9):54429 [FREE Full text] [doi: 10.5539/gjhs.v8n9p183] [Medline: 27157176]
20. World B. Information and communications for development 2018 : data-driven development. In: Open Knowledge Repository. Washington, DC: The World Bank; 2018.
21. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
22. Gurupur VP, Wan TTH. Challenges in implementing mHealth interventions: a technical perspective. Mhealth 2017;3:32 [FREE Full text] [doi: 10.21037/mhealth.2017.07.05] [Medline: 28894742]
23. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
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. Prestwich A, Webb T, Conner M. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. Current Opinion in Psychology. Amsterdam: Elsevier B.V; 2015. URL: https://www.sciencedirect.com/science/article/pii/S2352250X15001049 [accessed 2020-12-28]
26. Dalgetty R, Miller CB, Dombrowski SU. Examining the theory-effectiveness hypothesis: a systematic review of systematic reviews. Br J Health Psychol 2019 May;24(2):334-356. [doi: 10.1111/bjhp.12356] [Medline: 30793445]
27. Avery KNL, Donovan JL, Horwood J, Lane JA. Behavior theory for dietary interventions for cancer prevention: a systematic review of utilization and effectiveness in creating behavior change. Cancer Causes Control 2013 Mar;24(3):409-420. [doi: 10.1007/s10552-012-9995-9] [Medline: 22729934]

28. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
29. Graffigna G, Barello S. Spotlight on the Patient Health Engagement model (PHE model): a psychosocial theory to understand people's meaningful engagement in their own health care. Patient Prefer Adherence 2018;12:1261-1271 [FREE Full text] [doi: 10.2147/PPA.S145646] [Medline: 30050288]
30. Schaeffer C. Talk to Me. In: Oncology Issues. London: Taylor & Francis; Aug 10, 2017:45-52.
31. Tsoli S, Sutton S, Kassavou A. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. BMJ Open 2018 Feb 24;8(2):e018974 [FREE Full text] [doi: 10.1136/bmjopen-2017-018974] [Medline: 29478016]
32. Kraft MR, Androwich I. Interactive voice response technology: a tool for improving healthcare. NI 2012 (2012) 2012;2012:224 [FREE Full text] [Medline: 24199090]
33. Cohen-Cline H, Wernli KJ, Bradford SC, Boles-Hall M, Grossman DC. Use of interactive voice response to improve colorectal cancer screening. Med Care 2014 Jun;52(6):496-499. [doi: 10.1097/MLR.0000000000000116] [Medline: 24638119]
34. Franke KH, Krumkamp R, Mohammed A, Sarpong N, Owusu-Dabo E, Brinkel J, et al. A mobile phone based tool to identify symptoms of common childhood diseases in Ghana: development and evaluation of the integrated clinical algorithm in a cross-sectional study. BMC Med Inform Decis Mak 2018 Mar 27;18(1):23 [FREE Full text] [doi: 10.1186/s12911-018-0600-3] [Medline: 29580278]
35. Brinkel J, Dako-Gyeke P, Krämer A, May J, Fobil JN. An investigation of users' attitudes, requirements and willingness to use mobile phone-based interactive voice response systems for seeking healthcare in Ghana: a qualitative study. Public Health 2017 Mar;144:125-133. [doi: 10.1016/j.puhe.2016.11.017] [Medline: 28274374]
36. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. In: MIS Quarterly. Minneapolis and Saint Paul, Minnesota: Management Information Systems Research Center, University of Minnesota; 2003:425-478.
37. Graffigna G, Barello S, Bonanomi A. The role of Patient Health Engagement Model (PHE-model) in affecting patient activation and medication adherence: a structural equation model. PLoS One 2017;12(6):e0179865 [FREE Full text] [doi: 10.1371/journal.pone.0179865] [Medline: 28654686]
38. Glanz K, Rimer B, Viswanath K. Health behavior and health education: theory, research, and practice, 4th ed. Heal Behav Heal Educ Theory, Res Pract. Washington, D.C: American Psychological Association; 2008. URL: https://psycnet.apa.org/record/2008-17146-000 [accessed 2020-12-28]
39. Hou S. Health education: theoretical concepts, effective strategies and core competencies. In: Health Promotion Practice. Cairo: World Health Organization. Regional Office for the Eastern Mediterranean; Jun 19, 2014:619-621.
40. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
41. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
42. Reeder B, Hills RA, Turner AM, Demiris G. Participatory design of an integrated information system design to support public health nurses and nurse managers. Public Health Nurs 2014;31(2):183-192 [FREE Full text] [doi: 10.1111/phn.12081] [Medline: 24117760]
43. Information for cancer patients and the public. Cancer Information - Uganda Cancer Institute. URL: https://www.uci.or.ug/services/cancer-information/ [accessed 2020-12-28]
44. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
45. The use of focus group discussion methodology: insights from two decades of application in conservation. Qualitative methods for eliciting judgements for decision making. London: British Ecological Society; 2018. URL: https://doi.org/10.1111/2041-210X.12860 [accessed 2020-12-28]
46. Sampson H, Johannessen I. Turning on the tap: the benefits of using ‘real-life’ vignettes in qualitative research interviews. In: Qualitative Research. SAGE PublicationsSage UK: London, England; Jan 10, 2019:56-72.
47. Comprehensive cancer information. National Cancer Institute. URL: https://www.cancer.gov/ [accessed 2020-12-28]
48. Braun A. Biosafety in handling gene transfer vectors. Curr Protoc Hum Genet 2006 Aug;Chapter 12:Unit 12.1. [doi: 10.1002/0471142905.hg1201s50] [Medline: 18428388]
49. What is RQDA and what are its features? RQDA Project. URL: https://rqda.r-forge.r-project.org/ [accessed 2020-12-28]
50. Freedom to communicate. freePBX. URL: https://www.freepbx.org/ [accessed 2020-12-28]
51. HPE ProLiant DL380 Gen10 Server - Overview. Hewlett Packard Enterprise Support Center. URL: https://support.hpe.com/hpesc/public/docDisplay?docId=a00019682en_us&docLocale=en_US [accessed 2020-12-28]
52. SETU VTEP. Fixed VoIP to T1/E1 PRI Gateway. URL: http://www.matrixepabx.in/detail/8/setu_vtep.html [accessed 2020-12-28] [WebCite Cache ID http://www.matrixepabx.in/detail/8/setu_vtep.html]

53. Besacier L, Barnard E, Karpov A, Schultz T. Automatic speech recognition for under-resourced languages: a survey. In: Speech Communication. Amsterdam: Elsevier B.V; 2014:85-100.
54. Fadol AP, Mendoza TR, Lenihan DJ, Berry DL. Addressing the symptom management gap in patients with cancer and heart failure using the interactive voice response system: a pilot study. J Adv Pract Oncol 2018 Mar;9(2):201-214 [FREE Full text] [Medline: 30588354]
55. Brinkel J, May J, Krumkamp R, Lamshöft M, Kreuels B, Owusu-Dabo E, et al. Mobile phone-based interactive voice response as a tool for improving access to healthcare in remote areas in Ghana - an evaluation of user experiences. Trop Med Int Health 2017 May;22(5):622-630 [FREE Full text] [doi: 10.1111/tmi.12864] [Medline: 28278352]
56. Janevic MR, Aruquipa Yujra AC, Marinec N, Aguilar J, Aikens JE, Tarrazona R, et al. Feasibility of an interactive voice response system for monitoring depressive symptoms in a lower-middle income Latin American country. Int J Ment Health Syst 2016;10:59 [FREE Full text] [doi: 10.1186/s13033-016-0093-3] [Medline: 27688798]
57. Greaney ML, De Jesus M, Sprunck-Harrild KM, Tellez T, Bastani R, Battaglia TA, et al. Designing audience-centered interactive voice response messages to promote cancer screenings among low-income Latinas. Prev Chronic Dis 2014 Mar 13;11:E40 [FREE Full text] [doi: 10.5888/pcd11.130213] [Medline: 24625364]
58. Odeny TA, Newman M, Bukusi EA, McClelland RS, Cohen CR, Camlin CS. Developing content for a mHealth intervention to promote postpartum retention in prevention of mother-to-child HIV transmission programs and early infant diagnosis of HIV: a qualitative study. PLoS One 2014;9(9):e106383 [FREE Full text] [doi: 10.1371/journal.pone.0106383] [Medline: 25181408]

Abbreviations

DTMF: dual-tone multi-frequency
FGD: focus group discussion
HBM: health belief model
ISDN: Integrated Services Digital Network
IVR: interactive voice response
LMIC: low- and middle-income country
mHealth: mobile health
PBX: private branch exchange
UCI: Uganda Cancer Institute
UNCST: Uganda National Council for Science and Technology
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 02.07.20; peer-reviewed by S Six, M Adly, A Adly, A Adly, E Oghogho; comments to author 08.10.20; revised version received 11.11.20; accepted 15.11.20; published 26.01.21.

Please cite as:
Kabukye JK, Ilozumba O, Broerse JEW, de Keizer N, Cornet R
Implementation of an Interactive Voice Response System for Cancer Awareness in Uganda: Mixed Methods Study
JMIR Mhealth Uhealth 2021;9(1):e22061
URL: http://mhealth.jmir.org/2021/1/e22061/
doi:10.2196/22061
PMID:33496672

Original Paper

Configuration of Mobile Learning Tools to Support Basic Physical Assessment in Nursing Education: Longitudinal Participatory Design Approach

H Ösp Egilsdottir[1], MSN; Lena Günterberg Heyn[1], PhD; Espen Andreas Brembo[1], PhD; Kirsten Røland Byermoen[1], MSN; Anne Moen[2], Dr Polit; Hilde Eide[1], Dr Philos

[1]Science Centre Health and Technology, Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway
[2]Institute for Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
H Ösp Egilsdottir, MSN
Science Centre Health and Technology
Faculty of Health and Social Sciences
University of South-Eastern Norway
Grønland 58
Drammen, 3045
Norway
Phone: 47 94886406
Email: osp.egilsdottir@usn.no

Abstract

Background: As many students in higher education are skilled users of mobile technology, mobile learning (mLearning) can be a promising educational strategy to enhance their learning experience. mLearning might also be well suited for nursing students as they navigate between multiple learning contexts in their educational curriculum. As an educational strategy, mLearning may also reduce challenges caused by the theory-practice gap in nursing by supporting skills and knowledge transfer between the university and clinical settings. As the introduction of basic physical assessment skills (B-PASs) into Norwegian bachelor's degree education in nursing occurred quite recently, there is a lack of competence in supervision and teaching in both university and clinical settings. As such, mLearning appears to be a good strategy to support student B-PAS learning and knowledge transfer across learning contexts.

Objective: This study aims to explore and elicit the perspectives of students regarding the way in which a selection of digital learning resources supports B-PAS learning and application in clinical rotation, which of the selected digital learning resources are beneficial to include in a suite of mLearning tools, and how the selected digital learning resources could support the transfer of skills and knowledge from the academic to clinical context.

Methods: We used a longitudinal participatory design approach to co-design a suite of mLearning tools. The co-design processes took place in several workshops (WSs) over a period of 3 months: 2 WSs with first-year students (n=6), 3 WSs with second-year students (n=6), and 3 WSs with third-year students (n=8). The students evaluated several digital learning resources in both academic and clinical contexts. The digital learning resources included digital simulation with virtual patients, massive open online courses, and multimedia learning material. In the co-design WS, the potential and benefits of these digital learning resources for the learning and application of B-PASs were explored.

Results: The students reported that the digital learning resources stimulated learning in 7 different ways. They also emphasized the importance of including all selected and tested digital learning resources. Moreover, students supported the inclusion of additional learning material, such as multiple-choice tests and written assignments, aimed at providing feedback and contributing to knowledge development.

Conclusions: The co-design processes and collaboration with the nursing students provided insight into how a suite of mLearning tools may support the learning and application of B-PASs and human bioscience knowledge in clinical rotation. From the students' perspective, one of the strengths of the suite of mLearning tools was the range of content, as this met a broader range of student learning preferences regarding learning B-PASs. The suite of mLearning tools contributes to and supports skills training and knowledge transfer between multiple learning contexts.

(JMIR Mhealth Uhealth 2021;9(1):e22633) doi:10.2196/22633

KEYWORDS

learning; mobile phone; mobile application; education, nursing; students, nursing; education, clinical; nursing skills; physical examination; computer simulation; clinical competence

Introduction

Mobile Learning Opportunities in Higher Education

Students pursuing higher education today expect teaching methods that stimulate self-directed learning, active learning, peer learning, and the cocreation of knowledge [1,2]. Many of these students belong to the *digital natives* generation having grown up with mobile technology such as smartphones; this makes them skilled users, especially with regard to social interaction and finding information on the web [2,3]. Mobile technology used for mediating educational content offers flexibility in teaching methods; thus, mobile learning (mLearning) promotes and contributes to learning processes that are less constrained by time and context [4-6]. Research indicates that mLearning contributes to meaningful and comprehensive learning experiences, inviting students to select their preferred modalities and share the responsibility for their own learning processes [1,7,8].

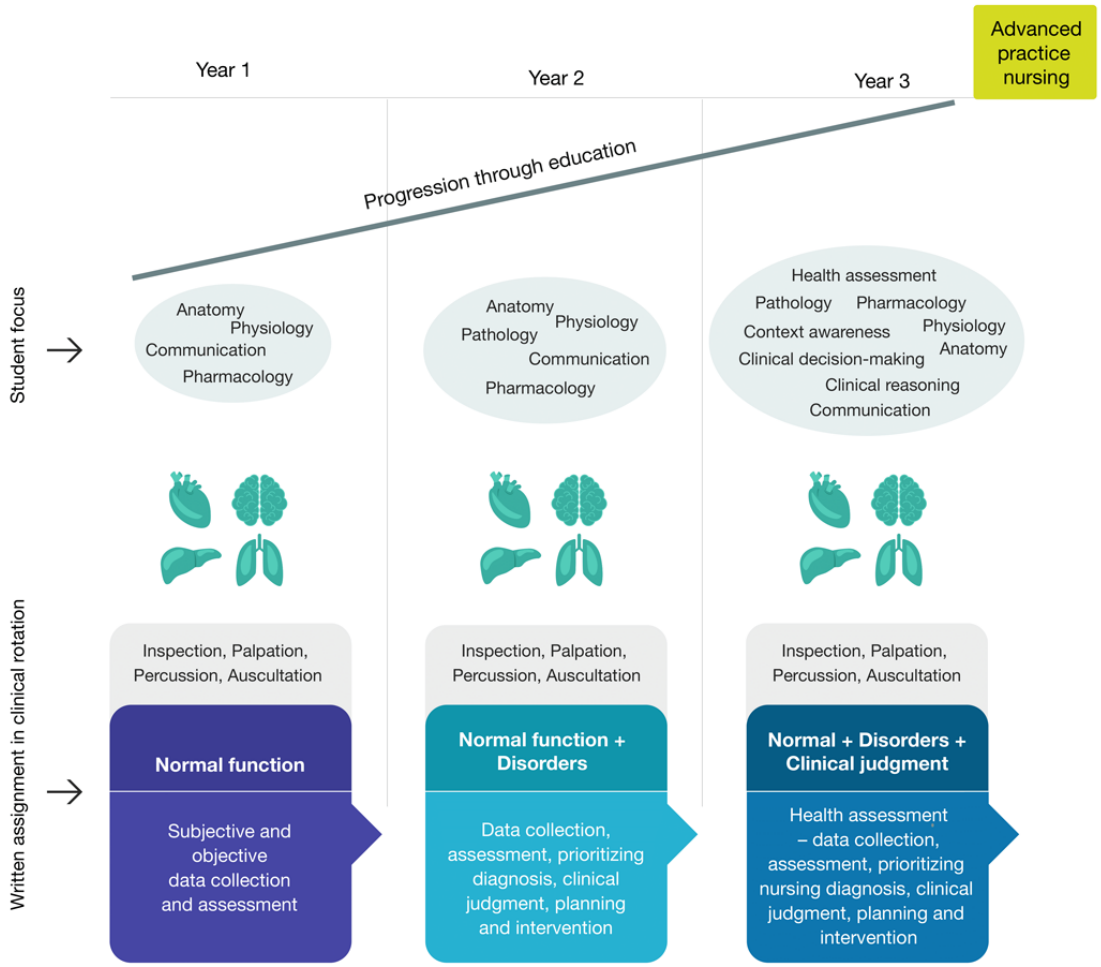
In nursing education, students navigate between multiple learning contexts during lectures and different clinical rotation periods. This requires students to handle and mitigate the differences between what is taught in universities and what they learn in the clinical rotation—this is commonly referred to as the theory-practice gap [9]. mLearning is shown to be a good strategy for reducing these challenges by supporting students in becoming less dependent on the learning context and increasingly self-directed in their own learning processes [10,11]. When considering the implementation of a suite of mLearning tools in nursing education, students should be core collaboration partners because of their role as end users [11,12]. Importantly, as nursing students are the ones who must navigate between multiple learning contexts, they should be the ones evaluating and testing mLearning and its possible content to determine what works for them and in what context [11]. As such, a participatory design involving co-design workshops (WSs) is a good method to highlight end user involvement and collaboration in the development and evaluation of the end product [12].

Rationale for Educational Focus in a Suite of mLearning Tools

Nurses base their clinical judgment on physical assessment and history taking to map their patients' health condition [13]. Physical assessment skills (PASs) are therefore one of the core competencies nurses must master [14]. Inadequate patient assessment might result in a failure to notice deteriorating patients and to initiate appropriate nursing interventions; this can threaten the safety of patients and result in adverse health outcomes [15,16]. However, performing adequate patient assessments can be challenging for novice nurses because of the complexity of health situations in the different patient groups that they encounter [16]. The nursing students learn PAS in preclinical courses at the university but apply these skills in clinical rotation. Therefore, learning and developing competence and confidence in performing PAS and the ability to articulate relevant knowledge are based on how these skills are taught on campus and on the supervision of preceptors during the clinical rotation [16,17].

Until recently, PASs were not a part of nursing education in Norway. To overcome the barriers identified in international studies [14], our university implemented selected PASs focusing on respiratory, peripheral circulation, abdominal, and neurological assessments in the curriculum in 2015. These PASs are considered to be the basic skills necessary for clinical competence for undergraduate nursing students and are, therefore, referred to as *Basic* PAS (B-PAS; Figure 1) [17].

Figure 1. Basic physical assessment skills progression model in Norwegian nursing education.



In a previous study, we identified that a lack of knowledge and practical use of B-PASs among the faculty and preceptors limited students in their learning of and performing these skills [17]. This points to a theory-practice gap with regard to developing these specific skills and underlying knowledge. This is concerning, as 50% of Norwegian nursing education takes place in clinical rotations in different contexts [18]. An additional result from this earlier study was that students identified mLearning as potentially enhancing their B-PAS learning processes. This finding is supported by other studies [19,20]. Research also highlights that mLearning can contribute to the articulation and integration of human bioscience knowledge (anatomy, physiology, pathophysiology, and pharmacology), which underpins knowledge-based basic physical assessment.

Thus, mLearning can support nursing students in learning and applying B-PASs in a clinical context without being solely dependent on guidance from their supervisors. mLearning may also contribute to better transfer of knowledge between multiple learning contexts, and it arguably works best in combination with other teaching methods and educational strategies [21-24].

There is a range of digital learning resources that are suitable for inclusion in a suite of mLearning tools highlighting B-PASs and relevant knowledge, for example, massive open online courses (MOOCs) [25], web-based apps, gaming [22], podcasts [26], multimedia presentations [27], and more comprehensive software involving immersive technology, such as digital simulation with virtual patients [28,29]. Digital simulation involves real people using digital and mobile technology to operate interactive virtual patient scenarios for educational purposes [30]. When reviewing the literature regarding factors facilitating or hindering the implementation of mLearning in medical and nursing education, Lall et al [31] not only identified high student satisfaction with mLearning but also revealed challenges related to the implementation process. These challenges existed both in universities and clinical settings and were related to faculty competence in using mobile technology, the acceptance of using mobile devices in different clinical settings, and students' access to the devices [31].

Therefore, the aim of this study is to co-design a suite of mLearning tools with nursing students to support their B-PAS learning and application. The specific objectives are to explore the following with students: (1) in what way does the selection of digital learning resources support B-PAS learning and application in clinical rotation; (2) which of the selected digital learning resources are beneficial when included in the suite of

mLearning tools; and (3) how the selected digital learning resources could support the transfer of knowledge from academic context to clinical settings.

Methods

Research Design

This qualitative longitudinal study was inspired by participatory design. Participatory design aims to involve end users in co-design and development processes with a specific purpose [12]. The co-design processes took place in several iterative WSs over a period of 3 months: 2 WSs with first-year students (n=6), 3 WSs with second-year students (n=6), and 3 WSs with third-year students (n=8). Through these WSs, the experiences and preferences of students were explored with the aim of informing the final selection of digital learning resources included in the suite of mLearning tools.

Recruitment and Sample

The nursing students were recruited from a bachelor's program in nursing at a large university in Norway. Information about the aim of this study and an invitation to participate were published on the learning management system (LMS) of the university, and a short presentation about this study was given orally on campus. Students signed up to participate by emailing the first author (ÖE) directly or via their course leader. The convenience sample consisted of 20 nursing students (14 women and 6 men) in the first, second, and third year of the program. The participants' ages varied from 20 to 50 years.

Procedure and Data Collection

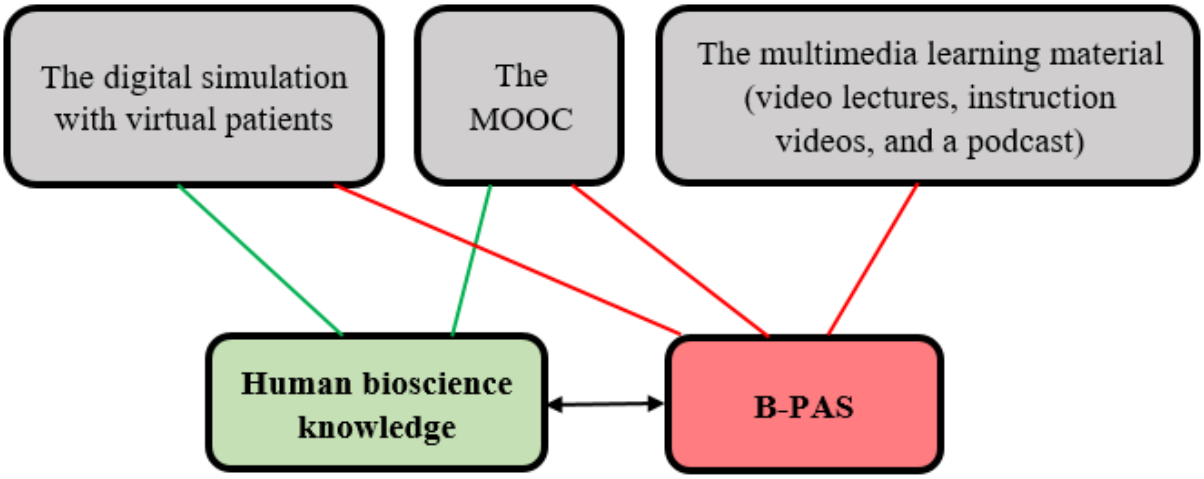
All the WSs were held in 2019 during the spring semester on the university campus. They were facilitated by the first author (ÖE), who was supported by one of the other researchers (EB, LH, or HE) during each WS. ÖE was also an active participant in the discussions. The main tasks for the facilitators were to involve all the students in the discussions, ensure that all planned topics were covered, and keep track of time and breaks. Each WS lasted a maximum of 2 hours, and the discussions were audio recorded. WSs 1 and 2 were transcribed verbatim, whereas WS 3 was worked with as audio files in which important statements were marked.

The Digital Learning Resources Suggested as Content in the Suite of mLearning Tools

A selection of digital learning resources highlighting B-PASs in different ways were introduced to and assessed by the participating nursing students. The digital learning resources included (1) a digital simulation program with virtual patients, (2) a MOOC, and (3) multimedia learning material (video lectures, instruction videos, and a podcast; Figure 2). These digital learning resources were new to most of the students.

The digital simulation program consists of several interactive virtual patient cases and can be used in groups or as an individual learning activity. One of the tasks in the digital simulation is to choose appropriate B-PASs and other interventions during assessment of virtual patients; the students also received feedback on their choice of assessments and the interventions performed in the program—fitting perfectly with the aim of the suite of mLearning tools. The MOOC is an individual e-learning course with 5 modules promoting students' assessment of vital signs and history-taking skills, notice of patient deterioration processes, and learning about more advanced skills (eg, auscultation of the lungs and heart). In addition, it was desirable to evaluate the selection of multimedia material, consisting of video lectures, instruction videos, and a podcast, all aiming to teach (verbally and visually) how to perform B-PASs on real patients. The podcast focused on giving the students perspectives on how to learn and focus on using these skills throughout their nursing education.

Figure 2. Learning focus for the different digital learning resources tested by the students. B-PAS: basic physical assessment skills; MOOC: massive open online course.



Content and Structure of Co-Design WSs

The researchers planned and framed the structure and content of the 3 WSs, which were generally based on the selected digital learning resources. Students engaged in the co-design processes by using and evaluating how the digital learning resources influenced their learning related to B-PASs and human bioscience knowledge during clinical rotation and in academic

courses. Figure 3 provides an overview of the processes in the co-design WSs.

The first WS included the introduction and discussion of 2 new learning resources: the digital simulation with virtual patients and the MOOC. We asked the students to evaluate and test these 2 digital learning resources individually before attending the second WS. In the second WS, we explored the experiences of students with using digital simulation and the MOOC. As B-PASs is built on knowledge of human biosciences, it was important to discuss with the students as to what type of learning material or assignments could facilitate coherence between theoretical knowledge of human biosciences and the application of these skills. The nursing students were asked to use and evaluate the multimedia learning materials (videos and podcast) before attending the next WS. In the third WS, we focused on discussing the students' feedback regarding the multimedia learning materials. Here, it was important to discuss the digital learning resources to be included and the ones to be excluded. In addition, we presented a pilot version of the suite of tools to give the students a visual example of how they might look in the LMS. This way, the students had the opportunity to give feedback and input regarding the first prototype. The participation of nursing students was relatively stable throughout the WSs; however, some students participated in only 1 WS (Table 1).

Figure 3. Co-design processes in the workshops. mLearning: mobile learning; MOOC: massive open online course.

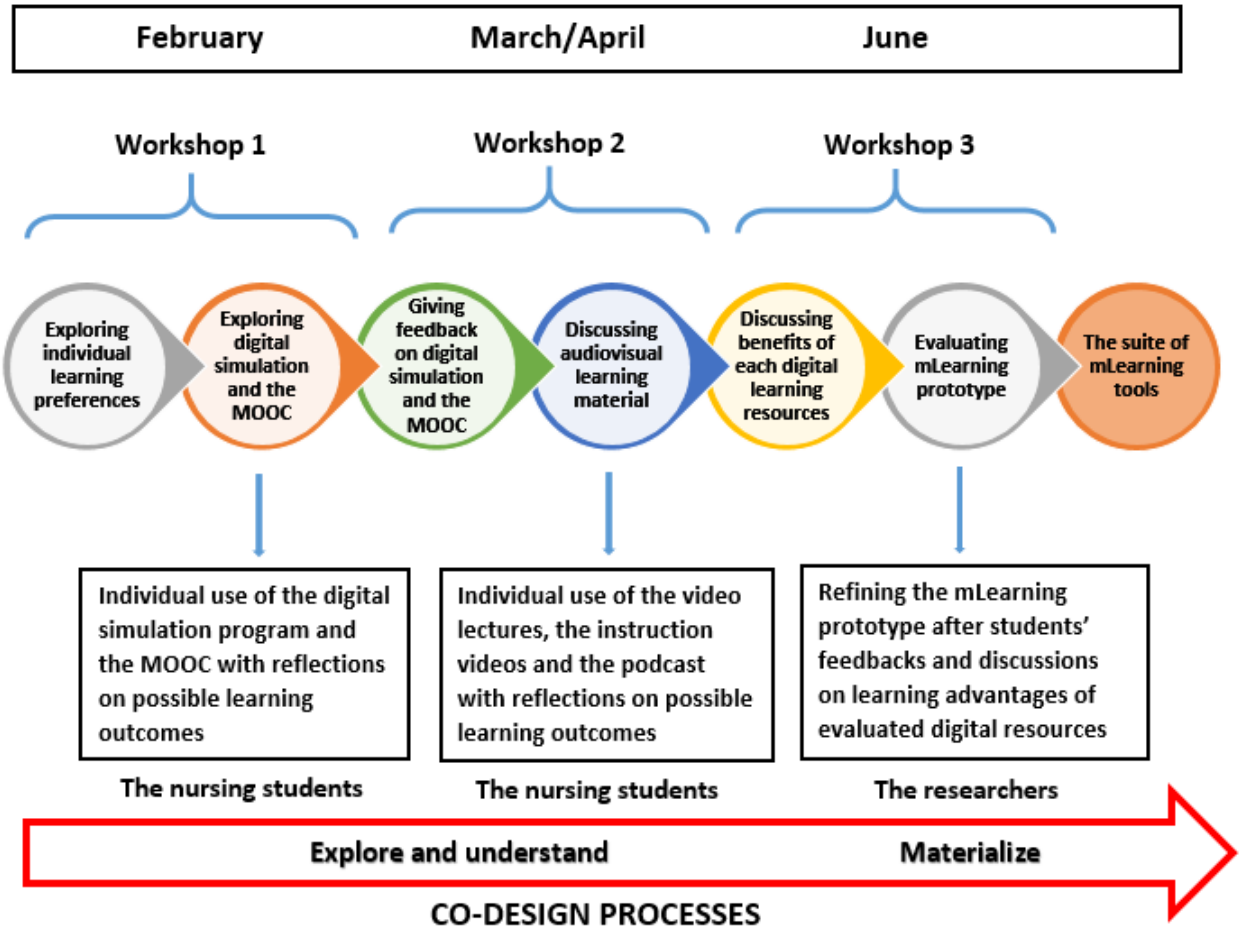


Table 1. Participation of nursing students in the workshops.

Student	WS[a] 1	WS 2	WS 3
First-year students			
Student 1	✓[b]	—[c]	—
Student 2	✓	—	—
Student 3	✓	✓	—
Student 4	✓	—	—
Student 5	✓	—	—
Student 6	—	✓	—
Second-year students			
Student 7	✓	✓	✓
Student 8	✓	✓	✓
Student 9	✓	✓	✓
Student 10	✓	—	—
Student 11	✓	✓	✓
Student 12	—	✓	✓
Third-year students			
Student 13	✓	✓	—
Student 14	✓	✓	✓
Student 15	✓	✓	✓
Student 16	✓	✓	✓
Student 17	✓	✓	✓
Student 18	✓	✓	—
Student 19	✓	✓	✓
Student 20	—	✓	—

[a]WS: workshop.

[b]Attending the workshop.

[c]Not attending the workshop.

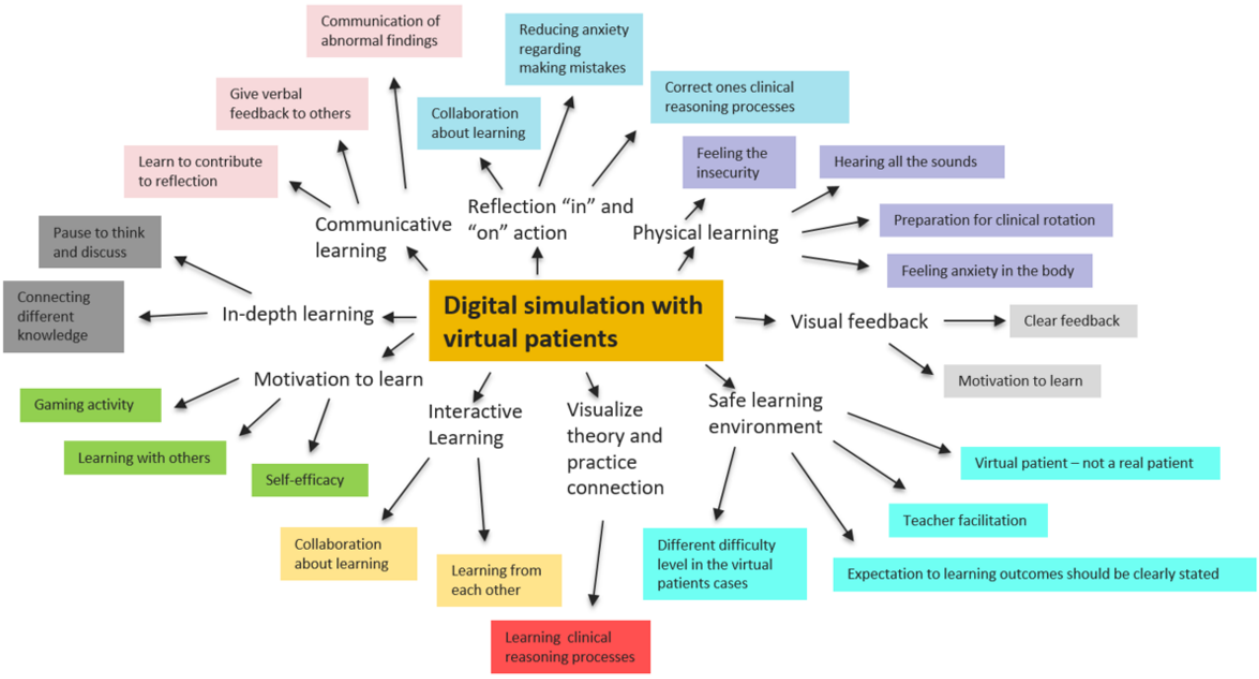
Research Ethics

The Norwegian Centre for Research Data approved the study (Reference number 462735). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. Reviewing the learning resources was not a requirement to attend WSs 2 and 3. The researchers were not involved in evaluating the academic or clinical courses of the participants, and participation in this study had no impact on grading in any of the courses.

Data Analyses

Thematic content analysis was used to analyze data [32], which comprised text materials (WSs 1 and 2) and audio files (WS 3). The research questions guided this process. The text material from WSs 1 and 2 (all student cohorts) and the audio files from WS 3 (second- and third-year students) were read and listened to completely, respectively. Further analysis of the data included important concepts and statements involving the design of and suggested content in the suite of mLearning tools. Figure 4 shows an example of how the digital resources (in this case, the digital simulation with virtual patients) were creatively mapped to explore core concepts throughout the analyses of the data.

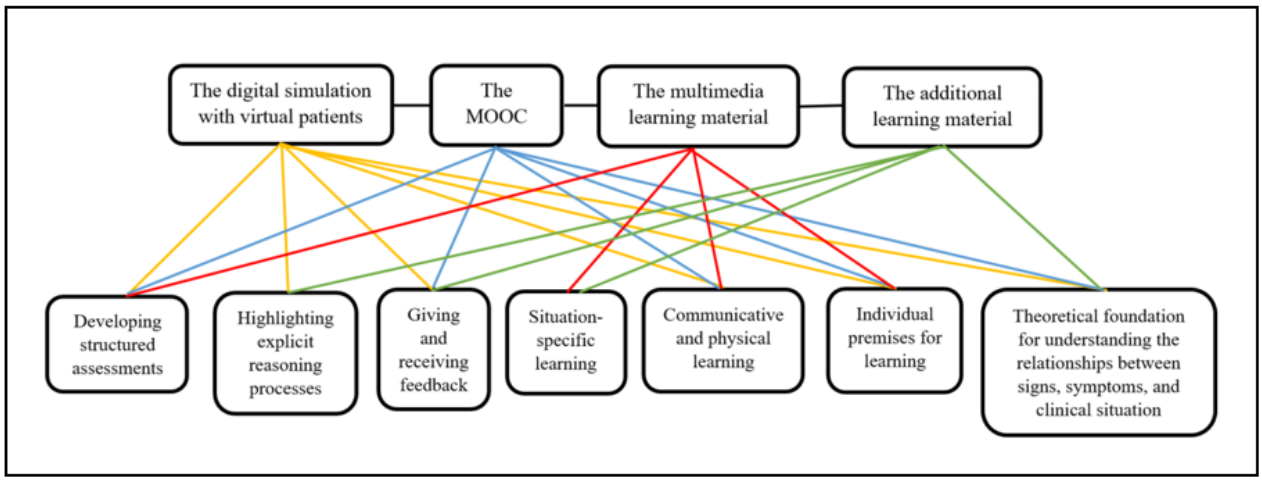
Figure 4. Example of a concept map in the analysis process.



Results

According to the students, the digital learning resources stimulated learning in 7 different ways: (1) developing structured assessment, (2) highlighting explicit reasoning processes, (3) giving and receiving feedback, (4) situation-specific learning, (5) communicative and physical learning, (6) individual premises for learning, and (7) the theoretical foundation for understanding the relationships between signs, symptoms, and clinical situation. Figure 5 shows how the digital learning resources contributed to student learning, both individually and collectively.

Figure 5. Contribution of digital learning resources to student learning. MOOC: massive open online course.



In the following sections, each digital learning resource and its influence on the learning process are presented. Students' own voices, in the form of quotations, are included in Textbox 1 to substantiate the data analyses. The content of the suite of mLearning tools is then presented.

Textbox 1. Students' quotations related to the evaluation and testing of digital learning resources; these quotes substantiate concept mapping and data analysis.

- The digital simulation with virtual patients:
  - Student 13 workshop (WS) 1: "I believe that this (digital simulation) will be very attractive for the students, it's a really cool way to learn and the students might be more motivated to learn anatomy and physiology, rather than sitting for six to seven hours listening to a lecture, which is really exhausting"
  - Student 17 WS 2: "There is always something that goes wrong. Then you have the opportunity to pause the simulation and review what you missed or what you did wrong. Luckily it is a fictive patient but you learn in what order you should do things"
  - Student 8 WS 1: "We get the opportunity to meet someone really ill in a totally safe environment"
- The Massive Open Online Course
  - Student 15 WS 3: "If the solution of the case is a stroke diagnosis and you didn't think of that, then there are obviously some really important cues that you missed"
- Multimedia learning materials
  - Student 14 WS 3: "It is really important to show the reasoning processes—it's not important what the conclusion is but just to show the reasoning processes"
  - Student 11 WS 2: "It would be perfect if lectures were available online. If you don't understand certain things and you are afraid to ask in the classroom, then you can just rewind and look at the material that is difficult to understand. You can also have the textbook beside you to try to figure it out"
  - Student 2 WS 2: "In order to recognize potential symptoms of a disease, you must know the normal body functions. Then you must know what to listen for—you cannot just listen to the lungs without knowing. It requires fundamental theoretical knowledge"
- Additional learning materials
  - Student 7 WS 3: "I would rather use my time and effort on something that gives me feedback on or an indication of my knowledge base, than on academic assignments in clinical rotation where I never get any feedback at all"
  - Student 15 WS 3: "There is something about being conscious regarding one's own feedback to other people. To be conscious about what word you use and how to give constructive feedback—and we will have to do that when we become supervisors for nursing students later"
- General design features for better end user experience
  - Student 11 WS 1: "A multimedia learning resource is an extremely effective learning tool within some courses—it is all about experimenting and finding out what works. But a well-designed digital learning resource can help to show how different knowledge is intertwined"

Digital Simulation With Virtual Patients

In general, all the students were excited about the digital simulation and the possibilities that this learning resource could offer for skills and knowledge development. The first-year students discussed how digital simulation could enhance their learning of human bioscience knowledge and the ability to connect it with other topics in nursing education. The second- and third-year students highlighted its usefulness related to the development of communication and clinical reasoning skills, as they learned to prioritize nursing interventions and focus on performing B-PASs. The nursing students also elaborated on their physical reactions in the situation. The sounds from the patient (eg, moaning in pain) and instruments (eg, changes in vital signs) made the students feel stressed, as they would have in a real patient situation. With this authenticity, the students felt that the digital simulation with virtual patients would contribute to physical and mentally preparedness for clinical rotation. The use of English in the digital simulation software created some challenges for most of the students; as such, they highlighted the value of working in their native language with regard to improving learning outcomes and the overall user experience.

After testing the digital simulation between WSs 1 and 2, the students emphasized the benefits of working together in a group rather than working with the virtual patient cases individually. The students also described how faculty was important when working in a group in the digital simulation sessions, with regard to promoting a safe learning environment, facilitating knowledge development, and collaborating within the group. The safe learning environment was viewed as a prerequisite for students' reflections *in* action and *on* action, influencing each other's learning processes by making it *safe* to actively participate in a discussion with one's peers.

Another important factor in the digital simulation was the opportunity to learn from making incorrect or suboptimal decisions without harming a real patient during the simulation. To enhance this learning outcome, the process of reflection during the simulation session and in the debriefing phase were crucial to understanding what went wrong. Through the digital simulation, the students also learned the importance of performing structured physical assessments; this included learning where to start gathering data for a thorough overview of the patient situation to initiate nursing interventions.

The nursing students also stressed that the digital simulation with virtual patients used in combination with theoretical classes and during clinical rotation periods might be a beneficial pedagogical strategy to enhance learning, especially in theoretical classes involving human bioscience knowledge. Some of the students in all 3 cohorts drew parallels to gaming experiences and discussed ways these may be helpful in the use of the digital simulation.

The MOOC

The structure of the 5 modules included in the MOOC appealed to the students. They found that elder care or community health care was the primary focus and suggested that the MOOC was best suited to the curriculum in the first and third years (in the Norwegian context). As some of the assignments were built on one specific clinical case, the students felt that a broader selection of cases would provide better knowledge development. The second- and third-year students especially liked how communication skills were emphasized in the MOOC. In their opinion, focusing on communication skills is as important as focusing on technical skills during clinical rotation. The nature of the assignments, such as listening to audio files, appealed to the second- and third-year students. Some of the assignments also required that they categorize patient data into objective or subjective data. In this manner, the MOOC supported students' learning regarding the value of working in a structured way when prioritizing data collection in clinical cases. Moreover, the clinical cases supported the cognitive processes involved in *catching the cues*, and for students, this was an important learning outcome of the MOOC.

Multimedia Learning Materials

The multimedia materials included video lectures, instruction videos, and a podcast. The students highlighted that the use of these learning materials supported different clinical situations, enabling them to selectively choose specific learning material before performing B-PASs with the patient. According to the students, this type of learning material can contribute to both knowledge development and self-efficacy. They found it helpful to have the learning outcomes clearly stated regarding the expected level of B-PAS performance. Students also valued the podcast format, which could be used to discuss clinical cases showing experienced nurses' clinical reasoning processes in action.

The students valued the functional possibilities offered by the multimedia materials, for example, they could pause the audio files to familiarize themselves with difficult concepts and then replay the audio files as often as necessary. Here, the auscultation skills were highlighted as especially challenging, particularly relating to the interpretation of lung and heart sounds. The students also suggested that the features in the multimedia materials could be simple animations, not necessarily actual patients, as the students have access to real patients during clinical rotation. Although they thought the use of animations instead of actual patients might contribute to a better understanding of the connection between human biosciences and the clinical situation, the students highlighted that the learning materials must be of high (sound and picture) quality.

Additional Learning Materials

The students found it helpful to include additional learning materials aimed at supporting knowledge transfer in and development of the suite of mLearning tools. This additional content consisted of tests with multiple-choice questions (MCQs), several clinical cases to choose among, and written assignments. The students also emphasized the inclusion of clearly stated learning outcomes related to this additional material, making it easier to understand the value of engaging with the learning material. They felt this could also contribute to knowledge development and self-efficacy. The students wanted access to the correct answers for the MCQ tests and the possible solutions for the clinical cases: they argued that this type of immediate feedback could help them understand their own level of performance and identify knowledge gaps. The students also suggested that it was important, for purposes of motivation and self-efficacy, to receive the results or feedback visually—perhaps with audio animation or a pop-up effect.

The students gave suggestions regarding how to structure this part of the suite of mLearning tools in a way that would trigger their curiosity and motivation to learn more. One suggestion was that correct answers in MCQ tests or clinical cases could *unlock* further advanced learning material. The students welcomed a written assignment targeting reflection on using the B-PAS during clinical rotation, preferably with a peer review; they felt that this would also provide them with experience in providing critical and constructive feedback. They did not want open peer review but preferred anonymity because of the different relationships between students and their own insecurities regarding giving peer feedback.

General Design Features for Better End User Experience

One important element of the co-design processes concerns the general design features that emerged from the WS discussions. The students emphasized that the suite of mLearning tools had to be easily accessible, have a logic structure, be compatible with smartphones, and be usable *on the go*. Students highlighted the importance and possibilities of linking the suite of mLearning tools to other relevant web-based resources. The second- and third-year students recommended limited access for first-year students to avoid overwhelming them with too much information. One solution that they suggested was to design a *lock* in the structure of the suite of mLearning tools that could be unlocked by the most motivated first-year students who are eager to learn more. It was also important for the students to have the possibility of being anonymous or to use avatars if the use of the tool was visible to other students.

The Suite of mLearning Tools Recommended From the Co-Design WSs

The co-design processes included all selected digital learning resources in the suite of mLearning tools aiming to support the B-PAS and knowledge development. The LMS of the university was used to structure the suite of mLearning tools, which enabled the students to access digital learning resources by using mobile devices. Table 2 provides an overview of the content available to the students and a short description of the different

digital resources. The content in the LMS had no fixed order, enabling students to access the content in which they were specifically interested. The co-design processes and the longitudinal research design allowed students to test and evaluate the selected digital resources when in different learning contexts, for example, in an academic context and in a clinical rotation context. In such cases, the student expressed that this suite of mLearning tools and the future use of mobile technology may have teaching and learning value in both theoretical and practical courses—a valuable educational component during challenging situations such as the current COVID-19 pandemic.

There was a small difference in what was planned for the first-year students to have access to. The written assignment was included for all years, but the element of peer review was excluded for the first-year students based on an assessment made by the older students and the research group. As none of the first-year students were able to attend the final WS, there was no opportunity to discuss with them the potential advantages or disadvantages of peer review. According to the second- and third-year students, the main strength of the suite of mLearning tools was that it offered different types of digital learning resources from which to choose. Thus, these students found it useful to include all of the selected and evaluated digital learning resources in the suite of mLearning tools.

Table 2. Overview of the structure and content of the suite of mobile learning tools in the learning management system.

Different digital learning resources	Elaboration of the content
	A short description of the digital simulation program, information about the sessions, and the virtual patient cases on which students can work. Also included are log-in details and information about whom to contact if there are problems with the log-in
	A brief description of the MOOC[a] and recommendation for which modules students should focus on, depending on which educational year they belong to. The students also find details here regarding the log-in process
	Detailed information about auscultation skills, divided into 2 sections: (1) lung sounds and (2) heart sounds. Each section contains links to YouTube videos and audio files with different sounds to which students may listen
	A total of 5 instruction videos with a nurse performing B-PASs[b] on a patient. Each video has a specific focus: the heart and peripheral circulation, the respiratory system, the abdomen, the neurological assessment, and recording vital signs. The duration of the videos is from 7:37 to 15:51 min
	A total of 4 video lectures in which each video has a specific focus: the heart and peripheral circulation, the respiratory system, the abdomen, and the neurological assessment. The duration of the videos is from 13:13 to 32:55 min
	• Two nurses (faculty members) talk about the origin of the physical assessment in nursing education and the differences between performing B-PASs as a RN[c] or as a nurse practitioner • A conversation between the faculty members and 2 newly graduated RNs, focused on working with B-PASs throughout the 3-year nursing program, and how they work with B-PASs as new RNs
	Brief information about how to structure professional communication about data gathered by mapping the patient health condition through the use of different communication tools and how to use these tools during clinical rotation
	Information about how to structure the professional documentation of the information gathered through mapping the patient health condition by using B-PASs in the different documentation systems used in clinical practice
	MCQ[d] aiming to support students' knowledge, to repeat and refresh bioscience knowledge, and to identify knowledge gaps
	Description of a written assignment targeting reflection on the use of B-PASs. Feedback is given by fellow students (a peer review for the second- and third-year students) in which all parties are anonymous
	Checklists summarizing the elements of every focus in B-PASs (eg, respiratory system and neurological assessment). Can also be used when students use B-PASs in clinical rotation

[a]MOOC: massive open online course.

[b]B-PAS: basic physical assessment skill.

[c]RN: registered nurse.

[d]MCQ: multiple-choice question.

Discussion

Principal Findings

This study explored the use of the participatory design method to co-design a suite of mLearning tools specifically aimed at supporting the learning and performance of B-PASs and integration of human bioscience knowledge. As such, this study expands the use of mLearning with its explicit focus on the learning and application of practical skills, thus contributing to a better understanding of how different digital learning resources, individually and collectively, influence students' learning and application of skills in nursing education.

The Contribution of Digital Learning Resources to Enhance Learning of Skills and Knowledge

The most important findings of this study are regarding how the digital learning resources in the suite of mLearning tools enhanced learning processes related to specific categories of skills and knowledge. The digital learning resources stimulated the learning and application of B-PASs from a multidimensional perspective: from physical (motor) learning to the articulation of integrated theoretical knowledge and the enhancement of context-specific learning processes (Figure 5). Access to the suite of mLearning tools appears to have impacted the development of self-efficacy related to B-PASs in nursing students. As such, the students might perform B-PASs more frequently, and perhaps better, while also applying human bioscience knowledge in clinical reasoning processes. This is in line with other studies showing that students value access to different kinds of mobile resources to support and enhance learning [6,8,30,33-35]. The results of this study are also interesting, as although nursing students already use a variety of mobile apps—for example, drug calculators, medical dictionaries, handbooks, and clinical skills guides [6,8,33]—our suite of mLearning tools still seems to have added value. Nevertheless, there is a need for further exploration of how mLearning should be implemented across the nursing curriculum; this could be done by critically exploring which learning processes mLearning aims to support. Although mLearning alone might not be the best educational strategy in nursing education, in combination with other activities, it might represent a good educational approach [17,23].

Digital simulation with virtual patients was the preferred learning resource by students. The students not only valued the possibility of a new and innovative learning activity that the digital simulation offered but also highlighted that the involvement of faculty was crucial to facilitate good learning processes in this context. This facilitation role also involved creating a safe environment that allowed reflection *in* and *on* actions during the simulation session. The role of facilitation in digital simulation with a focus on B-PASs, both for the teacher and the preceptor, needs to be further developed. Although digital simulation as a part of the suite of mLearning tools is an individual learning activity, it also seems to be important to plan the digital simulation as a group activity to increase the learning outcome. As we have shown earlier and, in this study, because of a lack of adequate supervision with regard to B-PASs [17], the suite of mLearning tools plays an important role in supporting the learning and development of these skills and knowledge. Research shows that digital simulation alone can strengthen self-efficacy, performance of clinical skills, decision making [28], clinical reasoning processes [35,36], and nontechnical skills [37] of students. This indicates that including digital simulation with virtual patients contributes to better learning of skills, and that reflection contributes to the articulation of human bioscience knowledge, promoting clinical reasoning skills.

Supporting Knowledge Transfer Between Different Learning Contexts for Skills Training

Another interesting finding of this study is how the co-design processes and the WS discussions revealed in what way digital learning resources can support knowledge transfer between different learning contexts. The suite of mLearning tools seemed to contribute to more seamless learning processes and thereby contributed to bridging the theory-practice gap in nursing, creating a more self-directed learning space for nursing students.

Lewin et al [37] argue that technology changes the boundaries of different types of learning (formal and informal) and learning spaces. This opens new ways to connect and combine different learning sites in higher education. As such, collaboration between universities and clinical settings may promote the creation of new knowledge and learning spaces, in which these institutions are equal partners in influencing learning and competence development in students. Chan et al [38] have termed this *seamless* learning: seamless learning processes enable students to navigate between different spaces and different roles, and to interact with different educational practices—for example, the university, the clinical setting, and the suite of mLearning tools [38]. As such, the suite of mLearning tools can promote transferability between learning contexts in nursing education. Therefore, further development and implementation should occur in close collaboration with students, the university, and the clinical setting.

However, the implementation of new and innovative teaching and learning strategies, such as the suite of mLearning tools, can face obstacles in both university and clinical settings. It is important for these obstacles to be identified and addressed to best support student learning. Attitudes of nursing staff, patients, and patients' families toward the use of mobile devices may hinder their use of these devices [8,30,33,39]. The nursing students in our study reported similar experiences. It was also important for the students to be able to access the learning materials anywhere, at any time. Hsu and Hsiang [6] recommend that mLearning should support offline activity to secure a better end user experience. Taking this into consideration when designing and using a suite of mLearning tools in nursing education might enhance students' experience of mobile technology as a pedagogical approach. This also highlights the importance of continuing to explore how mobile technology can be successfully implemented and support the teaching role of the faculty and the preceptors.

Limitations

This study has several limitations that must be noted. The first-year students were absent in WS 3, which limited the opportunity to discuss with them the structure and the final content of the suite of mLearning tools. This needs to be explored in future studies. In addition, the students participating in this study might not necessarily represent the diversity of the university's student population but rather those motivated students with the ability to participate in extracurricular activities. The use of the suite of mLearning tools should be tested by the entire student population to explore this further. Finally, the focus of the study was on the experience of the students with mLearning, not the perspectives of the faculty or

preceptors, representing another limitation that should be addressed through future research.

Comparison With Previous Work

The longitudinal design in this study offered opportunities to assess the different digital learning resources in 2 contexts: the university and clinical rotation. This process also strengthened the authenticity and the *here and now* experiences in the processes of assessing the feasibility and benefits of learning resources. O'Connor and Andrews [11] have shown the benefits of using a co-design approach with nursing students when designing an educational app supporting clinical skills in general and not specifically for B-PASs. Previous research in the field of mLearning related to PAS has only tested 1 app without involving students in the development process [6], whereas other studies have only focused on a single digital learning resource [24,26,28]. In contrast, this study explores multiple digital learning resources with high student involvement and identified the different impacts on students' learning. The literature also highlights the need to reassess the educational strategies used to teach PASs to ensure that these skills are applied in a clinical setting [17,40,41].

Conclusions

Students valued the invitation and opportunity to collaborate in co-design processes, thereby influencing the nursing education content. The longitudinal research design structuring the collaboration with students was essential to understanding what works and in which context. The nursing students viewed the suite of mLearning tools as beneficial for supporting B-PAS learning and their application during clinical rotation. Our findings indicate that one of the strengths of the suite of mLearning tools was its inclusion of all the different digital resources tested by the students; this variety in content met the different learning preferences and needs, enhancing B-PAS learning and human bioscience knowledge of the students. Therefore, the suite of mLearning tools may be a beneficial additional pedagogical strategy that supports knowledge and skills transfer between academic and clinical settings. Further studies are needed to explore different perspectives related to the use of mobile technology and mLearning (ie, faculty and preceptors), pedagogical strategies, and scaffolding learning material to better understand how mLearning can be used in nursing education.

Acknowledgments

The authors would like to gratefully acknowledge the contribution of the participating students in the co-design processes and development of the suite of mLearning tools.

Authors' Contributions

All authors contributed to the different phases of the research process. ÖE drafted the manuscript, and all coauthors contributed to data analyses and further writing of this manuscript, the final draft of which was read and approved by all authors.

Conflicts of Interest

None declared.

References

1. Gallegos C, Gehrke P, Nakashima H. Can mobile devices be used as an active learning strategy? Student perceptions of mobile device use in a nursing course. Nurse Educ 2019;44(5):270-274. [doi: 10.1097/NNE.0000000000000613] [Medline: 30394996]
2. Ministry of EducationResearch. Digitalisation Strategy for the Higher Education Sector 2017-2021. Government.no. 2018. URL: https://www.regjeringen.no/en/dokumenter/digitalisation-strategy-for-the-higher-education-sector-2017-2021/id2571085/ [accessed 2020-01-19]
3. Yot-Domínguez C, Marcelo C. University students' self-regulated learning using digital technologies. Int J Educ Technol High Educ 2017 Nov 17;14(1). [doi: 10.1186/s41239-017-0076-8]
4. Trede F, Markauskaite L, McEwen C, MacFarlane S. Education for Practice in a Hybrid Space: Enhancing Professional Learning with Mobile Technology. Spingapore: Springer Singapore Pte. Limited; 2019:E.
5. Karimi S. Do learners' characteristics matter? An exploration of mobile-learning adoption in self-directed learning. Comput Hum Behav 2016 Oct;63:769-776. [doi: 10.1016/j.chb.2016.06.014]
6. Hsu L, Hsiang H, Tseng Y, Huang S, Hsieh S. Nursing students' experiences of using a smart phone application for a physical assessment course: a qualitative study. Jpn J Nurs Sci 2019 Apr;16(2):115-124. [doi: 10.1111/jjns.12215] [Medline: 29947101]
7. Barak M, Hussein-Farraj R, Dori YJ. On-campus or online: examining self-regulation and cognitive transfer skills in different learning settings. Int J Educ Technol High Educ 2016 Nov 1;13(1). [doi: 10.1186/s41239-016-0035-9]
8. O'Connor S, Andrews T. Smartphones and mobile applications (apps) in clinical nursing education: a student perspective. Nurse Educ Today 2018 Oct;69:172-178. [doi: 10.1016/j.nedt.2018.07.013] [Medline: 30096510]
9. Zieber M, Wojtowicz B. To dwell within: bridging the theory-practice gap. Nurs Philos 2020 Apr;21(2):e12296. [doi: 10.1111/nup.12296] [Medline: 31887788]

10. Ally M, Prieto-Blázquez J. Quin és el futur de l'aprenentatge mòbil en l'educació? RUSC Rev de Univ Y Soc del Con 2014 Jan 15;11(1):142. [doi: 10.7238/rusc.v11i1.2033]
11. O'Connor S, Andrews T. Using co-design with nursing students to create educational apps for clinical training. Stud Health Technol Inform 2016;225:334-338. [Medline: 27332217]
12. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
13. Jarvis C. Physical Examination and Health Assessment 8th Edition. Philadelphia: Saunders; 2019.
14. Birks M, Cant R, James A, Chung C, Davis J. The use of physical assessment skills by registered nurses in Australia: issues for nursing education. Collegian 2013;20(1):27-33. [doi: 10.1016/j.colegn.2012.02.004] [Medline: 23678781]
15. Hoffman KA, Aitken LM, Duffield C. A comparison of novice and expert nurses' cue collection during clinical decision-making: verbal protocol analysis. Int J Nurs Stud 2009 Oct;46(10):1335-1344. [doi: 10.1016/j.ijnurstu.2009.04.001] [Medline: 19555954]
16. Levett-Jones T, Hoffman K, Dempsey J, Jeong SY, Noble D, Norton CA, et al. The 'five rights' of clinical reasoning: an educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. Nurse Educ Today 2010 Aug;30(6):515-520. [doi: 10.1016/j.nedt.2009.10.020] [Medline: 19948370]
17. Egilsdottir H, Byermoen KR, Moen A, Eide H. Revitalizing physical assessment in undergraduate nursing education - what skills are important to learn, and how are these skills applied during clinical rotation? A cohort study. BMC Nurs 2019;18:41 [FREE Full text] [doi: 10.1186/s12912-019-0364-9] [Medline: 31516382]
18. Rammeplan For Sykepleierutdanning. Ministry of Education and Research. 2008. URL: https://www.regjeringen.no/globalassets/upload/kd/vedlegg/uh/rammeplaner/helse/rammeplan_sykepleierutdanning_08.pdf [accessed 2020-01-19]
19. Breivik S, Tymi A. Hva innebærer systematisk klinisk undersøkelse i sykepleie? Sykepleien 2013(4):324-332. [doi: 10.4220/sykepleienf.2013.0132]
20. Jensen KT, Knutstad U, Fawcett TN. The challenge of the biosciences in nurse education: a literature review. J Clin Nurs 2018 May;27(9-10):1793-1802. [doi: 10.1111/jocn.14358] [Medline: 29575462]
21. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
22. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937. [doi: 10.2196/12937] [Medline: 30747711]
23. McCutcheon K, Lohan M, Traynor M, Martin D. A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. J Adv Nurs 2015 Feb;71(2):255-270. [doi: 10.1111/jan.12509] [Medline: 25134985]
24. Barisone M, Bagnasco A, Aleo G, Catania G, Bona M, Scaglia SG, et al. The effectiveness of web-based learning in supporting the development of nursing students' practical skills during clinical placements: a qualitative study. Nurse Educ Pract 2019 May;37:56-61. [doi: 10.1016/j.nepr.2019.02.009] [Medline: 31085383]
25. Lunde L, Moen A, Rosvold E. Learning clinical assessment and interdisciplinary team collaboration in primary care. MOOC for healthcare practitioners and students. Stud Health Technol Inform 2018;250:68. [Medline: 29857376]
26. Vogt M, Schaffner B, Ribar A, Chavez R. The impact of podcasting on the learning and satisfaction of undergraduate nursing students. Nurse Educ Pract 2010 Jan;10(1):38-42. [doi: 10.1016/j.nepr.2009.03.006] [Medline: 19781997]
27. Mayer RE. Multimedia Learning, 2nd ed. Cambridge, United Kingdom: Cambridge Univerity Press; 2009.
28. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: randomized controlled trial. J Med Internet Res 2019 Mar 18;21(3):e11529 [FREE Full text] [doi: 10.2196/11529] [Medline: 30882355]
29. Berg H, Steinsbekk A. Is individual practice in an immersive and interactive virtual reality application non-inferior to practicing with traditional equipment in learning systematic clinical observation? A randomized controlled trial. BMC Med Educ 2020 Apr 22;20(1):123 [FREE Full text] [doi: 10.1186/s12909-020-02030-7] [Medline: 32326948]
30. Cant R, Cooper S, Sussex R, Bogossian F. What's in a name? Clarifying the nomenclature of virtual simulation. Clin Simul Nurs 2019 Feb;27:26-30. [doi: 10.1016/j.ecns.2018.11.003]
31. Lall P, Rees R, Law GC, Dunleavy G, Cotič Z, Car J. Influences on the implementation of mobile learning for medical and nursing education: qualitative systematic review by the digital health education collaboration. J Med Internet Res 2019 Feb 28;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
32. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
33. Raman J. Mobile technology in nursing education: where do we go from here? A review of the literature. Nurse Educ Today 2015 May;35(5):663-672. [doi: 10.1016/j.nedt.2015.01.018] [Medline: 25665926]
34. Blum CA. Does podcast use enhance critical thinking in nursing education? Nurs Educ Perspect 2018;39(2):91-93. [doi: 10.1097/01.NEP.0000000000000239] [Medline: 29215390]
35. Koivisto J, Haavisto E, Niemi H, Katajisto J, Multisilta J. Elements explaining learning clinical reasoning using simulation games. International Journal of Serious Games, 2016 Dec 21;3(4):29-43. [doi: 10.17083/ijsg.v3i4.136]

36. Peddle M, Bearman M, Nestel D. Virtual patients and nontechnical skills in undergraduate health professional education: an integrative review. Clin Simul Nurs 2016 Sep;12(9):400-410. [doi: 10.1016/j.ecns.2016.04.004]
37. Lewin C, Lai K, van Bergen H, Charania A, Ntebutse JG, Quinn B, et al. Integrating academic and everyday learning through technology: issues and challenges for researchers, policy makers and practitioners. Tech Know Learn 2018 Aug 9;23(3):391-407. [doi: 10.1007/s10758-018-9381-0]
38. Chan T, Roschelle J, Hsi S, Kinshuk, Sharples M, Brown T, et al. One-to-one technology-enhanced learning: an opportunity for global research collaboration. Res Practice Tech Enhanced Learning 2006 Mar;01(01):3-29. [doi: 10.1142/S1793206806000032]
39. Martyn J, Larkin K, Sander T, Yuginovich T, Jamieson-Proctor R. Distance and devices - potential barriers to use of wireless handheld devices. Nurse Educ Today 2014 Mar;34(3):457-461. [doi: 10.1016/j.nedt.2013.04.021] [Medline: 23706774]
40. Fennessey AG, Wittmann-Price RA. Physical assessment: a continuing need for clarification. Nurs Forum 2011;46(1):45-50. [doi: 10.1111/j.1744-6198.2010.00209.x] [Medline: 21306395]
41. Birks M, James A, Chung C, Cant R, Davis J. The teaching of physical assessment skills in pre-registration nursing programmes in Australia: issues for nursing education. Collegian 2014;21(3):245-253. [doi: 10.1016/j.colegn.2013.05.001] [Medline: 25632720]

Abbreviations

B-PAS: basic physical assessment skill
LMS: learning management system
MCQ: multiple-choice question
mLearning: mobile learning
MOOC: massive open online course
PAS: physical assessment skill
WS: workshop

Edited by G Eysenbach; submitted 18.07.20; peer-reviewed by M Padilha, P Sousa; comments to author 10.08.20; revised version received 22.09.20; accepted 25.10.20; published 07.01.21.

Please cite as:
Egilsdottir HÖ, Heyn LG, Brembo EA, Byermoen KR, Moen A, Eide H
Configuration of Mobile Learning Tools to Support Basic Physical Assessment in Nursing Education: Longitudinal Participatory Design Approach
JMIR Mhealth Uhealth 2021;9(1):e22633
URL: https://mhealth.jmir.org/2021/1/e22633
doi:10.2196/22633
PMID:33410756

Original Paper

Implementing Mobile Health–Enabled Integrated Care for Complex Chronic Patients: Intervention Effectiveness and Cost-Effectiveness Study

Jordi de Batlle[1,2], PhD; Mireia Massip[1], BSc; Eloisa Vargiu[3], PhD; Nuria Nadal[4], MD; Araceli Fuentes[5], MD; Marta Ortega Bravo[6,7,8], MD; Felip Miralles[3], PhD; Ferran Barbé[1,2], MD; Gerard Torres[1,2], MD; CONNECARE-Lleida Group[9]

[1]Group of Translational Research in Respiratory Medicine, Institut de Recerca Biomedica de Lleida, Lleida, Spain
[2]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias CIBERES, Madrid, Spain
[3]Digital Health unit, Eurecat, Centre Tecnòlogic de Catalunya, Barcelona, Spain
[4]Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain
[5]Atenció Primària Àmbit Lleida, Lleida, Spain
[6]Research Support Unit Lleida, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Lleida, Spain
[7]Centre d'Atenció Primària Cappont, Gerència Territorial de Lleida, Institut Català de la Salut, Lleida, Spain
[8]Universitat de Lleida, Lleida, Spain
[9]Please see acknowledgements section for list of collaborators

Corresponding Author:
Gerard Torres, MD
Group of Translational Research in Respiratory Medicine
Institut de Recerca Biomedica de Lleida
80 Rovira Roure
Lleida, 25198
Spain
Phone: 34 973705372
Email: gtorres@gss.cat

Abstract

Background: Integrated care can generate health and social care efficiencies through the defragmentation of care and adoption of patient-centered preventive models. eHealth can be a key enabling technology for integrated care.

Objective: The aim of this study was to assess the effectiveness and cost-effectiveness of the implementation of a mobile health (mHealth)-enabled integrated care model for complex chronic patients.

Methods: As part of the CONNECARE Horizon 2020 project, a prospective, pragmatic, two-arm, parallel implementation trial was held in a rural region of Catalonia, Spain. During 3 months, elderly patients with chronic obstructive pulmonary disease or heart failure and their carers experienced the combined benefits of the CONNECARE organizational integrated care model and the eHealth platform supporting it, consisting of a patient self-management app, a set of integrated sensors, and a web-based platform connecting professionals from different settings, or usual care. We assessed changes in health status with the 12-Item Short-Form Survey (SF-12), unplanned visits and admissions during a 6-month follow up, and the incremental cost-effectiveness ratio (ICER).

Results: A total of 48 patients were included in the integrated care arm and 28 patients receiving usual care were included in the control arm (mean age 82 years, SD 7 years; mean Charlson index 7, SD 2). Integrated care patients showed a significant increase in the SF-12 physical domain with a mean change of +3.7 (SD 8.4) (P=.004) and total SF-12 score with a mean change of +5.8 (SD 12.8) (P=.003); however, the differences in differences between groups were not statistically significant. Integrated care patients had 57% less unplanned visits (P=.004) and 50% less hospital admissions related to their main chronic diseases (P=.32). The integrated care program generated savings in different cost scenarios and the ICER demonstrated the cost-effectiveness of the program.

Conclusions: The implementation of a patient-centered mHealth-enabled integrated care model empowering the patient, and connecting primary, hospital, and social care professionals reduced unplanned contacts with the health system and health costs,

and was cost-effective. These findings support the notion of system-wide cross-organizational care pathways supported by mHealth as a successful way to implement integrated care.

(JMIR Mhealth Uhealth 2021;9(1):e22135) doi:10.2196/22135

KEYWORDS

chronic disease; cost-benefit analysis; delivery of health care, integrated; mHealth; eHealth; quality of life

Introduction

The last decades have led to rapidly aging populations and an increased burden of chronic diseases [1]. In this scenario, health and social care providers struggle to contain costs while providing an adequate response to the population's care needs. Traditional care models suffer from care fragmentation. That is, the different care settings fail to communicate with each other effectively, and patients have to undergo repeated tests and changes in prescribed drugs, ultimately prompting feelings of starting anew after every transition [2]. Additionally, the focus is still on the diseases rather than on the patients, which leaves patients and their carers as passive actors [2]. Therefore, there is a need for a profound redesign of how care is provided to elderly patients with chronic conditions to ensure quality and sustainability [3]. Integrated care models aim to generate health and social care efficiencies through the defragmentation of care, promotion of collaboration and continuity of care across settings, adoption of patient-centered models, and prioritization of preventive models [4]. However, few of these models have attempted to and succeeded in simultaneously tackling all of the above-mentioned measures [4]. eHealth and mobile health (mHealth) can be the key enabling technologies allowing for such a paradigm shift [5].

The Personalised Connected Care for Complex Chronic Patients (CONNECARE) project is a Horizon 2020 European Union Research and Innovation project aiming to co-design, develop, deploy, and evaluate a novel smart and adaptive organizational integrated care model for complex chronic patients (CCP) [6]. From April 2016 to December 2019, the project successfully co-designed, by means of an iterative patient-centered process involving patients and stakeholders across different health settings, an organizational model for integrated care and an eHealth platform to support the model. The integrated care model promotes collaboration among professionals of different care settings (family physicians, hospital specialists, and social workers), prioritizes home-based prevention over institutional reactive care, and fosters patient empowerment. The use of a web-based platform offers a cross-setting web-based Smart Adaptive Case Management (SACM) system for professionals, and an mHealth self-management system with three-level monitoring features allowing for patient empowerment.

As part of the CONNECARE project, a novel mHealth-enabled integrated care model was implemented in Lleida, Spain. The existing care model in Lleida is limited by the scarce communication between professionals of different care settings, with different electronic medical record (EMR) systems in hospitals and primary care centers (Argos SAP and ECAP [7], respectively), and patients playing a passive role throughout the care path. We here describe the results of the implementation of an mHealth-enabled integrated care model for the community-based prevention of unplanned hospital-related events in CCP with a high risk for hospitalization.

Methods

Study Design

This was a prospective, pragmatic, two-arm, parallel implementation trial comparing care as usual to a 3-month mHealth-enabled integrated care intervention. The study was conducted from July 2018 to August 2019 in Lleida, which is a large rural area of over 4300 km^2, including two tertiary hospitals (University Hospital Arnau de Vilanova and University Hospital Santa Maria) and a network of 23 primary care centers spread across the whole territory, providing service to 400,000 citizens.

Target Population

Home-dwelling patients with chronic conditions and a history of hospitalizations were recruited for this study. The eligibility criteria were aged ≥55 years; admitted to hospital for a respiratory or cardiovascular event (ie, chronic obstructive pulmonary disease exacerbation or heart failure decompensation); living at home and discharged back to the community; no dementia or cognitive impairment (Global Deterioration Scale score < 5 [8]); Length, Acuity, Comorbidities and Emergency score > 7 [9]; and passing a basic technological test assessing home connectivity and patients' or carers' competences with the use of technology. The basic technological test is shown in Multimedia Appendix 1.

Recruitment

Patients were recruited during an unanticipated admission to the hospital through the emergency room (ER). They were identified based on EMR data and were contacted by a case manager before discharge. After the recruitment of patients to the intervention arm, an active search for a matched control with similar characteristics began. All patients and their carers, regardless of study arm, received a face-to-face explanation about the study.

Intervention

Patients in the intervention arm experienced an integrated care model, including (i) preliminary assessment of the patient's health status using several questionnaires, tests, and indices specific to their main chronic diseases and social needs; (ii) a self-management app, with status and performance reports, a virtual coach with customizable automated feedback, and full communication with the care team; (iii) a Fitbit Flex 2 digital activity tracker [10] and any additional sensor deemed necessary by the care team [11], including a digital pulse-oximeter, digital

scale, and digital blood pressure monitor, that were fully integrated into the self-management app; (iv) a patient profile in the SACM web-based platform, accessible to all members of the care team (family physicians, hospital specialists, and social workers), that was used for coordination and communication among professionals in the different settings, and to contact the patient when needed; and (v) assignment of a case manager in charge of supervising the whole process and serving as the main patient contact point. Additional details on the CONNECARE integrated care model and the supporting eHealth platform can be found in Multimedia Appendix 1. Patients in the control arm experienced care as usual, managed from primary care. After discharge from the initial 90 days of usual care for integrated care management, all patients were passively followed up for an additional 3 months.

Data Collection

Variables characterizing the patients were collected at recruitment using the SACM in tablet or desktop computers, including age, sex, main chronic disease(s), Charlson index of comorbidities [12], quality of life (QoL) as measured by the 12-Item Short-Form Survey (SF-12) [13], Barthel index for Activities of Daily Living [14], Hospital Anxiety and Depression scale [15], assessment of dwelling characteristics, main medications, Pfeiffer mental status questionnaire [16], and tobacco and alcohol consumption. The main outcomes were: (i) intervention effectiveness, as measured by the changes in the SF-12 health questionnaire's physical and mental domains (baseline vs discharge); (ii) use of health care resources after 6 months, and estimated associated costs based on Catalan Health Department official data [17]; and (iii) cost-effectiveness, based on the improvement in QoL relative to costs, assessed by means of the incremental cost-effectiveness ratio (ICER); all costs are described in US $ (conversion factor: 1 € = 1.21 US $). Additional details on cost estimations are described in Multimedia Appendix 1. The use of health care resources was collected from EMRs, which included hospital admissions, ER visits, visits to primary care, and visits to hospital specialists. Additionally, each admission or visit was assessed regarding its relation to the patient's chronic diseases as a binary variable (related or unrelated).

Statistical Analyses

Participants' baseline characteristics are summarized as n (%), mean (SD), or median (IQR) as appropriate. Comparisons between baseline characteristics of patients in the integrated care and control groups were performed using the χ^2 test, *t* test, or Kruskal-Wallis test, as appropriate. A paired *t* test was used to compare baseline to discharge values with respect to the SF-12 domains. Linear regression models were used to assess differences in the changes experienced by patients in the two groups. Negative binomial regression models were used to assess differences in the number of visits and admissions. Models were adjusted by age, sex, and Charlson comorbidity index. The ICER was calculated in relation to the SF-12 total score in three different scenarios: 100%, 150%, and 200% estimated cost of the integrated care program. Data analyses were conducted using Stata version 12.1 (StataCorp, College Station, TX, USA). The threshold for significance was P<.05.

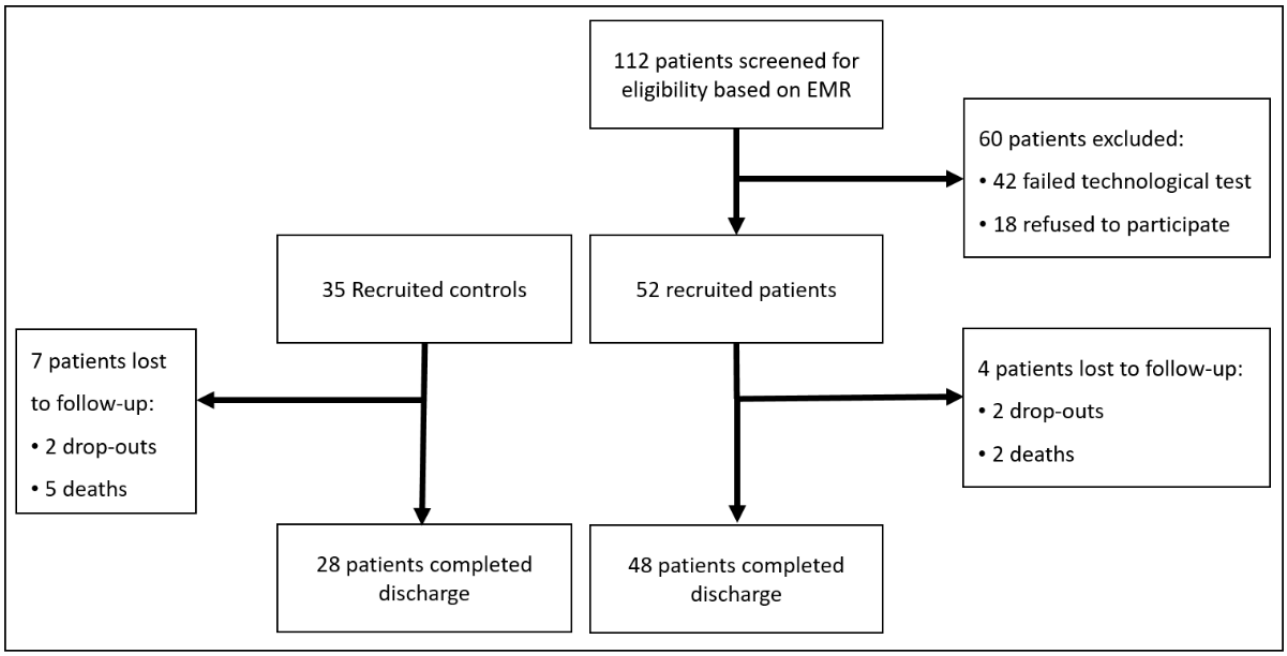
Ethical Considerations

This study was approved by Ethics Committee of Hospital Arnau de Vilanova (CEIC-1685) and all patients provided written informed consent. All collected data were handled and stored in accordance with current national and international legislation.

Results

Up to 112 patients were screened for eligibility. After excluding patients not meeting the inclusion criteria, 52 patients were recruited for the mHealth-enabled integrated care arm and 35 patients were recruited for the usual care arm. Final analyses were based on 48 integrated care and 28 control patients completing the follow up (Figure 1).

Figure 1. Study flowchart. EMR: electronic medical record.



The patients' baseline characteristics are shown in Table 1. There were no significant differences in mean age or Charlson index between the patients in the two arms.

Table 1. Baseline characteristics of patients in the usual care and integrated care (IC) arms.

Characteristic	Usual care (n=28)	IC (n=48)	*P* value[a]
Sex (male), n (%)	17 (61)	24 (50)	.37
Age (years), mean (SD)	82 (8)	82 (7)	.88
Charlson score, mean (SD)	7.4 (2.1)	6.7 (2.0)	.15
LACE[b] score, median (IQR)	15 (13-17)	14 (12-17)	.38
Barthel score, median (IQR)	90 (72.5-95)	90 (67.5-100)	.40
HAD[c] anxiety score, mean (SD)	4.9 (3.5)	4.3 (2.7)	.44
HAD depression score, mean (SD)	5.6 (2.9)	5.7 (2.3)	.95
Pfeiffer intact intellectual functioning, n (%)	21 (75)	37 (77)	.67
GDS[d], no cognitive decline, n (%)	27 (96)	44 (92)	.25

[a]χ^2 test, *t* test, or Kruskal-Wallis equality-of-populations rank test, as appropriate.

[b]LACE: Length, Acuity, Comorbidities, and Emergency score.

[c]HAD: Hospital Anxiety and Depression scale.

[d]GDS: Global Deterioration Scale.

Table 2 shows the changes in QoL (SF-12 domains) from baseline to discharge. Patients in the integrated care arm showed a significant increase in the SF-12 physical domain and total SF-12 scores. The differences in QoL between integrated care and control patients favored the integrated care patients but did not achieve statistical significance.

Table 2. Changes in health status in the usual care and integrated care arms.

SF-12[a] score	Baseline, mean (SD)	Discharge, mean (SD)	Change, mean (SD)	*P* value[b]
Physical				
Usual care	29.6 (8.3)	31.6 (9.0)	+2.0 (7.5)	.16
Integrated care	29.0 (7.3)	32.7 (9.4)	+3.7 (8.4)	.004
Difference	–0.6 (2.0)	+1.1 (2.0)	+1.7 (2.9)	.21
Mental				
Usual care	47.0 (13.5)	45.8 (15.5)	–1.2 (11.9)	.59
Integrated care	51.8 (9.9)	53.9 (11.5)	+2.0 (11.2)	.21
Difference	+4.2 (3.7)	+9.2 (3.7)	+5.0 (5.2)	.10
Total				
Usual care	76.6 (13.9)	77.4 (20.5)	+0.8 (14.7)	.77
Integrated care	80.8 (11.9)	86.6 (16.3)	+5.8 (12.8)	.003
Difference	+4.2 (3.7)	+9.2 (3.7)	+5.0 (5.2)	.10

[a]SF-12: 12-Item Short-Form Survey.

[b]Paired *t* test comparing baseline and discharge measures; linear regression predicting the difference on baseline and discharge measures according to intervention arm, adjusted by age, sex, and Charlson index.

Table 3 shows that integrated care patients had 57% less unplanned visits, representing a significant difference. Integrated care patients also experienced a 50% reduction in hospital admissions related to their main chronic diseases, although this difference was not statistically significant.

Table 3. Total use of health services during the follow-up period.

Health service	Usual care (n=28), mean (SD)	Integrated care (n=48), mean (SD)	*P* value[a]	Adjusted *P* value[b]
All unplanned visits	2.3 (3.1)	1.0 (1.1)	.001	.004
Unplanned visits related to chronic disease	0.9 (1.2)	0.4 (0.6)	.01	.04
All hospital admissions	0.5 (0.8)	0.4 (0.6)	.35	.50
Hospital admissions related to chronic disease	0.4 (0.7)	0.2 (0.5)	.18	.32

[a]Negative binomial regression model.

[b]Negative binomial regression model adjusted by age, sex, and Charlson comorbidity index.

The results for the within trial costs and cost-effectiveness analyses for all unplanned visits and hospital admissions are shown in Table 4, and those for unplanned visits and hospital admissions related to the patient's main chronic diseases are shown in Table S1 of Multimedia Appendix 1. The integrated care program generated savings from US $584 to $1434 per patient, depending on the scenarios. The integrated care program was cost-effective according to the ICER, performing better in terms of QoL while reducing overall expenses.

Table 4. Within trial costs (average cost per patient) and cost-effectiveness considering all unplanned visits and hospital admissions in three integrated care (IC) program cost scenarios.

Related cost	Usual care (n=28), US $	IC (n=48), US $	Difference	ICER[a]
Unplanned visits[b]	173.62	78.02	–95.60	N/A[c]
Hospital admissions[b]	3069.94	2404.15	–665.79	N/A
Total medical costs per patient	3243.56	2482.17	–761.39	N/A
Scenario 1: 100% IC program costs				
CONNECARE program	0	85.92	85.92	N/A
Total costs per patient	3243.56	2568.09	–675.47	–135.64
Scenario 2: 150% IC program costs				
CONNECARE program	0	128.89	128.89	N/A
Total costs per patient	3243.56	2611.06	–632.50	–127.01
Scenario 3: 200% IC program costs				
CONNECARE program	0	171.84	171.84	
Total costs per patient	3243.56	2654.01	–589.55	–118.39

[a]ICER: incremental cost-effectiveness ratio; incremental cost associated with 1 additional point gain in 12-Item Short-Form Survey (SF-12).

[b]Costs based on the Catalan Institute of Health official pricing (CVE-DOGC-A-13051031-2013).

[c]N/A: not applicable.

Discussion

Principal Results

The prospective assessment of the implementation of an mHealth-enabled integrated care program for CCP management showed a reduction in the number of unplanned contacts with the health system, generated substantial savings for the health system without having any negative impact on QoL or clinical outcomes, and demonstrated cost-effectiveness.

Strengths and Limitations

A key strength of this study was the effort to involve, from the very beginning, all of the stakeholders from different organizations that would be actors in a large-scale deployment of the mHealth-enabled integrated care program. This is especially relevant as the lack of cooperation between organizations, teams, or professions is a recurrent barrier toward the implementation of integrated care [4]. Other relevant strengths included: (i) the involvement of informal carers in the integrated care process, as close relatives of patients are usually the link between the patients and the health system, and such informal carers were key for facilitating the use of the self-management app in patients with a mean age over 80 years; (ii) the use of a self-management app (including a virtual coach with customizable automated feedback, full communication with the care team, and active monitoring), as such apps can significantly enhance doctor-patient relationships [18] and detect worsening in patient conditions or frailty [19]; (iii) the promotion and assessment of patients' physical activity, as mobility impairment is found in one third of people over 65 years [20]; (iii) the implementation region, a large rural area of over 4300 km^2, which can benefit the most from

community-based integrated care initiatives that preclude unnecessary travels to the hospital; and (iv) the prospective study design.

In terms of limitations, although the organizational model did not experience substantial changes throughout the implementation period, the supporting technological platform was in a permanent process of refinement and addition of new functionalities. This implied that the integrated care experience was richer in patients who were recruited near the end of the implementation study compared with that of patients recruited at the very beginning. Similarly, this had a considerable impact on the health care professionals, who had to cope with a platform in constant development and that was not fully integrated with the existing EMR. However, directly participating in a dynamic development and implementation process allowed the professionals to feel engaged and propose changes, and for new features to be developed, which ultimately resulted in not a single professional dropping out of the implementation study. Other limitations were: (i) the relatively small number of patients involved in this first phase of the deployment, which ultimately affected statistical power; (ii) the assumptions held while determining the costs of the mHealth-enabled integrated care program intervention, as well as the lack of assessment of indirect (societal) savings; and (iii) the strategy of centralizing the entry points to the integrated care program in the hospital (after ER admission), as it is important that system-wide cross-organizational care pathways consider multiple entry points [21]. In this regard, upcoming phases of the implementation will consider additional entry points such as the primary care centers.

Comparison With Existing Literature

The impact of the implementation of the integrated care model was assessed in three domains: (i) patients' QoL, (ii) use of health services, and (iii) economics. In the first domain, QoL, the integrated care model performed slightly better than usual care, with positive differences-in-differences values but not reaching statistical significance. This result is in line with a 2017 umbrella review concluding that integrated care interventions showed mixed results in terms of improving patients' QoL [22]. In this sense, it must be noted that the short duration of the implemented intervention (3 months) could have limited its capacity to affect the overall QoL. Nevertheless, the short duration did not preclude obtaining excellent results in terms of the use of health services. The integrated care model reduced the number of unplanned visits by 57% and the number of hospital admissions related to the main chronic disease of each patient by 50%. Half of the published reviews on integrated care interventions between 2000 and 2015 reported significant reductions in hospital activity, ranging from 15% to 50% [23]. This places the reported results for the integrated care program within the top margin of positive results, and supports the notion of system-wide cross-organizational care pathways as a successful way to implement integrated care in contrast to smaller and narrow interventions [2]. Finally, regarding the economic impact, integrated care generated savings from US $584 to $1434 per patient and was cost-effective. This is in line with reviews stating the potential cost-effectiveness of integrated care for the management of chronic diseases [24]. Our results are also in line with savings found in other chronic diseases such as diabetes mellitus (from –1508 to +299 Euro; approximately US $–1809 to +359) or schizophrenia (from –3860 to +614 Euro; approximately US $–4632 to 737) [25].

Implications for Research and Practice

The need for people-centered integrated care, capable of providing an adequate response to the needs of growing populations of older people with chronic conditions while keeping costs sustainable, has been clearly stated by the World Health Organization [26]. In the frame of the CONNECARE project, an mHealth-enabled integrated care model was co-designed through an iterative process that involved all of the key stakeholders: patients, hospital and primary care medical and technical staff, social carers, managers, developers, and researchers. This coproduction multidisciplinary team had a clear focus on the patient and was ready to consider system-wide cross-organizational care pathways. The resulting mHealth-enabled integrated care model was thus perfectly aligned with the 2015 World Health Organization report on aging and health [1], which states that care providers should ensure that: (i) the assessment of individual impairments/declines in capacity is used to inform the development of a comprehensive care plan, and all domains are assessed together; (ii) interventions encouraging physical exercise are included in the care plans; and (iii) the presence of any impairment/decline in capacity triggers actions for the medical assessment of associated diseases. Moreover, we identified key features to be included in successful integrated care models for elder CCP: (i) the involvement of informal caregivers, being key in the adoption of mHealth tools such as self-management apps and sensors, and making the overall user experience very satisfactory [27]; (ii) the enablement of a common web-based platform for the coordination of care across settings and patient follow up; and (iii) the enhancement of communication channels for patients, which reduced the need of face-to-face appointments for quick consultations or questions. This latter aspect is especially relevant when patients are depending on others for travel to the general practitioners' offices or primary care centers.

Conclusion

The implementation of a patient-centered mHealth-enabled integrated care model empowering patients and connecting primary, hospital, and social care professionals reduced unplanned contacts with the health system and health costs, and was cost-effective. This supports the notion of system-wide cross-organizational care pathways using mHealth tools as a successful way to implement integrated care.

Acknowledgments

This work was supported by the European Union's Horizon 2020 Research and Innovation Program (under grant agreement GA-689802). JdB acknowledges receiving financial support from the Catalan Health Department (Pla Estratègic de Recerca i Innovació en Salut [PERIS] 2016: SLT002/16/00364) and Instituto de Salud Carlos III (ISCIII; Miguel Servet 2019: CP19/00108), cofunded by the European Social Fund (ESF), "Investing in your future." The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The authors would like to express their sincerest gratitude to all of the participating patients and their carers, as well as all of the involved health and social care professionals, who were key in the co-design of the integrated care model and its implementation in the region of Lleida.

Authors' Contributions

JdB, EV, NN, FM, FB, and GT participated in the conceptualization of project. MM, EV, AF, MB, and GT conducted data collection. MM, EV, and JdB participated in data curation. JdB conducted all statistical analyses. JdB wrote the original draft of the manuscript. All authors reviewed the final manuscript. JdB, EV, FM, FB, and GT secured funding for the project. The CONNECARE-Lleida group at Institut de Recerca Biomedica de Lleida, Lleida, Spain: Maria Aguilà Balastegui, Sandra Alexandre Loxano, Laila Al-Jouja Llorente, Tomás Alonso Sancho, Enrique Aparicio Bañeres, Ana Arce Vila, Jose Maria Baron Burriel, Ramon Bascompte Claret, Albert Bigorda Sague, Emilia Blanco Ponce, Maria Boldú Franque, Àngels Bosch Roig, Carmen Bravo Santiago, Alba Capdevila Sarramona, Aida Castelló Corretge, Jordi Colomina Morales, Montse Coma Gassó, Fina Cregenzan Ortiz, Dolors Del Pozo Garcia, Reis Drudis Morell, Mireia Falguera Vilamajó, Pere Farre Pagés, Yolanda Fauria Garcia, Anabel Fusalba Canales, Jara Gayan Ordas, Sergi Godia Lopez, Irene Gomez Companys, Jessica Gonzàlez Gutierrez, Anna Gort Oromí, Carme Jorge Tufet, Mercé Lavega Llorens, Laia Llort Samsó, Maria Rosa Lopez Cervelló, Belen Malla Clua, Josep Maria Marsol Mas, Teresita Martí Ribes, Diana Martin Capella, José Maria Martínez Barriuso, Esther Mateus Solé, Ramon Mazana Novellon, Petra Merino De los Santos, Miquel Mesas Julio, Sonia Minguet Vidal, Nuria Moles Porta, Luis Miguel Montaña Esteban, Dolors Morera Roset, Meritxell Moyà Oro, Irene Muñoz Del Campo, Francisco Nicolás Sánchez, Inés Ortiz Catalán, Mireia Ortiz Valls, Sonia Ortiz Congost, Jose Maria Palacin Peruga, Francesc Pallisó Folch, Eugeni Paredes Costa, Pablo Pastor Pueyo, Ana Pérez Sainz, Antonio Plana Blanco, Anna Planas Hiraldo, Pepita Pont Aldoma, Marife Quelle Alonso, Rebeca Ramirez Molinero, Maria Àngels Revés Juanbaro, Anna Ribé Miró, Eva Ribó Caubet, Rebeca Rodriguez Corbaton, Marina Rué Florensa, Oscar Sacristán García, Irene Sanmartí Forns, Maria Cruz Sanz Martinez, Neus Sendra Bordes, Montse Torra Riera, Maria Cruz Urgelés Castillón, Laia Utrillo Montagut, Montse Vidal Ballesté.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online supplement.
[DOCX File , 1567 KB - mhealth_v9i1e22135_app1.docx]

References

1. World Report on Ageing and Health. World Health Organization. 2015. URL: https://www.who.int/ageing/events/world-report-2015-launch/en/ [accessed 2020-12-09]
2. Harvey G, Dollard J, Marshall A, Mittinty MM. Achieving Integrated Care for Older People: Shuffling the Deckchairs or Making the System Watertight For the Future? Int J Health Policy Manag 2018 Apr 01;7(4):290-293 [FREE Full text] [doi: 10.15171/ijhpm.2017.144] [Medline: 29626395]
3. World Health Organization. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. Geneva, Switzerland: World Health Organization (WHO); 2017.
4. Dates M, Lennox-chhugani N, Sant H, Pereira A, Tedeschi M. Health system performance assessment – Integrated Care Assessment (20157303 HSPA). Luxembourg: European Commission; 2018. URL: https://ec.europa.eu/health/sites/health/files/systems_performance_assessment/docs/2018_integratedcareassessment_en.pdf [accessed 2020-12-09]
5. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: Toward new models for chronic disease management. Maturitas 2015 Sep;82(1):22-27. [doi: 10.1016/j.maturitas.2015.03.015] [Medline: 25891502]
6. The CONNECARE project. URL: http://www.connecare.eu/ [accessed 2020-03-19]
7. eCAP. URL: http://salutweb.gencat.cat/ca/ambits_actuacio/linies_dactuacio/tecnologies_informacio_i_comunicacio/ecap/ [accessed 2020-03-19]
8. Reisberg B, Ferris S, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982 Sep;139(9):1136-1139. [doi: 10.1176/ajp.139.9.1136] [Medline: 7114305]

9. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010 Apr 06;182(6):551-557 [FREE Full text] [doi: 10.1503/cmaj.091117] [Medline: 20194559]
10. Fitbit Flex 2 digital activity tracker. URL: http://www.fitbit.com/in/flex2 [accessed 2020-08-25]
11. Withings. URL: http://www.withings.com/ [accessed 2020-08-25]
12. Charlson ME, Pompei P, Ales KL, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
13. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
14. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988;10(2):61-63. [doi: 10.3109/09638288809164103] [Medline: 3403500]
15. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
16. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975 Oct;23(10):433-441. [doi: 10.1111/j.1532-5415.1975.tb00927.x] [Medline: 1159263]
17. CVE-DOGC-A-13051031-2013: Resolution SLT/353/2013, of February 13, on the review of public prices for the services provided by the Catalan Health Institute. URL: https://www.comb.cat/Upload/Documents/4635.pdf [accessed 2020-03-19]
18. Józsa T. The relevance of mobile applications helping in doctor-patient relationships. Br J Gen Pract 2020 Mar;70(692):109. [doi: 10.3399/bjgp20X708377] [Medline: 32107220]
19. Aznar-Tortonda V, Palazón-Bru A, la Rosa DMF, Espínola-Morel V, Pérez-Pérez BF, León-Ruiz AB, et al. Detection of frailty in older patients using a mobile app: cross-sectional observational study in primary care. Br J Gen Pract 2020 Jan;70(690):e29-e35 [FREE Full text] [doi: 10.3399/bjgp19X706577] [Medline: 31685541]
20. Mitra S, Sambamoorthi U. Disability prevalence among adults: estimates for 54 countries and progress toward a global estimate. Disabil Rehabil 2014;36(11):940-947. [doi: 10.3109/09638288.2013.825333] [Medline: 23962193]
21. Borgermans L, Devroey D. A Policy Guide on Integrated Care (PGIC): Lessons Learned from EU Project INTEGRATE and Beyond. Int J Integr Care 2017 Sep 25;17(4):8 [FREE Full text] [doi: 10.5334/ijic.3295] [Medline: 29588631]
22. Flanagan S, Damery S, Combes G. The effectiveness of integrated care interventions in improving patient quality of life (QoL) for patients with chronic conditions. An overview of the systematic review evidence. Health Qual Life Outcomes 2017 Sep 29;15(1):188 [FREE Full text] [doi: 10.1186/s12955-017-0765-y] [Medline: 28962570]
23. Damery S, Flanagan S, Combes G. Does integrated care reduce hospital activity for patients with chronic diseases? An umbrella review of systematic reviews. BMJ Open 2016 Nov 21;6(11):e011952 [FREE Full text] [doi: 10.1136/bmjopen-2016-011952] [Medline: 27872113]
24. Cronin J, Murphy A, Savage E. Can chronic disease be managed through integrated care cost-effectively? Evidence from a systematic review. Ir J Med Sci 2017 Nov;186(4):827-834. [doi: 10.1007/s11845-017-1600-5] [Medline: 28477328]
25. Desmedt M, Vertriest S, Hellings J, Bergs J, Dessers E, Vankrunkelsven P, et al. Economic impact of integrated care models for patients with chronic diseases: a systematic review. Value Health 2016;19(6):892-902 [FREE Full text] [doi: 10.1016/j.jval.2016.05.001] [Medline: 27712719]
26. Global strategy on people-centred and integrated health services. World Health Organization. Geneva, Switzerland; 2015 Mar. URL: https://www.who.int/servicedeliverysafety/areas/people-centred-care/global-strategy/en/ [accessed 2020-12-09]
27. Vargiu E, Dorado A, Massip M, Gómez-Martínez M, Torres G, de Batlle J, et al. User Experience in CONNECARE: Results and Lessons Learnt in Lleida. XPA & Health Com 2019;2 [FREE Full text]

Abbreviations

CCP: complex chronic patients
CONNECARE: Personalised Connected Care for Complex Chronic Patients
EMR: electronic medical records
ER: emergency room
ICER: incremental cost-effectiveness ratio
mHealth: mobile health
QoL: quality of life
SACM: Smart Adaptive Case Management
SF-12: 12-Item Short-Form Survey

Edited by G Eysenbach; submitted 04.07.20; peer-reviewed by I Cano, J Roca; comments to author 08.08.20; revised version received 31.08.20; accepted 29.10.20; published 14.01.21.

Please cite as:
de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega Bravo M, Miralles F, Barbé F, Torres G, CONNECARE-Lleida Group Implementing Mobile Health–Enabled Integrated Care for Complex Chronic Patients: Intervention Effectiveness and Cost-Effectiveness Study
JMIR Mhealth Uhealth 2021;9(1):e22135
URL: https://mhealth.jmir.org/2021/1/e22135
doi:10.2196/22135
PMID:33443486

Original Paper

Measuring Daily Compliance With Physical Activity Tracking in Ambulatory Surgery Patients: Comparative Analysis of Five Compliance Criteria

Ryan Kelly[1], PhD; Simon Jones[2], PhD; Blaine Price[3], BA, BSc, MSc; Dmitri Katz[3], PhD; Ciaran McCormick[3], MSc; Oliver Pearce[4], MD

[1]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[2]Department of Computer Science, University of Bath, Bath, United Kingdom

[3]School of Computing and Communications, The Open University, Milton Keynes, United Kingdom

[4]Trauma and Orthopaedics Department, Milton Keynes University Hospital, Milton Keynes, United Kingdom

Corresponding Author:
Blaine Price, BA, BSc, MSc
School of Computing and Communications
The Open University
Walton Hall
Milton Keynes, MK7 6AA
United Kingdom
Phone: 44 1908653701
Email: b.a.price@open.ac.uk

Abstract

Background: Physical activity trackers such as the Fitbit can allow clinicians to monitor the recovery of their patients following surgery. An important issue when analyzing activity tracker data is to determine patients' daily compliance with wearing their assigned device, using an appropriate criterion to determine a valid day of wear. However, it is currently unclear as to how different criteria can affect the reported compliance of patients recovering from ambulatory surgery. Investigating this issue can help to inform the use of activity data by revealing factors that may impact compliance calculations.

Objective: This study aimed to understand how using different criteria can affect the reported compliance with activity tracking in ambulatory surgery patients. It also aimed to investigate factors that explain variation between the outcomes of different compliance criteria.

Methods: A total of 62 patients who were scheduled to undergo total knee arthroplasty (TKA, ie, knee replacement) volunteered to wear a commercial Fitbit Zip activity tracker over an 8-week perioperative period. Patients were asked to wear the Fitbit Zip daily, beginning 2 weeks prior to their surgery and ending 6 weeks after surgery. Of the 62 patients who enrolled in the study, 20 provided Fitbit data and underwent successful surgery. The Fitbit data were analyzed using 5 different daily compliance criteria, which consider patients as compliant with daily tracking if they either register >0 steps in a day, register >500 steps in a day, register at least one step in 10 different hours of the day, register >0 steps in 3 distinct time windows, or register >0 steps in 3 out of 4 six-hour time windows. The criteria were compared in terms of compliance outcomes produced for each patient. Data were explored using heatmaps and line graphs. Linear mixed models were used to identify factors that lead to variation between compliance outcomes across the sample.

Results: The 5 compliance criteria produce different outcomes when applied to the patients' data, with an average 24% difference in reported compliance between the most lenient and strictest criteria. However, the extent to which each patient's reported compliance was impacted by different criteria was not uniform. Some individuals were relatively unaffected, whereas others varied by up to 72%. Wearing the activity tracker as a clip-on device, rather than on the wrist, was associated with greater differences between compliance outcomes at the individual level (P=.004, r=.616). This effect was statistically significant (P<.001) in the first 2 weeks after surgery. There was also a small but significant main effect of age on compliance in the first 2 weeks after surgery (P=.040). Gender and BMI were not associated with differences in individual compliance outcomes. Finally, the analysis revealed that surgery has an impact on patients' compliance, with noticeable reductions in activity following surgery. These reductions affect compliance calculations by discarding greater amounts of data under strict criteria.

Conclusions: This study suggests that different compliance criteria cannot be used interchangeably to analyze activity data provided by TKA patients. Surgery leads to a temporary reduction in patients' mobility, which affects their reported compliance when strict thresholds are used. Reductions in mobility suggest that the use of lenient compliance criteria, such as >0 steps or windowed approaches, can avoid unnecessary data exclusion over the perioperative period. Encouraging patients to wear the device at their wrist may improve data quality by increasing the likelihood of patients wearing their tracker and ensuring that activity is registered in the 2 weeks after surgery.

Trial Registration: ClinicalTrials.gov NCT03518866; https://clinicaltrials.gov/ct2/show/NCT03518866

(JMIR Mhealth Uhealth 2021;9(1):e22846) doi:10.2196/22846

KEYWORDS

activity tracking; adherence; compliance; surgery; total knee arthroplasty

Introduction

Background

Commercial activity trackers hold great promise for enabling clinicians to understand the physical activity of patients following surgery. Devices such as Fitbit or Apple Smart Watch, which typically present physical activity as a daily step count [1-3], have been shown to provide clinically viable data that are more accurate than patients' self-reports [4,5]. In light of these capabilities, researchers have explored how activity trackers can help to monitor patient recovery, particularly in cases involving ambulatory surgery [6,7], where early mobility is thought to result in better outcomes [8].

A necessary precursor to determining a patient's activity level is compliance analysis [9,10]. The principal aim of compliance analysis is to determine whether a patient provided reliable data on any given day, within a defined tracking period. What counts as reliable data is determined by setting a threshold for inclusion and discarding data that do not meet the threshold. As an example, a patient might be instructed to wear a tracking device every day for a minimum of 8 hours and would be considered compliant on days when the protocol was followed [9]. Any noncompliant days are then excluded from the data set, ensuring that only valid days are used in subsequent analyses [10]. Calculating each patient's compliance helps to ensure the validity, quality, and trustworthiness of activity tracking data [10-13]. Compliance analysis also supports reliable inferences based on the acquired data [14], including assessment of patient recovery [15,16] and activity levels [17].

A crucial step in compliance analysis involves selecting an appropriate criterion for data filtering. The literature harbors a range of criteria that differ in terms of how they define a valid day and hence filter data [10]. For example, quantity-based measures filter data based on absolute step counts [18], whereas time-based measures filter data based on criteria such as hours of device wear per day [19]. Compliance criteria also vary in their leniency. For instance, the criterion ">0 steps" considers a day as valid if at least one step is recorded within a 24-hour window [20]. Conversely, the criterion "≥10 hours" requires at least one step to be registered in 10 distinct hours over the day [21]. Problematically, the variety of available compliance criteria can leave researchers unsure which to use when working with activity data. It has also been shown that rates of data inclusion change when different compliance criteria are used to analyze an identical data set [10]. Strict criteria can often result in many otherwise valid days being discarded [10]. These issues motivate the need for researchers to investigate the impact of different compliance criteria when analyzing activity data.

In this work, we investigate the effect of applying different compliance criteria to activity data acquired from patients undergoing ambulatory surgery. We focus on total knee arthroplasty (TKA, ie, knee replacement) as it is a procedure that relates directly to movement and physical functioning [22]. TKA often results in pain and swelling [23], causing temporary reductions in mobility [24] that subside over time as functionality returns to the joint [6,7]. Tracking TKA patients' activity is therefore valuable to clinicians [25] and commercial systems are being developed that help clinicians to monitor patient recovery by collecting activity data [26]. However, analyzing compliance in TKA patients is not straightforward because these patients often have limited mobility following surgery. Selecting an overly strict compliance criterion to filter data may therefore result in the unnecessary exclusion of days, simply because a patient was undergoing recovery. Moreover, a patient could be compliant with device usage without registering steps due to reduced mobility, leaving ambiguity as to whether the lack of steps indicates that the device was not worn or whether there was no movement. These issues point toward a need to understand how different compliance criteria can affect the analysis of activity data from TKA patients; how compliance analyses should account for fluctuations in mobility over the perioperative period; and how researchers should select an appropriate criterion for analyzing activity data from TKA patients, such that recorded data are not unnecessarily discarded on the basis of a low step count.

Aims of This Study

This study aimed to understand the impact of applying 5 different compliance criteria to the activity tracking data of TKA patients. We further sought to investigate causes of variation in compliance outcomes through analysis of patients' tracked data.

Methods

Setting

This study was a field study in which TKA patients were asked to wear a commercial Fitbit Zip activity tracker (Fitbit, Inc.) to record their everyday physical activity, in the form of step count,

over an 8-week perioperative period. The perioperative period comprised the 2 weeks before the patient's surgery, the day of surgery, and the 6-week postoperative period (57 days in total). The study was conducted in the United Kingdom as part of a planned clinical trial (Trial Registration No. NCT03518866) that explored associations between early mobility and patient-reported outcomes following TKA (not reported here). The trial involved calculating each patient's compliance with wearing their assigned tracker. This task brought the effects of different compliance criteria to our attention and motivated this analysis.

Ethics and Recruitment

The study occurred over a 3-year period at the Open University in collaboration with Milton Keynes University Hospital (MKUH), a large public hospital in the United Kingdom. All procedures received approval from the Open University's Human Research Ethics committee (ID: HREC/2014/1635/Price/1), the R&D Department at MKUH, and the NHS Health Research Authority London (Surrey Borders Research Ethics Committee, Reference: 15/LO/0649).

Participants were recruited by the sixth author (OP), who identified patients from the hospital's elective orthopedic operating list. Inclusion criteria were as follows: adults undergoing TKA, ability to speak and understand conversational English, no cognitive impairments (eg, dementia), and no medical conditions unrelated to the individual's surgery that may affect pain levels or ability to participate (eg, severe neurological disorder, acute cancer, psychiatric disorder, or infections).

Patients who met the inclusion criteria were invited to participate during an outpatient appointment with their consultant. An information pack was provided for each patient to take away and read (Multimedia Appendix 1). The pack contained an overview of the study, a consent form, a prestudy questionnaire, and the contact details of the researchers. A member of the research team (BP or LS, see "Acknowledgments") followed up with the patient via telephone 1 week after their consultant appointment. Those who indicated willingness to participate were asked to bring the signed consent form and completed questionnaire with them to a presurgery clinic.

Participants

A total of 62 patients volunteered for the study; 12 participants were excluded from the study because their scheduled surgery was canceled. A further 30 individuals were removed, either due to technical issues related to the Fitbit or because they withdrew from the study.

The final sample comprised 20 participants (7 men and 13 women) who underwent ambulatory surgery and provided Fitbit data. The mean age of the participants was 64.5 years (SD 8.94; range 37-76). All had mean hospital stays of 2 days (range 1-4 days) following their surgery. Preoperative health was assessed using the ASA-PS (American Society of Anesthesiologists Physical Status) classification [27]. One patient was ASA 1, 18 patients were ASA 2, and 1 patient was ASA 3. This is broadly representative of typical comorbidities for a UK population requiring TKA, who may also have arthritis in other joints [28]. Medication use was not assessed as its heterogeneity was not thought to lend itself to analysis of physical activity following TKA.

Materials and Data Collection Setup

Participants were provided with a commercial Fitbit Zip activity tracker. This tracker was among the most popular commercial activity trackers when our study began, and was chosen due to its simplicity, high durability, and validity for capturing step data [29-31]. Furthermore, the Fitbit Zip has a 90-day battery life, avoiding the need for patients to charge the device (a known contributor to noncompliance [32]).

Participants who took part before January 2017 (N=8) were given the Fitbit Zip with a clip-on housing and were advised to wear it on their clothing (eg, belt loop), waist band, or brassiere strap. Some participants informally reported forgetting to unclip the device from their clothing, so the remaining participants (N=12) were given a plastic housing for the Zip that allowed them to wear the tracker around their wrist.

To avoid the need for participants to manage the synchronization of data collection from the Fitbit [19], a set of custom-made 'Fitboxes' was created by the research team. Each Fitbox was designed to be plugged into a network port on the participant's home router. The Fitbox was a custom 3D-printed plastic box containing a Fitbit wireless dongle attached to a Raspberry Pi computer. The computer ran a Python script that used the Galileo library [33] to capture data from the Fitbit without user intervention. The computer scanned for the patient's Fitbit every 10 minutes and synchronized the data to Fitbit's cloud server via Bluetooth. Data were then extracted from Fitbit's service for hosting on our own server.

Procedure

After each participant had consented to participate, a Fitbit Zip and Fitbox were delivered to their home by postal mail, arriving 2 weeks prior to their scheduled surgery. The participant was asked to plug the Fitbox into their internet router and begin wearing the Fitbit immediately.

Participants were requested to wear the Fitbit for the following 8 weeks, including their scheduled date of surgery. The data collection was periodically monitored by a member of the research team (BP) to check if there were any technical issues. In 2 cases a faulty Fitbox was replaced.

At the end of the 8-week period, participants were informed that they could stop using the Fitbit and were asked to bring the equipment to a scheduled appointment with their consultant. Participants were thanked and debriefed about the purpose of the research.

Data Analysis

Minute-by-minute step counts for each patient's perioperative period were acquired using the Fitbit API. The data were analyzed using scripts that applied the 5 different compliance criteria to each participant's data to determine daily compliance statistics. An average compliance outcome over the perioperative period was then derived for each criterion, and for each participant. The data were subsequently plotted using heatmaps

and line graphs to investigate periods of the day in which patients registered their steps. Statistical tests were performed using JASP to explore factors that explain differences between calculated rates of compliance. In this analysis, all data from each participant were included as our aim was to explore differences between rates of compliance under each criterion, and how these differences impact data retention.

Compliance Criteria

The 5 compliance criteria in our study are listed in Table 1. These criteria were chosen because they represent plausible approaches to assessing compliance in TKA patients, and because they have been used extensively in the literature on activity tracking [10].

As Table 1 illustrates, each criterion defines a different threshold for what counts as a valid day. The first 2, *>0 steps* and *>500 steps*, are quantity-based measures that assess compliance based on absolute thresholds for step count. To be considered compliant on a given day, a person must register either at least one step or more than 500 steps, respectively. The criterion *>0 steps* is very lenient and may therefore be useful for analyzing data from low-mobility populations, but can be problematic as it assumes there is no error in data collection. Activity trackers often register data from hand or arm movements [34], meaning that even the smallest incidental movement could be mistaken for a valid day of wear. Criteria like *>500 steps* address this problem by setting a higher minimum requirement for what counts as a valid day, though this may be too high to capture data from surgery patients during their recovery period. Patients may in fact be compliant with device use, while still taking fewer than 500 steps per day after surgery.

The next threshold, *≥10 hours*, requires a person to register at least one step in 10 different hours of the day. Previous work has noted that this criterion is often used in health informatics, but it is among the most stringent and can lead to high rates of data exclusion [10].

The final 2 criteria, *3-a-day* and *3-of-4 windows*, are time-based measures that consider a day as valid if the patient registers data in 3 predefined periods. In *3-a-day*, a patient must register at least one step in 3 windows anchored to the morning, afternoon, and evening [35]. In *3-of-4* windows, the day is broken down into 4 equal windows of 6 hours each, and a person is required to register data in at least three of them to be compliant [36]. Both of these criteria assess compliance based on continued wear over the course of a day. However, the *3-a-day* criterion contains windows of unequal size, which may be problematic when assessing the compliance of people who work unusual hours or rotated shift work. The *3-of-4* windows criterion addresses this limitation by allowing the patient to register data in a more flexible schedule.

Table 1. Definitions of compliance for the 5 criteria used in this study.

Compliance criterion	Definition	Example study
>0 steps	A day is considered valid if the tracker registered at least one step, ie, any data whatsoever.	Epstein et al [20]
>500 steps	A day is considered valid if the tracker registered more than 500 steps in a day.	Meyer et al [37]
≥10 hours	A day is considered valid if the tracker registered data in at least ten different 1-hour windows.	Sirard and Slater [21]
3-a-day	A day is considered valid if the tracker registered data in 3 predefined periods: 3 am to 11 am, 11 am to 3 pm, and 3 pm to 3 am.	Meyer et al [35]
3-of-4 windows	A day is considered valid if the tracker registered data in at least three of four periods: 12 am to 6 am, 6 am to 12 pm, 12 pm to 6 pm, and 6 pm to 12 am.	Barak et al [36]

Results

Analysis Overview

We first consider how the 5 criteria produce different compliance outcomes across the entire sample. We then consider how the 5 criteria affect rates of calculated compliance for individual patients. This allows us to interrogate whether the difference in compliance outcomes is consistent across the sample or whether certain patients are more affected by changing criteria. We then consider how variations between compliance calculations vary over the perioperative period, investigating the extent to which variation in compliance can be explained by demographic factors and whether the Fitbit was a clip-on or wrist-worn device.

Impacts of the Compliance Criteria Across the Sample

Table 2 shows the mean compliance for each criterion, averaged across the 8-week perioperative period for all 20 participants. Compliance can range from 0.00 to 1.00, with an outcome of 1.00 indicating 100% compliance.

Table 2 reveals that there is a difference in compliance outcomes between the criteria. For example, >0 steps gives a mean compliance of 0.80 (SD 0.17) across the sample, whereas ≥10 hours is lower at 0.56 (SD 0.25).

The histograms in Table 2 illustrate the proportion of the sample with each compliance level, showing that different criteria result in a different distribution of outcomes. For example, the distribution of compliance rates across all patients when using >0 steps is skewed toward 1.0, whereas the distribution for the ≥10 hours criterion is skewed toward 0.2. These outcomes reflect the nature of the measures, that is, >0 steps is the most lenient criterion whereas ≥10 hours is the strictest [10]. They also illustrate how different criteria lead to different rates of data exclusion, that is, a greater number of days are considered as invalid under the strictest criterion.

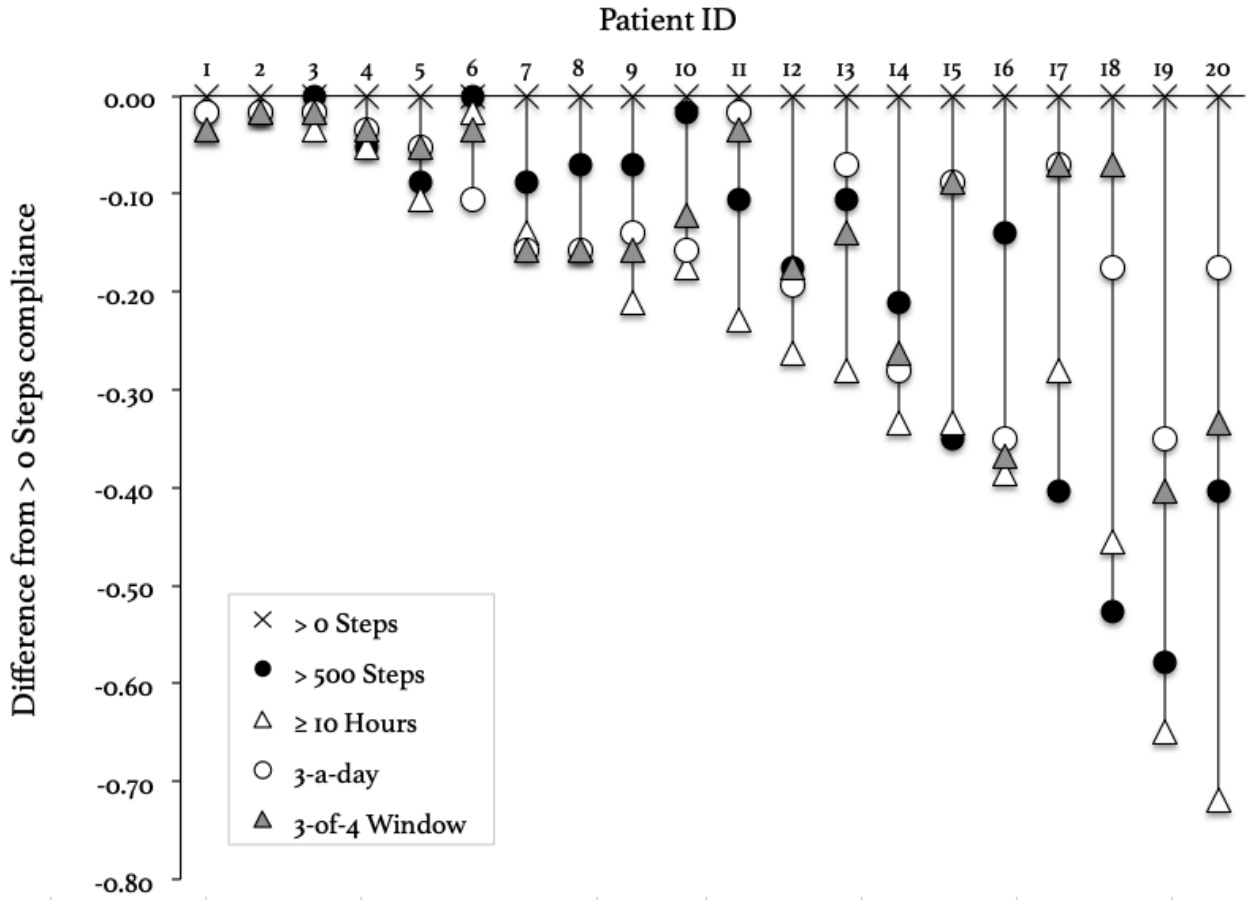
Table 2. Descriptive statistics for calculations of daily compliance with activity tracking across the patient sample, using each of the 5 compliance criteria. The histograms illustrate the proportion of patients with each compliance level under that criterion.

Compliance criterion	Mean (SD)	Minimum	Maximum	Range	Histogram
>0 steps	0.80 (0.17)	0.40	1.00	0.60	
>500 steps	0.63 (0.22)	0.26	1.00	0.74	
≥10 hours	0.56 (0.25)	0.02	0.96	0.94	
3-a-day	0.67 (0.22)	0.05	0.93	0.88	
3-of-4 windows	0.67 (0.22)	0.04	0.93	0.89	

Impacts of the Compliance Criteria Between Patients

Figure 1 provides a visual comparison of the compliance outcomes for each patient, showing the difference between >0 steps and the other 4 criteria. We use >0 steps as a baseline to illustrate the differences, as it produces the highest compliance outcome for all patients. In Figure 1, patients are listed according to the deviation between the compliance measures, ranging from smallest to greatest (left to right, respectively). Full data on each patient's outcomes, along with demographic information, can be found in Multimedia Appendix 2.

Figure 1. Differences in compliance calculations for the patients in our study. The scores are plotted against the most lenient criterion, >0 Steps, and show the difference between this criterion and the other measures.



We first consider differences between patients in terms of their calculated compliance with activity tracking, focusing on outcomes calculated *within* particular criteria. Overall, the data show that there are individual differences between patients in terms of calculated compliance. For example, 5 patients achieved 100% compliance under the >0 steps criterion, whereas patient 8 scored just 49%. Similarly, 3 patients had 98% or higher compliance under the >500 steps criterion, whereas others fared much worse. This reiterates evidence from the literature, that is, some people are highly compliant whereas others exhibit breaks or lapses in tracking [20].

We next consider differences between compliance criteria and how they change for each patient. A salient observation from Figure 1 is that the change produced from switching compliance criteria is not uniform across patients. Changing criteria can make minimal difference for some patients but has a substantial impact for others. For example, the outcomes for patients 1-4 are largely unaffected, whereas patient 20's compliance moves from 100% valid days under >0 steps to just 28% under ≥10 hours. This equates to a 72% difference between the 2 criteria, meaning that much of the patient's data would be excluded from further analysis under the stricter threshold.

Furthermore, we observed that changing criteria can *increase* reported compliance for some patients while *decreasing* reported compliance for others. This is illustrated by P16 and P17, and how their compliance changes when switching from >500 steps to ≥10 hours. P17 records a 12% *increase* in compliance when changing to ≥10 hours, whereas P16's compliance *decreases* by 24%.

The observed disparity in the direction of change begs the question of why different compliance criteria result in changes between patients, and why some patients are more affected by changes than others. From examining our data, we observed an apparent trend whereby patients with wrist-worn devices appeared to exhibit less variation between compliance outcomes (ie, lower SD), whereas those with clip-on devices exhibit greater variation across the 5 criteria.

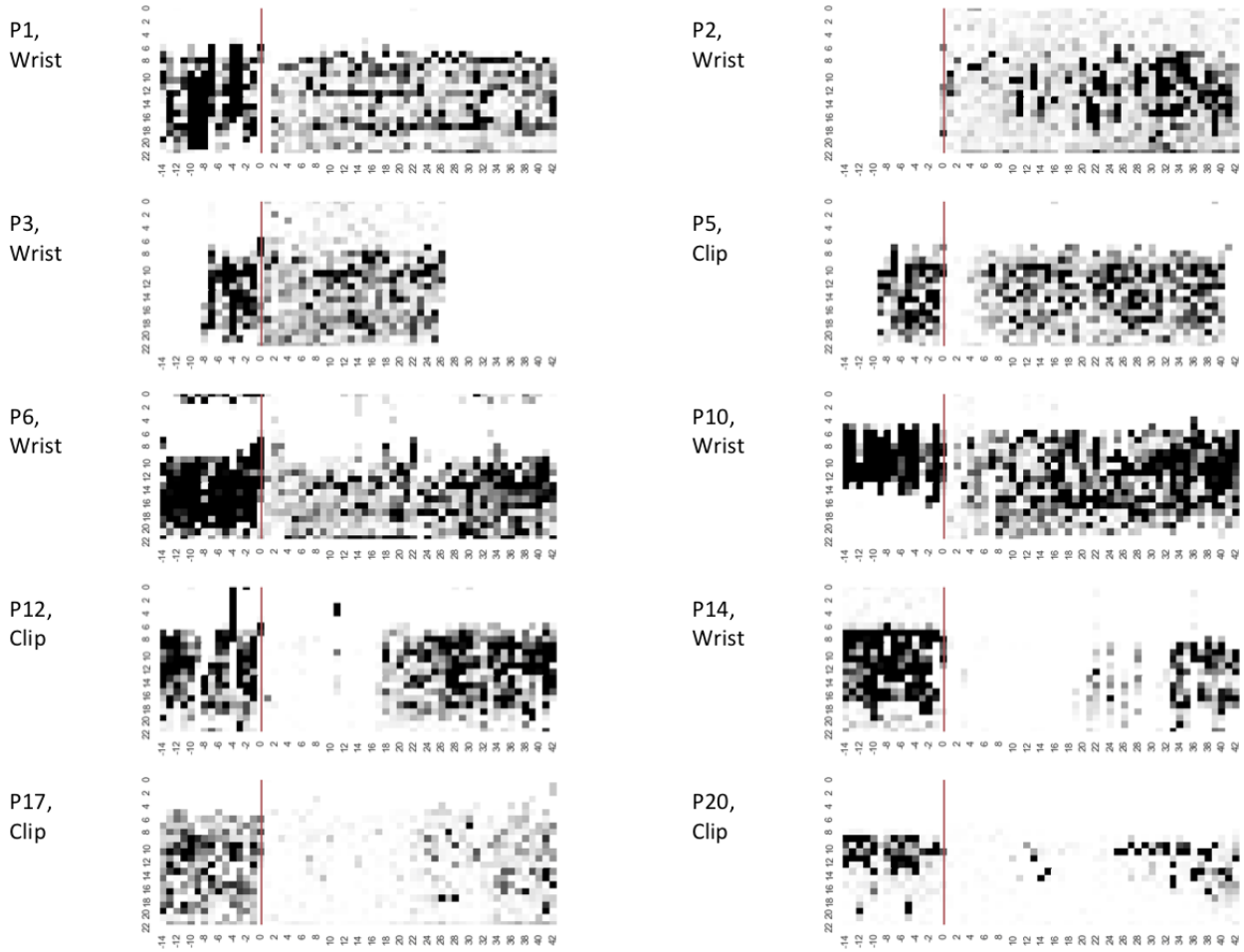
Statistical analysis of SD between compliance calculations using point biserial correlation revealed a statistically significant correlation (r=.616, P=.004). Hence, participants with a lower SD between compliance outcomes (ie, those toward the left side of Figure 1) were more likely to be using a wrist-worn Fitbit device than those with a higher SD (ie, those toward the right side of the graph, who were more likely to be wearing a clip-on device).

We also investigated whether age, BMI, and gender influenced compliance outcomes. Pearson correlation analysis revealed no significant relationship between age and SD of compliance outcomes (r=.178, P=.453). Likewise, Pearson correlation showed no significant relationship between BMI and SD (r=.147, P=.535). Point biserial correlation analysis revealed no significant relationship between gender and SD of compliance outcomes (r=.393, P=.086).

Physical Activity Patterns Throughout the Perioperative Period

To explore why the 5 criteria produce different compliance outcomes between patients, we generated heatmaps that visualize each patient's raw step count data across the perioperative period. Figure 2 shows heatmaps for 10 patients in our sample who exhibit different patterns of activity. We selected these patients to illustrate how different activity patterns produce changes under different compliance measures. Heatmaps for all 20 patients are included in Multimedia Appendix 3.

Figure 2. Heatmaps for 10 participants, illustrating the variations in patterns of physical activity throughout the day and across the 8-week perioperative period (from 2 weeks presurgery to 6 weeks postsurgery). The x-axis represents the day, beginning 2 weeks prior to surgery (day -14) and ending 6 weeks after surgery (day 42). The red line indicates the day of surgery (day 0). The y-axis represents the hour of the day. The color of cells corresponds to the number of steps recorded in a given hour, ranging from white (0) to black (500+).



The heatmaps resulted in 3 main observations.

First, there are general differences before and after surgery, with all of the participants exhibiting a noticeable change in their behavior after the red line, which indicates the day of surgery. Most patients (eg, P12, P14, and P17) exhibit a decline in the amount of data collected, whereas some others (such as P2) record more data. It is also notable that some patients have clear gaps in their data collection record. For example, P2 failed to record any data before the surgery, and P14 has many missing days following the surgery. These gaps are evidence of failing to wear the Fitbit and are considered as noncompliant days under all of the criteria in our analysis.

The second observation is that overall levels of recorded activity decrease after surgery among patients who appeared to be wearing their Fitbit during the preoperative period. For some patients this is illustrated by noticeable gaps in activity after surgery (eg, P5 and P12), whereas for others there is a simple reduction in the amount of data collected. Activity levels then typically increase over time, returning toward and sometimes exceeding a presurgery level, as illustrated by more dark cells appearing in the final weeks (eg, P2, P6, P12, and P17). This pattern is consistent with the gradual restoration of physical activity after TKA [6,7], but has implications for assessing compliance because the stricter criteria (eg, >500 steps and ≥10 hours) are likely to be insensitive to patients' reduced activity following surgery.

A third observation is that some patients' *distribution* of activity over the day remains the same following surgery, but for others it changes over time. For patients including P1, P5, P14, and P17, there is a continuation in the person's distribution of activity from before and after surgery, even though the difference represents an overall reduction in the level of activity. By comparison, P10's hours of activity end around 4 pm in the preoperative period but extend further into the evening after surgery.

The changes in activity level and distribution are important because they provide an explanation for why some patients may become gradually more compliant under certain criteria while remaining stable under others. In addition, the heatmaps suggest a clear impact of surgery on the patients' activity level, providing a potential explanation as to why some individuals are considered as less compliant under stricter measures.

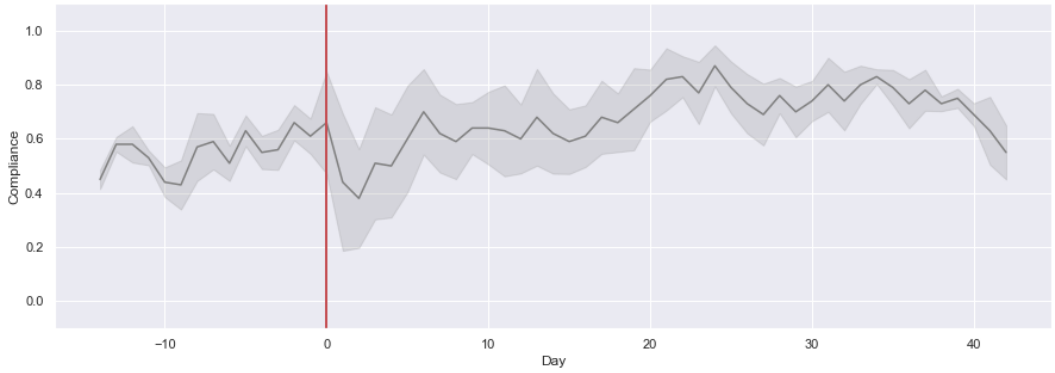
Variations in Compliance Criteria Agreement Throughout the Perioperative Period

Our observations of the heatmap data highlighted temporal variations in patients' recorded activity, with salient periods at which notable changes occur. The heatmaps also highlight

specific periods of interest for which physical activity is likely to be altered, that is, immediately after surgery and throughout the recovery period. Thus, we investigated temporal variations in compliance criteria agreement, that is, how closely the 5 outcomes align with one another over time, and how this alignment varies throughout the perioperative period. This allows us to further understand the impact of surgery on compliance agreement across the sample.

Figure 3 shows the daily compliance rates for all 5 criteria, with the red line marking the day of surgery. The data suggest there was a trend toward higher average compliance toward the end of the perioperative period. There is a noticeable drop-off in mean compliance immediately following surgery, and a slight decrease around the 40-day mark.

Figure 3. Daily compliance rates over time. The percentage of the sample that were compliant on any given day is illustrated by the y-axis. The dark gray line represents the mean average across all 5 of the compliance criteria. The height of the gray shaded region indicates the standard deviation between measures over time. The red vertical line represents the day of surgery.



In terms of agreement between the criteria, it can be seen that there is very little difference between measures in the presurgery phase, as reflected by the narrower shaded region for SD. After the surgery there is a 25% drop in the average compliance, with a corresponding increase in SD between criteria. Over the next 3 weeks, the mean increases while the SD decreases, as indicated by the shrinking height of the shaded area. Toward the end of the perioperative period, in the final 2 weeks, there is a dip in the average compliance but the deviation between measures is small and stable.

Figure 4 shows specifically how patients' compliance appears to vary over time when applying each of the criteria. There are several important observations from this figure. The first is that compliance under the >0 steps criterion is generally higher than the other criteria following surgery, with a very noticeable difference for the first 20 days of the recovery period. This means that there are many instances in which patients were registering at least some steps over the day, but the total number was often insufficient to be considered compliant on stricter thresholds.

Figure 4. Comparison of daily compliance rates for all 5 compliance criteria.

The second main observation from Figure 4 is that the >500 steps, ≥10 hours, 3-a-day, and the 3-of-4 window thresholds appear to follow each other reasonably closely over the entire 8-week period. However, these 4 thresholds cause a large drop in calculated compliance after surgery, and then show a gradual increase over time. The implication of this is that the patient is registering a small number of steps in the days after his/her surgery, but those steps are not sufficient to be considered compliant under stricter activity thresholds.

A final observation is that there is a slight increase in compliance at the point of surgery under the >0 steps criterion. This likely indicates fewer cases of participants forgetting to wear the Fitbit, and could be caused by hospital staff reminding participants to wear the device, or by the fact that TKA may make the idea of

monitoring physical activity more salient. Alternatively, surgery presents a change to routine and thus there may be fewer daily activities that are distracting from wearing the tracker.

Taken together, Figures 3 and 4 illustrate the plausible impact of temporary impairment on participants' compliance under different criteria. Many days appear to be considered as valid under the most lenient criterion (>0 steps) but would be discarded under the other measures.

Analysis of Mean Calculated Compliance

To explore the preceding observations quantitatively, we subdivided the perioperative period into distinct 2-week stages (excluding the day of surgery) and calculated 2-week compliance figures for each patient, using the 5 criteria. This allowed us to explore whether there are differences between the 2-week stages (defined as *presurgery*, *weeks 0-2*, *weeks 2-4*, and *weeks 4-6*). We then used linear mixed models to investigate changes in calculated compliance between the 5 criteria, using demographic information as independent variables. Each patient's average compliance for the 2-week stages is shown in Multimedia Appendix 4.

Specifically, our model compared the effects of *compliance criteria*, *gender*, *age*, *BMI*, *stage*, and *device type* (*clip or wrist-worn tracker*) on *2-week compliance*. ASA was excluded because there was insufficient variance in scores between participants. In each analysis the independent variables were set as fixed effects. Participant ID was set as a random effect. The reported β estimates indicate the value by which 2-week compliance varies for the respective condition. We report the standard error (SE) for the estimates and test for significance at α=.05.

We found a significant main effect of *compliance criteria* on *2-week compliance* ($F_{4,22.51}$=16.816, *P*<.001). Compliance scores were significantly higher (approximately 16%) for >0 steps (β=.157, SE=0.022, *P*<.001). Compliance scores were significantly lower for >500 steps (β=–.041, SE=0.017, *P*=.027), ≥10 hours (β=–.104, SE=0.019, *P*<.001), and 3-a-day (β=–.031, SE=0.014, *P*=.045).

We also found a significant main effect of *stage* on *2-week compliance* ($F_{3,18}$=4.166, *P*=.021). Compliance scores were significantly higher (approximately 10%) in the week 4-6 postsurgery *stage* (β=.102, SE=0.041, *P*=.024). This matches our observations of Figures 3 and 4, and indicates that patients were gradually regaining mobility and recording more data as the perioperative period progressed.

We also found a small but significant main effect of *age* on *2-week compliance* ($F_{1,15.48}$=5.004, β=.009, SE=0.004, *P*=.040). Specifically, older patients were more likely to have higher compliance, regardless of criteria. This may be because health issues are more salient for older adults [38], potentially increasing their compliance with using an activity tracker.

Finally, we found significant interaction effects between *criteria* and *stage* ($F_{12,245.89}$=5.631, *P*<.001). Compliance scores were significantly higher for the >0 steps criterion in weeks 0-2 postsurgery (β=.115 , SE=0.019, *P*<.001). Compliance scores were significantly lower for >500 steps (β=–.067, SE=0.019, *P*<.001) and 10 hours (β=–.054 , SE=0.019, *P*=.005) in weeks 0-2 postsurgery. Aligning with Figures 3 and 4, this outcome demonstrates that the weeks following surgery are associated with reductions in compliance under stricter criteria.

Variations Over Time by Device Type

Based on our earlier finding that deviation between measures corresponds to device type (wrist vs clip), we examined the temporal variations for each device independently. Figure 5A shows the proportion of the clip-worn group that was compliant on each day. It can be seen that there is close agreement between the criteria during the presurgery phase. At the point of surgery there is a noticeable increase in deviation between criteria, which gradually decreases over time.

Figure 5. Temporal variations in compliance for the 5 criteria over the perioperative period. (A) The proportion of the clip-worn group that was compliant on each day. (B) The variation between criteria for the clip-worn group. (C) The proportion of the wrist-worn group that was compliant on each day. (D) The variation between criteria for the wrist-worn group.

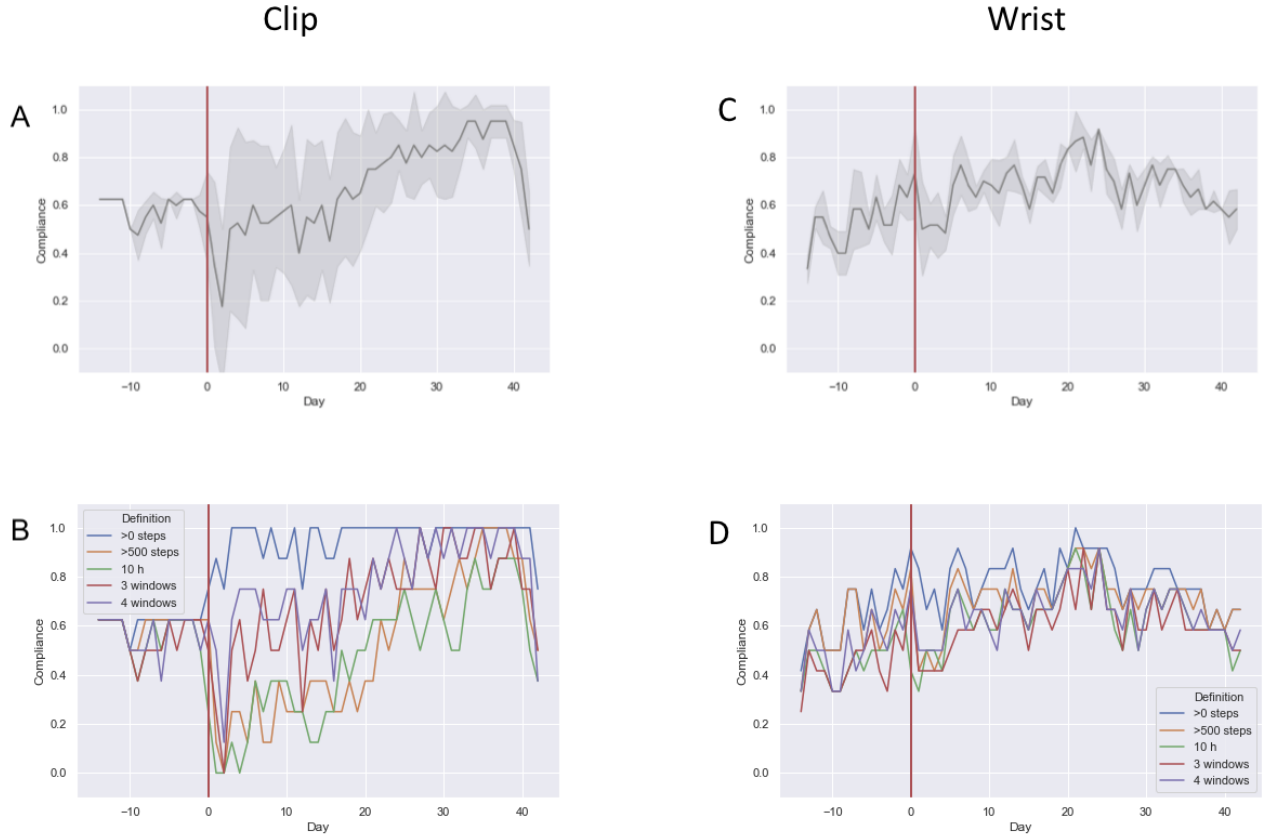


Figure 5B again shows that compliance is particularly sensitive to criterion selection during the first 3-4 weeks of the postoperative period, with >0 steps close to 100% throughout. The >500 steps and 3-of-4 window outcomes are closely aligned during this phase, as are ≥10 hours and 3-a-day. However, the overall range is very large.

Figure 5C shows the proportion of the wrist-worn group that was compliant on each day. Contrasting Figure 5A (clip group) with Figure 5C (wrist group), we see that the wrist-worn group exhibit far lower deviation between measures throughout the entire period, and only a small increase in deviation immediately following surgery. For the wrist group, the deviation level remains consistently low from approximately 2 weeks postsurgery. Figure 5D also shows close agreement between criteria. This reveals that the deviation between measures is a more prominent issue for the clip device group, but that the issue only becomes apparent at the point of surgery.

To investigate these observations quantitatively, we compared the effects of *gender*, *age*, *BMI*, *stage*, and *device type* (*clip or wrist-worn tracker*) on the SD between compliance outcomes using linear mixed models. We again separated the perioperative period into 4 distinct 2-week stages. (The SDs for each patient in the 4 stages are shown in Multimedia Appendix 4.) ASA was again excluded. The independent variables were set up as fixed effects and ID was set as a random effect. The β estimates indicate the value by which SD between compliance outcomes changes.

Overall, we found a significant main effect of *stage* on SD between compliance outcomes, $F_{3,54}$=11.421, P<.001. Deviation between outcomes was significantly higher (approximately 8%) in the first 2 weeks postsurgery (β=.082, SE=0.016, P<.001).

We also found a significant interaction effect between *device type* and *stage* ($F_{3,54}$=9.976, P<.001). For wearers of the clip-on device, deviation between measures was significantly higher (approximately 6%) in the 2 weeks immediately following surgery (β=.064, SE=0.016, P<.001). These outcomes provide further evidence of how surgery appears to impact patients' compliance, with reductions in activity in weeks 0-2 causing some patients to miss the requirements of stricter thresholds, resulting in a greater quantity of data being discarded. This effect is especially pronounced among those who wore the tracker as a clip-on device (Figure 5B).

Discussion

Principal Findings

The first finding of this study was that the 5 compliance criteria provide different compliance outcomes when applied to the patients' data. We found a 24% difference in average rates of data retention between the most lenient criterion (>0 steps) and the most stringent (≥10 hours), with other measures falling between these 2 points. This finding illustrates that different compliance criteria cannot be used interchangeably to analyze data from TKA patients, and dovetails with the work of Tang et al [10], who observed similar effects after applying 4 criteria

to 9 distinct activity tracking data sets. A unique feature of our work is the examination of TKA patients, who are experiencing temporary impairment to their mobility due to surgery. Our study reinforces the need to carefully consider an appropriate criterion for assessing compliance and shows that this is especially crucial when working with a surgery population, given impairments to their mobility.

Our second main finding was that differences in compliance outcomes were not uniform across the sample. The data of some patients were largely unaffected by the use of different criteria, whereas others varied considerably. The most extreme case resulted in 72% of a patient's data being excluded when switching between 2 different criteria. Such a change would represent a clinically significant difference in the patient's recorded activity level [17,39]. Our investigation of potential causes revealed that patients who wore the Fitbit as a clip-on had greater variation between compliance outcomes, compared with those who wore the device on their wrist. One explanation for this finding may be related to how the activity tracker is treated once it is in the home. Anecdotally, our patients mentioned attaching the clip-on Fitbit to their clothes and forgetting to remove it at the end of the day. Using wrist-worn trackers may therefore result in higher quality data because they are less likely to be forgotten by patients and because a wrist-worn device does not need to be removed as often (eg, while asleep).

The third main contribution from our study lies in demonstrating the impact of surgery on calculations of compliance. We found that patients exhibit changes in tracking behavior and fluctuations in physical activity after surgery. Statistical analyses also showed that the deviation between compliance outcomes was significantly higher (P<.001) in the 2 weeks following surgery. Specifically, compliance was calculated as significantly higher (P<.001) when using >0 steps, which is also the most lenient criterion. The implication of these findings is that analyses of compliance need to account for reductions in physical mobility following a surgical procedure. Stricter measures of compliance may exclude data on the basis of low activity, even though this activity may in fact be representative of the patient's capabilities following surgery. Although the >0 steps criterion is vulnerable to recording incidental step data from arm movements [34], the presence of these "steps" may help to show the tracker was worn and may be preferable to discarding days entirely. This consideration is important for clinicians who plan to use physical activity data from TKA patients, and for systems designed to support decisions based on its use.

Implications for Activity Tracking Studies

On the basis of our study, we suggest that compliance analysis among TKA patients should begin by using the >0 steps criterion as an initial data filter. This enables identification of days on which the tracker was worn, allowing noncompliant days to be removed from calculations of step count. Next, a windowed approach that has a minimal activity target, such as 3-of-4 windows [36], should be used to determine whether the patient recorded data across the day.

Regarding methodology, we echo Tang et al [10] and encourage future researchers to clearly describe the approach taken when calculating and reporting compliance. Researchers should also state the criterion used to determine a valid day and consider how other criteria may affect the data. Failure to describe the compliance measure may undermine the perceived validity of research using activity tracking data.

Finally, future studies should consider using secondary data sources to acquire ground truth about whether a patient was wearing an activity tracker. Many activity trackers now include heart rate monitors, and although these do not guarantee improvements in compliance [40], the additional data source could be triangulated with step counts to determine wear time. A limitation of these devices, however, is that they have a shorter battery life than trackers such as the Fitbit Zip, and hence may impact compliance because patients need to recharge the device [32]. Ecological momentary assessment [41] techniques, delivered via a participant's smartphone, could also be used to collect self-reports about whether a tracker is being worn.

Design Implications

The results of this study can inform presurgery and postsurgery monitoring systems (eg, [26]) that support clinical decision making on the basis of activity data. Our study showed that the underlying threshold used to calculate compliance can affect rates of data retention. Decisions made on the basis of these data may be flawed if the system excludes a large proportion of valid days due to notionally low activity. We suggest that systems should show the exact number of steps recorded and how these steps were distributed over the day, enabling clinicians to assess compliance at the absolute and temporal levels. Noncompliant days should be illustrated as gaps in the patient's record, which may still be useful to support decision making [14] and encourage patients to increase their compliance with tracker use.

Clinical Implications

Based on this study, clinical trials involving activity trackers must account for the fact that stringent compliance criteria may lead to the exclusion of potentially useful data from TKA patients. Selecting a lenient criterion can address this concern. Similarly, clinical monitoring systems (eg, [26]) will present unreliable data without an appropriate compliance measure.

Researchers should also be aware that patients undergoing TKA have a relatively high dropout rate in the context of use of wearables. Our study was conducted in the UK, where the waiting times for surgery mean there is a substantial time gap from being listed for surgery and the surgery itself. Trials should recognize the potential for patient forgetfulness if the activity trackers are handed out early in the patient journey. Likewise, trials should recognize that the postoperative period is often painful, with the consequent possibility for patients to be distracted from using the wearable as it may not be high on their list of priorities.

Limitations

The number of analyzed participants is a limitation of this study. Our sample of 20 patients was sufficient to investigate

differences and illustrate patterns but needs to be verified with larger sample sizes. A cohort of over 100 patients is likely to be needed to draw clinical conclusions about correlations between early activity, compliance, and other factors such as pain, analgesia use, and late outcomes. Knee replacement patients are known to have an approximate 20% dissatisfaction rate past 1 year [23]. Early activity may be a predictor in this likely multifactorial problem, and future work should seek higher statistical power to shine a light on this.

Another limitation is that, because of the open invitation for participation in the study, patients with greater concern about physical activity may have accepted the invitation. Rates of compliance with device use may be different among those who are less concerned about physical activity.

Lastly, our study was not designed to investigate differences between wearing the tracker on different parts of the body. This means that participants were not randomly assigned to wear the tracker in particular places, and thus we cannot completely rule out other latent factors as explanations for compliance differences between wrist-worn and clip-on trackers. Possible influences include gradual improvements in administering the study protocol, or changes in staff, at the collaborating hospital. Our analysis led us to explore demographic factors and the placement of the Fitbit as these were available to us based on the data we collected. Other factors that may impact patients' behavior (eg, smoking status) should be included in future work.

Conclusions

This study aimed to understand how different criteria affect calculated compliance with activity tracking in TKA patients. Our findings suggest that different compliance criteria cannot be used interchangeably to analyze activity data provided by patients following ambulatory surgery. Instead, reductions in postsurgery mobility necessitate the use of lenient compliance criteria, such as >0 steps, combined with a windowed approach. These criteria can account for temporary mobility impairments while also tracking wear over the course of a day. Encouraging patients to wear the device at their wrist, and using secondary sources of data as ground truth, can increase confidence in compliance outcomes by ensuring that activity is detected and by increasing patients' actual wear time.

Acknowledgments

This research was part-funded by UK EPSRC grants EP/L021285/1 and EP/P01013X/1 and ERC grant 291652. We thank Lesley Simpson, Carole Jellicoe, Erica Witherall, Aran Nanthakumar, and Raaid Ismail for their assistance with data collection.

Authors' Contributions

BP and OP conceived the original study. BP, OP, and CM collected data. RK, SJ, and CM conducted data analysis. RK, BP, and SJ conceived and drafted the paper. RK, SJ, and DK revised the paper in response to reviewers' comments. All authors provided input to the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant Information Sheet.
[PDF File (Adobe PDF File), 132 KB - mhealth_v9i1e22846_app1.pdf]

Multimedia Appendix 2
Table 3. Comparison of five compliance criteria applied to the activity tracking data of 20 different patients. A person is considered to be compliant if they meet the criterion on a given day. The calculation is provided as a percentage of compliant days over the 8-week perioperative period.
[DOCX File , 16 KB - mhealth_v9i1e22846_app2.docx]

Multimedia Appendix 3
Heatmaps for all 20 Participants.
[DOCX File , 245 KB - mhealth_v9i1e22846_app3.docx]

Multimedia Appendix 4
Table 4: Mean and Std Dev comparison.
[DOCX File , 14 KB - mhealth_v9i1e22846_app4.docx]

References

1. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
2. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S582-S588. [Medline: 16294121]
3. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of ubifit garden. New York, NY: Association for Computing Machinery; 2008 Apr Presented at: CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 2008; Florence, Italy p. 1797-1806. [doi: 10.1145/1357054.1357335]
4. Verlaan L, Bolink SAAN, Van Laarhoven SN, Lipperts M, Heyligers IC, Grimm B, et al. Accelerometer-based Physical Activity Monitoring in Patients with Knee Osteoarthritis: Objective and Ambulatory Assessment of Actual Physical Activity During Daily Life Circumstances. Open Biomed Eng J 2015;9:157-163 [FREE Full text] [doi: 10.2174/1874120701509010157] [Medline: 26312077]
5. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016 Jul;48:36-41 [FREE Full text] [doi: 10.1016/j.gaitpost.2016.04.025] [Medline: 27477705]
6. Twiggs J, Salmon L, Kolos E, Bogue E, Miles B, Roe J. Measurement of physical activity in the pre- and early post-operative period after total knee arthroplasty for Osteoarthritis using a Fitbit Flex device. Med Eng Phys 2018 Jan;51:31-40. [doi: 10.1016/j.medengphy.2017.10.007] [Medline: 29117912]
7. Van der Walt N, Salmon LJ, Gooden B, Lyons MC, O'Sullivan M, Martina K, et al. Feedback From Activity Trackers Improves Daily Step Count After Knee and Hip Arthroplasty: A Randomized Controlled Trial. J Arthroplasty 2018 Nov;33(11):3422-3428. [doi: 10.1016/j.arth.2018.06.024] [Medline: 30017217]
8. Mahmood A, Barklie L, Pearce O. Ambulatory Surgery. Early outpatient pain scores in hip and knee arthroplasty. Could these be early predictors of painful joint replacements? 2017 Jan;23.1:19-21 [FREE Full text]
9. Tang LM, Day M, Engelen L, Poronnik P, Bauman A, Kay J. Daily & Hourly Adherence: Towards Understanding Activity Tracker Accuracy. New York, NY: Association for Computing Machinery; 2016 May Presented at: CHI EA '16: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 2016; San Jose, CA p. 3211-3218. [doi: 10.1145/2851581.2892438]
10. Tang LM, Meyer J, Epstein DA, Bragg K, Engelen L, Bauman A, et al. Defining Adherence: Making Sense of Physical Activity Tracker Data. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-22. [doi: 10.1145/3191769]
11. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information Quality Challenges of Patient-Generated Data in Clinical Practice. Front Public Health 2017;5:284 [FREE Full text] [doi: 10.3389/fpubh.2017.00284] [Medline: 29209601]
12. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S531-S543. [Medline: 16294116]
13. Tudor-Locke C, Burkett L, Reis JP, Ainsworth BE, Macera CA, Wilson DK. How many days of pedometer monitoring predict weekly physical activity in adults? Prev Med 2005 Mar;40(3):293-298. [doi: 10.1016/j.ypmed.2004.06.003] [Medline: 15533542]
14. Tang LM, Kay J. Harnessing Long Term Physical Activity Data—How Long-term Trackers Use Data and How an Adherence-based Interface Supports New Insights. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2017 Jun 30;1(2):1-28. [doi: 10.1145/3090091]
15. Twiggs J, Roe J, Salmon JL, Miles B, Theodore W. Patient-specific activity level benchmarks during recovery for total knee arthroplasty. Orthopaedic Proceedings 2017;99-B(Suppl_6):55 [FREE Full text]
16. Price BA, Kelly R, Mehta V, McCormick C, Ahmed H, Pearce O. Feel My Pain: Design and Evaluation of Painpad, a Tangible Device for Supporting Inpatient Self-Logging of Pain. New York: ACM; 2018 Apr Presented at: 2018 CHI Conference on Human Factors in Computing Systems (CHI '18); 2018; Montreal URL: https://doi.org/10.1145/3173574.3173743 [doi: 10.1145/3173574.3173743]
17. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
18. Chastin SFM, Dall PM, Tigbe WW, Grant MP, Ryan CG, Rafferty D, et al. Compliance with physical activity guidelines in a group of UK-based postal workers using an objective monitoring technique. Eur J Appl Physiol 2009 Aug;106(6):893-899. [doi: 10.1007/s00421-009-1090-x] [Medline: 19488779]
19. Purta R, Mattingly S, Song L, Lizardo O, Hachen D, Poellabauer C, et al. Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits. New York: ACM; 2016 Sep Presented at: 2016 ACM International Symposium on Wearable Computers (ISWC '16); 2016; Heidelberg URL: https://doi.org/10.1145/2971763.2971767 [doi: 10.1145/2971763.2971767]
20. Epstein DA, Kang JH, Pina LR, Fogarty J, Munson SA. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. New York: ACM; 2016 Sep Presented at: 016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16); 2016; Heidelberg URL: https://doi.org/10.1145/2971648.2971656 [doi: 10.1145/2971648.2971656]

21. Sirard JR, Slater ME. Compliance with wearing physical activity accelerometers in high school students. J Phys Act Health 2009;6 Suppl 1:S148-S155 [FREE Full text] [doi: 10.1123/jpah.6.s1.s148] [Medline: 19998861]
22. Webber SC, Strachan SM, Pachu NS. Sedentary Behavior, Cadence, and Physical Activity Outcomes after Knee Arthroplasty. Med Sci Sports Exerc 2017 Jun;49(6):1057-1065. [doi: 10.1249/MSS.0000000000001207] [Medline: 28099297]
23. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient Dissatisfaction Following Total Knee Arthroplasty: A Systematic Review of the Literature. J Arthroplasty 2017 Dec;32(12):3854-3860. [doi: 10.1016/j.arth.2017.07.021] [Medline: 28844632]
24. Carr AJ, Robertsson O, Graves S, Price AJ, Arden NK, Judge A, et al. Knee replacement. Lancet 2012 Apr 07;379(9823):1331-1340. [doi: 10.1016/S0140-6736(11)60752-6] [Medline: 22398175]
25. Appelboom G, Yang AH, Christophe BR, Bruce EM, Slomian J, Bruyère O, et al. The promise of wearable activity sensors to define patient recovery. J Clin Neurosci 2014 Jul;21(7):1089-1093. [doi: 10.1016/j.jocn.2013.12.003] [Medline: 24534628]
26. MyMobility Website.: Zimmer Biomet; 2020. URL: https://www.zimmerbiomet.com/medical-professionals/zb-connect/mymobility.html [accessed 2021-01-15]
27. Doyle DJ, Garmon EH. American society of anesthesiologists classification (ASA class).: StatPearls Publishing; 2020 Jul 04. URL: https://www.ncbi.nlm.nih.gov/books/NBK441940/ [accessed 2021-01-15]
28. Whitehouse SL, Blom AW, Taylor AH, Pattison GTR, Bannister GC. The Oxford Knee Score; problems and pitfalls. Knee 2005 Aug;12(4):287-291. [doi: 10.1016/j.knee.2004.11.005] [Medline: 15993604]
29. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
30. Hermsen S, Moons J, Kerkhof P, Wiekens C, De GM. Determinants for Sustained Use of an Activity Tracker: Observational Study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
31. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
32. Bielek P, Tomlein A, Kratky P, Mitrik S, Barla M, Bielikova M. Move2Play: an innovative approach to encouraging people to be more physically active. New York: ACM; 2012 Jan Presented at: 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12); 2012; Miami URL: https://doi.org/10.1145/2110363.2110374 [doi: 10.1145/2110363.2110374]
33. Allard B. Galileo Python Library. 2016 Sep 18. URL: https://pypi.org/project/galileo/ [accessed 2021-01-15]
34. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106 [FREE Full text] [doi: 10.2196/mhealth.6321] [Medline: 28801304]
35. Meyer J, Wasmann M, Heuten W, El Ali A, Boll S. Identification and Classification of Usage Patterns in Long-Term Activity Tracking. New York: ACM; 2017 May Presented at: 2017 CHI Conference on Human Factors in Computing Systems (CHI'17); 2017; Denver URL: https://doi.org/10.1145/3025453.3025690 [doi: 10.1145/3025453.3025690]
36. Barak S, Wu SS, Dai Y, Duncan PW, Behrman AL, Locomotor Experience Applied Post-Stroke (LEAPS) Investigative Team. Adherence to accelerometry measurement of community ambulation poststroke. Phys Ther 2014 Jan;94(1):101-110 [FREE Full text] [doi: 10.2522/ptj.20120473] [Medline: 24029299]
37. Meyer J, Schnauber J, Heuten W, Wienbergen H, Hambrecht R, Appelrath HJ, et al. Exploring Longitudinal Use of Activity Trackers. : IEEE; 2016 Oct Presented at: 2016 IEEE International Conference on Healthcare Informatics (ICHI); 2016; Chicago p. 198-206 URL: http://doi.org/10.1109/ICHI.2016.29 [doi: 10.1109/ichi.2016.29]
38. Drewnowski E, Monsen E, Birkett D, Gunther S, Vendeland S, Su J, et al. Health Screening and Health Promotion Programs for the Elderly. Disease Management & Health Outcomes 2003;11(5):299-309 [FREE Full text] [doi: 10.2165/00115677-200311050-00003]
39. Boyer KA, Kiratli BJ, Andriacchi TP, Beaupre GS. Maintaining femoral bone density in adults: how many steps per day are enough? Osteoporos Int 2011 Dec;22(12):2981-2988. [doi: 10.1007/s00198-011-1538-9] [Medline: 21318440]
40. Faust L, Purta R, Hachen D, Striegel A, Poellabauer C, Lizardo O, et al. Exploring Compliance: Observations from a Large Scale Fitbit Study. New York: ACM; 2017 Apr Presented at: 2nd International Workshop on Social Sensing; 2017; Pittsburgh p. 55-60 URL: https://doi.org/10.1145/3055601.3055608 [doi: 10.1145/3055601.3055608]
41. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]

Abbreviations

MKUH: Milton Keynes University Hospital
SE: standard error
TKA: total knee arthroplasty

Edited by L Buis; submitted 24.07.20; peer-reviewed by M Asgari Mehrabadi, AS Poncette, I Bouloukaki; comments to author 17.08.20; revised version received 12.10.20; accepted 06.11.20; published 26.01.21.

Please cite as:
Kelly R, Jones S, Price B, Katz D, McCormick C, Pearce O
Measuring Daily Compliance With Physical Activity Tracking in Ambulatory Surgery Patients: Comparative Analysis of Five Compliance Criteria
JMIR Mhealth Uhealth 2021;9(1):e22846
URL: http://mhealth.jmir.org/2021/1/e22846/
doi:10.2196/22846
PMID:33496677

Original Paper

Smartphone App (2kmFIT-App) for Measuring Cardiorespiratory Fitness: Validity and Reliability Study

Adria Muntaner-Mas[1,2], PhD; Antonio Martinez-Nicolas[3,4], PhD; Alberto Quesada[2], BSc; Cristina Cadenas-Sanchez[2], PhD; Francisco B Ortega[2,5], PhD

[1]Department of Physical Education and Sports, Faculty of Education, University of Balearic Islands, Palma, Spain

[2]PROmoting FITness and Health Through Physical Activity Research Group, Department of Physical Education and Sports, Faculty of Sport Sciences, University of Granada, Granada, Spain

[3]Chronobiology Research Group, Department of Physiology, Faculty of Biology, University of Murcia, Murcia, Spain

[4]Ciber Fragilidad y Envejecimiento Saludable, Madrid, Spain

[5]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Adria Muntaner-Mas, PhD
Department of Physical Education and Sports
Faculty of Education
University of Balearic Islands
Valldemossa Road, km 7.5
Palma
Spain
Phone: 34 971172875
Email: adria.muntaner@uib.es

Abstract

Background: There is strong evidence suggesting that higher levels of cardiorespiratory fitness (CRF) are associated with a healthier metabolic profile, and that CRF can serve as a powerful predictor of morbidity and mortality. In this context, a smartphone app based on the 2-km walk test (UKK test) would provide the possibility to assess CRF remotely in individuals geographically distributed around a country or continent, and even between continents, with minimal equipment and low costs.

Objective: The overall aim of this study was to evaluate the validity and reliability of 2kmFIT-App developed for Android and iOS mobile operating systems to estimate maximum oxygen consumption (VO2max) as an indicator of CRF. The specific aims of the study were to determine the validity of 2kmFIT-App to track distance and calculate heart rate (HR).

Methods: Twenty participants were included for field-testing validation and reliability analysis. The participants completed the UKK test twice using 2kmFIT-App. Distance and HR were measured with the app as well as with accurate methods, and VO2max was estimated using the UKK test equation.

Results: The validity results showed the following mean differences (app minus criterion): distance (–70.40, SD 51.47 meters), time (–0.59, SD 0.45 minutes), HR (–16.75, SD 9.96 beats/minute), and VO2max (3.59, SD 2.01 ml/kg/min). There was moderate validity found for HR (intraclass correlation coefficient [ICC] 0.731, 95% CI –0.211 to 0.942) and good validity found for VO2max (ICC 0.878, 95% CI –0.125 to 0.972). The reliability results showed the following mean differences (retest minus test): app distance (25.99, SD 43.21 meters), app time (–0.15, SD 0.94 seconds), pace (–0.18, SD 0.33 min/km), app HR (–4.5, 13.44 beats/minute), and app VO2max (0.92, SD 3.04 ml/kg/min). There was good reliability for app HR (ICC 0.897, 95% CI 0.742-0.959) and excellent validity for app VO2max (ICC 0.932, 95% CI 0.830-0.973). All of these findings were observed when using the app with an Android operating system, whereas validity was poor when the app was used with iOS.

Conclusions: This study shows that 2kmFIT-App is a new, scientifically valid and reliable tool able to objectively and remotely estimate CRF, HR, and distance with an Android but not iOS mobile operating system. However, certain limitations such as the time required by 2kmFIT-App to calculate HR or the temperature environment should be considered when using the app.

(JMIR Mhealth Uhealth 2021;9(1):e14864) doi:10.2196/14864

KEYWORDS

exercise test; mobile apps; reproducibility of results; physical fitness; telemedicine; cardiorespiratory fitness

Introduction

Cardiorespiratory fitness (CRF) indicates the global cardiovascular, respiratory, and musculoskeletal capacity required to perform prolonged exercise [1]. There is strong evidence suggesting that higher levels of CRF are associated with a healthier metabolic profile [2-4], and that CRF is a powerful predictor of morbidity and mortality [1,5,6] Given the well-known relevance of CRF to general health status, its assessment has been strongly recommended in the recent American Heart Association scientific statement, proposing the assessment of CRF as a clinical vital sign [7]. Additionally, CRF assessment is important for testing the success of an intervention and for monitoring purposes.

The maximal oxygen consumption (VO_2max) is an objective measure of CRF and has been considered to be its best indicator [8]. The American College of Sports Medicine provides guidelines, including different methods for CRF testing and ergo spirometry conducted during incremental maximal exercise tests on treadmills or a cycle ergometer, as these methods are considered the gold-standard measures of CRF [9]. In this context, the 2-km walk test (UKK test) is a widely used field-based battery of fitness tests such as the adult version of the EUROFIT battery and more recently the ALPHA fitness test battery for adults [10]. The UKK test has a distinct characteristic from most other field-based tests aiming to indirectly estimate VO_2max (eg, the 6-min walk test). After completing the exertion (ie, walking 2 km as fast as possible), the tester records the time spent (which for a given distance is a measure of performance as an indicator of walk speed) plus the physiological response to that exertion (ie, the heart rate [HR]). If either of these two parameters changes, the estimation of VO_2max would also change, making the test very sensitive to detect small changes in the CRF level, which is important for assessing the effectiveness of an intervention for monitoring purposes. Despite these advantages of the UKK test for measuring CRF, to properly perform this test, a valid and reliable HR monitor, stopwatch, and instrument (eg, measuring tape or GPS) are required to measure a 2-km route. It would be very useful and practical if a valid and reliable smartphone app could replace all of these instruments. Moreover, if a sports specialist, clinician, or researcher wants to assess the CRF level of several individuals, the individuals requiring the test need to visit the assessment center, which imposes a geographic limit as to who can be tested.

In this context, an app based on the UKK test would provide the possibility to assess CRF remotely for individuals who are geographically distributed around a country or continent, or even among those living on different continents, with minimal equipment and costs. The app would require the calculation of distance traveled, time, and HR at the end of the walk. To our knowledge, there is no currently validated app to estimate VO_2max through the equations provided by the developers of the UKK test [11]. We have also not found any fitness app that integrates a measure of fitness performance (ie, walking speed estimated from measuring a 2-km distance and the time spent to complete it) and the HR measured into an estimation of VO_2max using the camera of a mobile phone, which would remove the need for an additional monitoring device (eg, HR chest band).

Notably, some authors have demonstrated the validity of apps for measuring distance or HR in isolation. For instance, Benson et al [12] and Gordon et al [13] demonstrated the validity of a GPS-enabled iPhone app to track exercise distance. In addition, Martinez-Nicolas et al [14] revised the Runkeeper app and suggested its suitability for tracking distance. Otherwise, over the last few years, smartphone apps have gained the ability to measure HR by detecting the pulse using photoplethysmographic (PPG) imaging [15]. In this sense, Mitchell et al [16] examined the accuracy of Instant Heart Rate (Azumio) for pulse rate measurement, which was compared to that of an FT7 Polar HR monitor. The app was proven to be valid and reliable at rest and immediately postexercise. Likewise, Poh and Poh [17] showed strong agreement between HR assessments obtained using the Cardiio app against a Food and Drug Administration–eared pulse oximeter at rest and after moderate to vigorous exercise. Based on this background, we decided to develop an app ad hoc, named 2kmFIT-App, that can unite the measure of the distance walked, particularly the 2 km of the UKK test, the time needed to complete the 2-km walk, and the HR at the end of the test using the phone camera.

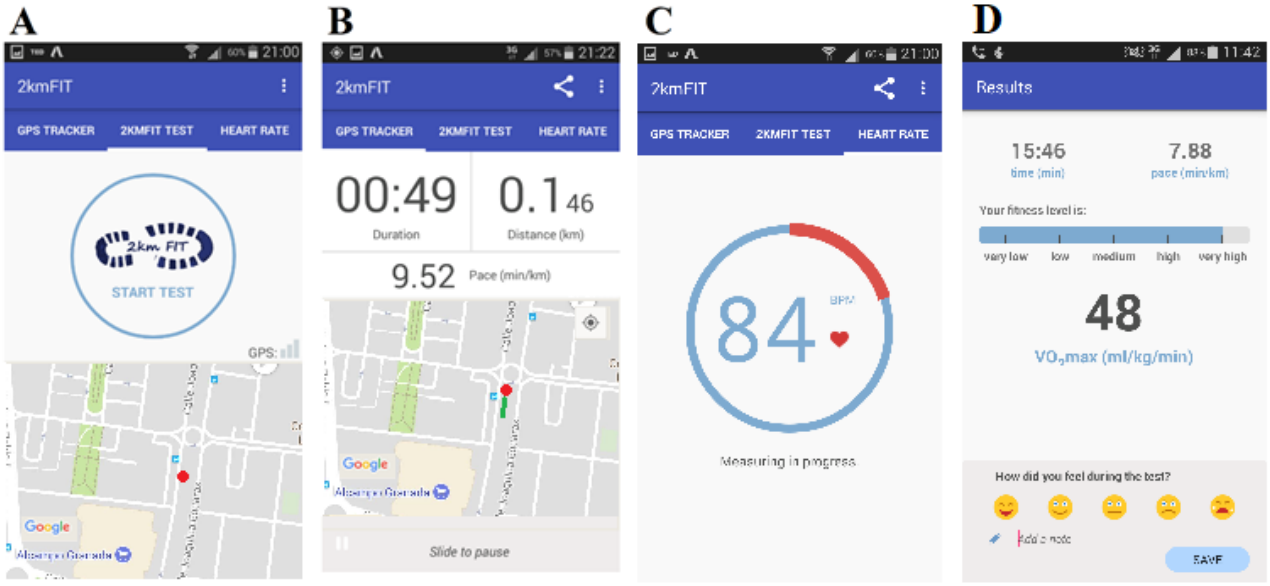
The overall aim of this study was to evaluate the validity and reliability of 2kmFIT-App developed ad hoc for Android and iOS mobile operating systems to estimate VO_2max, in comparison with VO_2max calculated following the original instructions and instruments of the UKK test. The specific aims of the study were to determine the validity of 2kmFIT-App to track distance and calculate HR versus a measuring wheel and HR monitor, respectively.

Methods

Development of 2kmFIT-App

We developed 2kmFIT-App (Figure 1), which is able to track distance through the GPS of the smartphone, time, and HR using PPG imaging from the phone camera. The app was created to run on both major mobile operating systems. Android Studio 2.2.3 and Java 1.8_112 were used to develop the Android app, and the ButterKnife, Timber, AppIntro, DBFlow, Saripaar, Google Maps, CardView, and MPAndroidChart frameworks (Android Inc, USA) were used to design the required functionalities of the Android app. Likewise, Xcode 8.3.3 and Swift 3 were used for the Mac operating system to develop the iOS app, and the AVFoundation, MapKit, QuartzCore, Charts, Realm, and PermissionScope frameworks were utilized to create the required functionalities of the iOS app.

Figure 1. Main screens of the 2kmFIT-App Android version.



Different versions of 2kmFIT-App were developed, both for Android (versions 1 and 2) and iOS (versions 1, 2, and 3), aiming to enhance precision in measurements. Finally, the app was programmed with the equations of Oja et al [18] based on the original UKK test and to estimate the VO_2max in both mobile operating systems. The CRF reference values of Rodriguez et al [19] were also programmed into the app. 2kmFIT-App has been registered in Intellectual Property Registry (Safe Creative register number 1904040536262). More detailed information on the structure and content of 2kmFIT-App can be found in Multimedia Appendix 1.

Validation Protocol

Design

The validation evaluation of 2kmFIT-App was performed in two phases: in-laboratory validation (phase 1) and field-test validation and reliability (phase 2). A Samsung Galaxy SIII Neo (Android 4.4.2) and an iPhone 6s plus (iOS 11.3) were used for testing 2kmFIT-App during the two phases on the Android and iOS mobile operating system, respectively.

In-Laboratory Validation (Phase 1)

During this preliminary validation phase, different HR measurements were taken with 2kmFIT-App using the Android (version 1 and 2) and iOS (version 1, 2, and 3) versions, at two different exercise intensities (rest and moderate intensity). The HR measurement was tested because it is the most novel and challenging smartphone feature integrated into the app. In addition, during the UKK test, HR is taken only when the participant reaches the end of the 2-km walk. The objective of this first phase was to ensure that 2kmFIT-App could measure HR at the end of the test with a reasonable margin of error. One member of the research group examined the accuracy of 2kmFIT-App for HR measurement using both mobile operating systems under three different conditions: at rest, and with moderate and high exercise intensities. Additionally, four commercial iOS-based HR apps available in the App Store market were also tested to gather information on the accuracy of existing HR apps and to compare the HR accuracy of our developed app (see Multimedia Appendix 1 for detailed information).

Field Test Validation and Reliability (Phase 2)

Operating System

The second phase examined the validity and reliability of 2kmFIT-App using only the Android mobile operating system (version 2) in field conditions with 20 study participants. The iOS version of 2kmFIT-App was determined to have poor accuracy for HR measurement at rest and at medium intensity, and was therefore not included in phase 2. Accordingly, all findings presented herein refer only to the app running on the Android operating system.

Study Design and Participants

In this second phase, we used a cross-sectional design to test the validity of 2kmFIT-App by comparing the results obtained through the app data against criterion measures (UKK test). Additionally, we developed a test-retest design for testing its reliability in a repeated-measures analysis. A convenient sample of 20 healthy adults (25% female), who were mainly students from the University of Granada (Spain), were recruited for this study. We estimated the sample size needed for detecting correlation coefficients between the HR and VO_2max assessed with the app and with the criterion methods equal to or higher than 0.7, with a standard α error of 5%. Our power calculation model showed that with 17 participants we would have 95% power to detect the expected correlations between methods. We finally included 20 participants to have some additional residual power. The study was approved by the Human Research Ethics Commission of the University of Granada (ref: 280/CEIH/2017) and abides by the bioethical principles set out by the Declaration of Helsinki. Participants received information about the characteristics of the study and data management. Participants also provided written informed consent to participate in the study. Data from the volunteers were included in a database and

were protected according to Spanish Law 15/1999 of December 13, 1999.

Instruments

A professional TM12 measuring wheel from Top Measure was considered an accurate criterion measure for distance on the inner line of an athletic track. Criterion HR measurements were collected by an RS300X HR monitor from Polar (Kempele, Finland) and a smartphone was used as a stopwatch reference (LG G2 Mini, Android version 5.0.2). Weight (kg) and height (m) were obtained in one step using an electronic scale with an integrated stadiometer (Seca 769 scale with Seca 220 stadiometer, Hamburg, Germany) without shoes, in light clothing, and the Frankfort plane. BMI was calculated using the formula of weight/height^2 (kg/m^2). 2kmFIT-App was used during the performance of the UKK test as the instrument to be validated (see Multimedia Appendix 1 for a full description of the 2kmFIT-App structure and content).

Testing Protocol

The test was performed by the participant from 9 AM to 5 PM on an athletics track made of an artificial surface of polyurethane in an outdoor setting. The weather was mainly sunny, and temperatures ranged between 8°C and 26°C. All measures were collected from March 6 to 22, 2017 at the Faculty of Sports Sciences of the University of Granada (Granada, Spain). Participants performed the test twice (test-retest) with a 1-week interval between assessments. Before starting measurements, the experimental protocol was individually explained to participants, and then their weight and height were measured immediately afterward. The participants were then asked to wear the Polar RS300X watch and the chest strap. A user profile was created, and the name of the participant, along with their gender, birth date, weight, and height were entered into the "users" section of 2kmFIT-App.

HR Instructions

At the starting line of the test, participants were instructed on how to perform a valid HR measurement with 2kmFIT-App following the instructions shown in Textbox 1. Once the participants felt familiar with the process and obtained two valid HR measures with the app, they were considered to be sufficiently trained to perform the UKK test.

Textbox 1. Instructions for heart rate (HR) measurement using 2kmFIT-App.

1. Stay standing, relax, and breathe normally.
2. Try not to swallow while measuring.
3. Place the index fingertip vertically on the camera.
4. Try not to move your finger, and apply constant pressure on the camera lens.
5. Do not apply excessive pressure.
6. Tap the button (with your other hand) on the screen to start the HR measurement and at this moment the camera's flashlight will turn on.
7. Keep the phone stable until the HR measurement ends.
8. When finishing the walked distance, it is recommended to tap the camera lens with the finger as quickly as possible to reduce the time of HR recovery.
9. If the fingertip or ambient temperature is cold, the use of gloves during the walked distance is recommended.

UKK Test Instructions and Procedures

The tester reminded the participants of the instructions for performing the test following the original protocols and recommendations [18]. The app was then initialized (Figure 1A) and the UKK test was started. The distance was tracked using the app's GPS feature (Figure 1B). When the participants approached the 2-km mark (1.8 km), the app emitted three beeps indicating that the test will soon be complete, and participants had to get ready for HR measurement using the smartphone camera. The app then indicated the end of the test with another audio signal. At this moment, the participants stopped walking and measured their HR using 2kmFIT-App, as detailed in the "HR instructions" section (Figure 2). Once 2kmFIT-App finished the HR measurement, the following outcomes were collected from the app: time spent to complete the test (ie, as 2 km estimated by the app), HR at the end of the test, and VO_2max calculated by 2kmFIT-App. For the HR measurement, the tester also noted the outcomes obtained using criterion instruments on the provided datasheet. The outcomes were the three times recorded from the stopwatch (T0, T1, T2) and HR measured by the Polar monitor at three separate moments: P0, at the end of the walk (at the moment the final beep emitted by the app); P1, when the app started to measure HR; and P2, when the HR measurement was completed and shown in the app. 2kmFIT-App took an average of 25.85 (SD 0.98) seconds to obtain HR measurements. Finally, the criterion VO_2max was calculated using the original equations of the UKK test and all of the criterion measures [11].

Figure 2. Correct placement of the index fingertip on different locations of the camera lens to obtain a correct heart rate measurement by photoplethysmographic imaging.

Statistical Analysis

For the field-test validation and reliability (phase 2), the data of the UKK test from the 20 participants in total (2 trials) are presented as the mean (SD). A one-sample *t* test was used to evaluate whether the mean difference (ie, systematic error) between the app criterion measures was significantly different ($P<.05$) from zero (reference). The mean absolute percentage error (MAPE) was also calculated as the mean difference of the 2kmFIT-App outcome minus the respective criterion outcome (×100/mean criterion outcome). To complement this analysis, the intraclass correlation coefficient (ICC) was calculated between the criterion and 2kmFIT-App measurements, and between the test and rest. Interpretation of ICC values was based on standardized guidelines, in which a value less than 0.5 indicates poor reliability, 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. Additionally, the Pearson product-moment correlation coefficient was calculated and linear regressions were performed to analyze the validity of 2kmFIT-App with criterion references and test-retest reliability. The agreement between the traditional UKK test and 2kmFIT-App measures was examined using the Bland and Altman method [20]. The mean difference (error) and 95% limits of agreement (error, 1.96 SD) were calculated. Results were graphically examined by plotting the differences against their mean. Bland-Altman analysis was performed with SigmaPlot 12.5 for Windows. Statistical analyses were performed using the Statistical Package for the Social Sciences software version 23.0 (SPSS Inc, Chicago, IL, USA). The level of statistical significance was set at $P<.05$.

Results

Participants' Characteristics

Twenty participants were included in phase 2 for field-test validation and reliability analysis. Of note, 6 participants (from the test and retest) were excluded due to the following reasons: the participant failed to place the fingertip in the correct position when measuring (n=1), the smartphone did not provide data (n=2), or the smartphone froze during the test (n=3). Therefore, 16 men and 4 women ranging in age from 19 to 29 years (mean 24.96, SD 2.33 years) were included in the analysis. Descriptive characteristics of the participants are shown in Table 1, and their 2-km walk UKK test and retest results are shown in Table 2. During phase 1, only the Android version of 2kmFIT-App showed accuracy in HR measurement at medium intensity; therefore, no data are shown for iOS in phase 2 (see Multimedia Appendix 1 for these results).

Table 1. Characteristics of the sample (N=20).

Characteristic	Mean (SD)
Age (years)	24.96 (2.33)
Weight (kg)	69.73 (10.52)
Height (m)	1.74 (0.07)
BMI (kg/m^2)	23.03 (2.29)

Table 2. Test and retest results of the 2-km walk UKK test (N=20).

Variable	Overall (test and retest), mean (SD)	Test, mean (SD)	Retest, mean (SD)	*P* value[a]
Outside temperature (°C)	15.80 (3.28)	18.40 (4.32)	13.20 (4.58)	.001
App distance[b] (meters)	1942.60 (38.31)	1929.61 (51.47)	1955.60 (34.91)	.01
Pace (min/km)	8.29 (0.88)	8.37 (0.90)	8.20 (0.88)	.03
Time (minutes)				
Criterion[c]	16.57 (1.75)	16.75 (1.79)	16.39 (1.77)	.03
App[d]	16.09 (1.82)	16.16 (1.93)	16.02 (1.83)	.50
Time (seconds)				
P0-P2[e]	25.85 (4.99)	25.40 (6.48)	26.30 (5.89)	.59
P1-P2[f]	19.24 (4.95)	18.78 (5.87)	19.70 (6.34)	.57
Heart rate (beats/minute)				
Criterion P0[g]	142.08 (17.37)	142.35 (16.54)	141.80 (19.48)	.81
Criterion P1[h]	139.90 (17.66)	140.65 (16.77)	139.15 (19.92)	.53
Criterion P2[i]	125.90 (19.78)	127.55 (16.84)	124.25 (23.89)	.23
App[j]	123.35 (21.49)	125.60 (18.32)	121.10 (26.05)	.15
VO_2max[k] (ml/min/kg)				
Criterion P0[l]	38.71 (5.17)	38.16 (5.27)	39.27 (5.20)	.007
Criterion P1[l]	38.95 (5.25)	38.35 (5.29)	39.56 (5.35)	.004
Criterion P2[l]	40.49 (5.71)	39.79 (5.73)	41.20 (5.87)	.006
App[m]	42.21 (5.93)	41.75 (6.06)	42.67 (6.18)	.19

[a]Paired-samples *t* test between test and retest.

[b]Real distance at which the test measured by means of the app ends.

[c]Estimated time of the UKK test if mean walking speed up to 2 km as measured by the criterion method would have been maintained.

[d]Time taken to perform the test measured by the app.

[e]P0-P2: time difference between the final beep emitted for the app and showing the heart rate on the app screen.

[f]P1-P2: time difference between the start of heart rate measurement and the end of heart rate calculation by means of the app.

[g]Polar heart rate immediately when the test finished.

[h]Polar heart rate when starting the measurement with the app.

[i]Polar heart rate at the end of the measurement with the app.

[j]Heart rate measured by the app at the end of the test.

[k]VO_2max: maximum oxygen consumption.

[l]Criterion VO_2max P0, P1, and P2: VO_2max estimated with the hypothetical arrival at the UKK test finish line considering walking speed and heart rate at P0, P1, and P2, respectively.

[l]VO_2max estimated by the app calculation.

Field Test Validation and Reliability (Phase 2)

Validity

Table 3 shows the mean differences among app minus criterion measures within the app, along with the ICC and MAPE values, which express accuracy as a percentage of the error. Of note, distance, time, and HR were underestimated by the app, and consequently overestimated VO_2max. The systematic error was significantly different from zero for all studied outcomes (P<.001) except for HR at P2. In addition, there was moderate validity found for HR at P0 and good validity found for VO_2max at P0 based on the ICC.

Table 3. Validity of 2km FIT-App (Android version) in comparison with criterion measures (UKK test).

Outcomes	2kmFIT-App measure, mean (SD)	Criterion measure (UKK test), mean (SD)	Difference (app – criterion), mean (SD)	P [a]	MAPE[b] (%)	ICC[c] (95% CI)	r [d]
Distance (meters)	1929.6 (51.47)	2000.0 (0.00)	–70.40 (51.47)	<.001	–3.52	N/A[e]	
Time (min)	16.16 (1.93)	16.75 (1.79)	–0.58 (0.44)	<.001	–3.58	0.971 (0.928-0.988)	0.97
Heart rate (beats/minute)							
P2[f] – P0[g]	N/A	142.35 (16.54)	–16.75 (9.96)	<.001	–11.77	0.731 (–0.211 to 0.928)	0.84
P2 – P1[h]	N/A	140.65 (16.77)	–15.05 (9.56)	<.001	–10.70	0.770 (–0.197 to .0939)	0.86
P2	125.60 (18.32)	127.55 (16.84)	–1.95 (7.83)	.28	–1.53	0.947 (0.869-0.979)	0.90
VO_2max[i] (ml/min/kg)							
P2 – P0	N/A	38.16 (5.27)	3.59 (2.01)	<.001	9.41	0.878 (–0.125 to 0.972)	0.95
P2 – P1	N/A	38.35 (5.29)	3.40 (2.03)	<.001	8.88	0.887 (–0.102 to 0.974)	0.94
P2	41.75 (6.06)	39.79 (5.73)	1.96 (1.89)	<.001	4.93	0.948 (0.569-0.986)	0.95

[a]One-sample t test: the intertrial difference was entered as a dependent variable; the P value indicates whether the mean difference is significantly different from 0 for all measures.

[b]MAPE: mean absolute percentage error; calculated as ([study outcome – criterion measure]/criterion measure) × 100.

[c]ICC: intraclass correlation coefficient; model two-way mixed and single measures.

[d]Pearson correlation coefficient.

[e]N/A: not applicable.

[f]Polar heart rate when the measurement of the app is completed and shown.

[g]Polar heart rate immediately when the test finished.

[h]Polar heart rate when starting the measurement with the app.

[i]VO_2max: maximum oxygen consumption; estimated with the hypothetical arrival at the UKK test finish line considering walking speed and heart rate at P0, P1, and P2, respectively.

Figures 3 and 4, and Figures S1 and S2 in Multimedia Appendix 1 show the Bland-Altman plots used to evaluate the agreement between the estimated HR and VO_2max measured through 2kmFIT-App against criterion measures. The MAPE (Table 3) revealed that 2kmFIT-App underestimated distance, time, and HR at P0, and consequently overestimated VO_2max at P0.

Figure 3. Agreement and regression plots for heart rate as measured by the app (2kmFIT-App, Android version) versus criterion. In the Bland-Altman plot, the central line represents the mean difference (systematic error) and the upper and lower dotted lines represent the 95% limits of agreement (mean difference, SD 1.96 of the differences). Criterion HR P0: polar HR taken immediately when the test finished; R: Pearson correlation coefficient; R2: coefficient of determination; bpm: beats per minute.

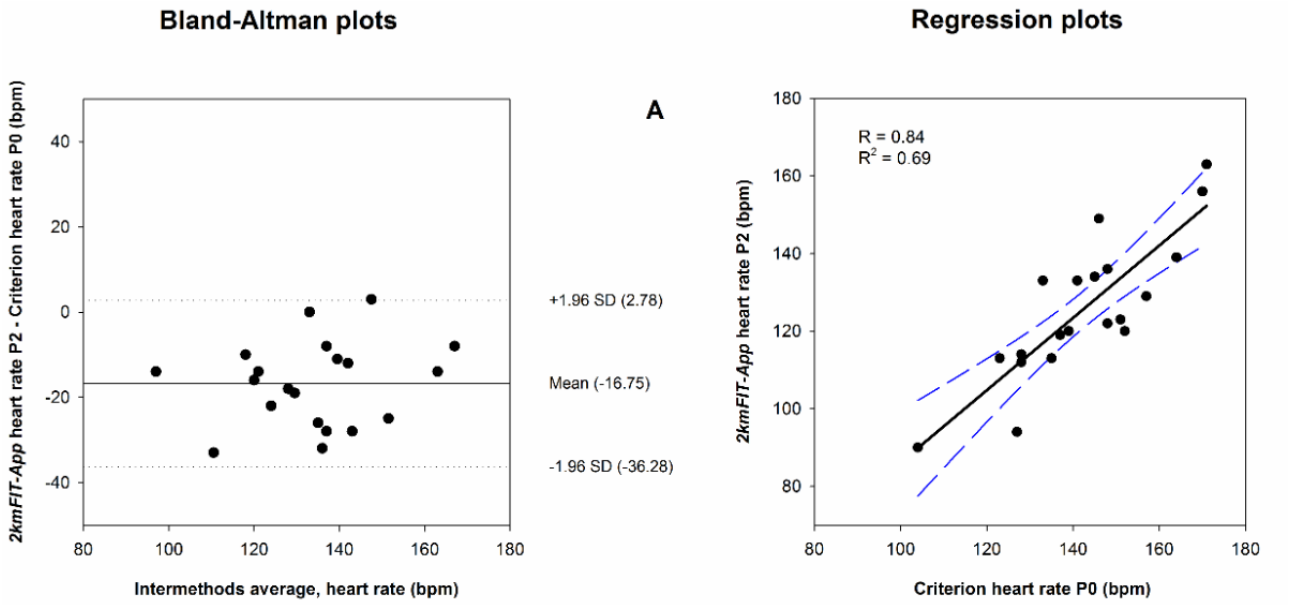
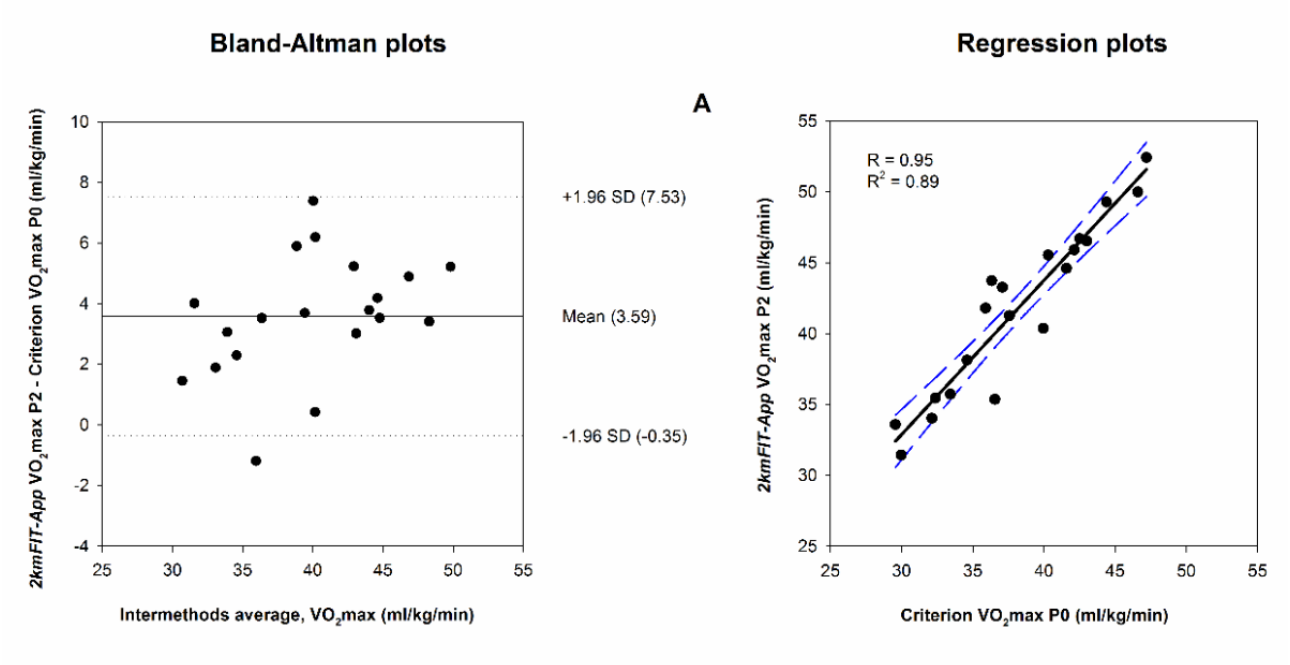


Figure 4. Agreement between the estimated app (2kmFIT-App, Android version) maximal oxygen consumption (VO2max) versus criterion VO2max. The central line represents the mean difference (systematic error) between app and criterion measures. Upper and lower dotted lines represent the 95% limits of agreement (mean difference, SD 1.96 of the differences). R: Pearson correlation coefficient; R2: determination coefficient. Agreement between app VO2max and (A) criterion VO2max P0 (VO2max estimated to the hypothetical arrival at the UKK test finish line considering walking speed and heart rate at P0).



Reliability

Table 4 shows the mean differences among app minus test-rest measures within the app for assessing reliability, along with the ICC values. The systematic error was significantly different from zero for distance and pace. There was good reliability as revealed by the ICC for app HR and excellent reliability for app VO_2max.

Table 4. Reliability of 2kmFIT-App (Android version) based on test-retest analysis with the app.

Variable	Test, mean (SD)	Retest, mean (SD)	Difference (retest – test), mean (SD)	*P* value[a]	ICC[b] (95% CI)	*r*[c]
Distance (meters)	1929.61 (51.47)	1955.60 (34.91)	25.99 (43.21)	.01	0.620 (0.077-0.848)	0.56
Time (minutes)	16.16 (1.93)	16.02 (1.83)	–0.15 (0.94)	.50	0.935 (0.839-0.974)	0.88
Pace (min/km)	8.37 (0.90)	8.20 (0.88)	-0.18 (0.33)	.03	0.957 (0.870-0.984)	0.93
Heart rate P2[d] (beats/minute)	125.60 (18.32)	121.10 (26.05)	–4.5 (13.44)	.15	0.897 (0.742-0.959)	0.87
VO_2max P2[e] (ml/kg/min)	41.75 (6.06)	42.67 (6.18)	0.92 (3.04)	.19	0.932 (0.830-0.973)	0.88

[a]One-sample *t* test.

[b]ICC: intraclass correlation coefficient.

[c]*r:* Pearson correlation coefficient.

[d]Heart rate P2: Polar heart rate when the measurement of the app is completed and shown.

[e]VO_2max P2: maximum oxygen consumption considering polar heart rate when the measurement of the app is completed.

Discussion

Principal Findings

The objective of this study was to analyze the validity and reliability of a smartphone app (2kmFIT-App) for measuring distance, HR, and CRF using Android and iOS mobile operating systems. 2kmFIT-App was demonstrated to be valid and reliable with the Android mobile operating system for measuring CRF, HR, and distance in comparison with CRF estimated by the UKK test, HR measured by a standard Polar HR monitor, and distance measured with a measuring wheel, respectively. First, our validity analysis revealed that 2kmFIT-App (Android, version 2) underestimated distance (–70.40 meters) and HR (–1.95 beats/minute), and overestimated VO_2max (3.59 ml/min/kg) with an ICC higher than 0.73 for all variables. Second, test-retest reliability showed that 2kmFIT-App (Android, version 2) overestimated distance (25.99 meters) and VO_2max (0.92 ml/min/kg), but underestimated HR (–4.5 beats/minute). Third, the iOS version of 2kmFIT-App did not obtain accurate measures of HR at medium exercise intensity, thereby its field-test validation and reliability were not further investigated. Collectively, our investigation highlights the potential of 2kmFIT-App as a new and portable device for safely measuring CRF with a low margin of error in the Android mobile operating system.

In-Laboratory Validation (Phase 1)

One of the major challenges of this investigation was the validity of 2kmFIT-App iOS versions to measure HR. None of the iOS versions we developed achieved precise HR measurements at medium exercise intensities. This lack of accuracy was also found when we tested other commercially available iOS-based HR apps (see Multimedia Appendix 1). This finding concurs with the study of Bouts et al [21] who did not find strong correlations between an electrocardiogram and two iOS-based HR apps (Instant Heart Rate: HR monitor; Runtastic Heart Rate Monitor). More significant bias in HR measurement was found at medium exercise intensities. Although identifying the technological reason for this error is complex, we can speculate a justification for this outcome. The iPhone 6 and later models include a hybrid infrared radiation filter. This filter is designed to reflect or block infrared wavelengths and is usually used to enhance poor lighting conditions. In resting conditions, the difference between oxygenated (red color) and deoxygenated (blue color) blood is low; however, during medium or higher exercise intensities, the contrast becomes higher. Therefore, a possible hypothesis is that the hybrid infrared filter preprocesses the finger image captured (diminishing the contrast between oxygenated and deoxygenated blood), which increases noise in the measurement of HR. Specifically, it could be assumed that the more variability in the blood volume, which occurs at higher frequencies of HR, the greater the measurement error will be.

Field Test Validation and Reliability (Phase 2)

Our results revealed high levels of agreement between 2kmFIT-App and criterion references, with an ICC ranging from 0.73 to 0.97. Although ICC and *r* values suggested good validity, we also found a systematic error. The criterion-related validity analyses suggested that the app measure of distance provided lower values than the criterion distance, which could be interpreted as a slight underestimation. Specifically, 2kmFIT-App underestimated the distance by 70.40 meters, demonstrating a <4% MAPE from the reference (2 km). Similar to this finding, Benson et al [12] obtained an underestimation of 80 meters (trial 1) and 49 meters (trial 2) with the Motion X GPSTM app against sport-specific global GPS with a criterion distance of 2.4 km. The systematic error between repeated measures (reliability) taken through 2kmFIT-App demonstrated a relatively low error in the distance (25.99 meters). Of note, our investigation was performed in an open-air area (on an athletic track); however, some aspects such as the relatively dense environment (eg, tall buildings, dense vegetation, urban canyons), the manner of carrying the smartphone (pocket or arm), and walking in a straight line or making circles might influence satellite fixing, and therefore precision in distance

measurement [22]. Despite these variables, the interest in using apps as stand-alone physical activity monitors is increasing [23]; thereby, our app has proven accuracy in track distance, indicating its validity for determining the walked distance.

2kmFIT-App showed a high degree of validity for measuring HR against Polar RS300X at the end of the UKK test. The systematic error between the app HR and criterion HR at P2 was –1.95 beats/minute, indicating an underestimation of the app with a MAPE <2%. The accuracy of 2kmFIT-App was equivalent regardless of the HR at which the UKK test was finished. Moreover, it is worth mentioning that absolute differences over 20 beats/minute were not observed in any measurement, and only 14.63% (6 measures) reached a difference of up to 10 beats/minute. These results do not concur with those of Coppetti et al [24] who found more than 20% absolute differences at over 20 beats/minutes using some commercially available apps. Furthermore, in contrast to our data, Coppetti and colleagues showed an overestimation for the Instant Heart Rate app (4.52 beats/minute) and Heart Fitness app (1.96 beats/minute) measured at resting conditions. Nevertheless, our results are in line with those of Yan et al [25] who found slight underestimation by the Cardiio smartphone app at different exercise intensities, thereby confirming our findings.

The reliability results (retest minus test) indicated the good reliability of 2kmFIT-App with an ICC of 0.89. Of note, this study was conducted under uncontrolled light and temperature conditions, which might have influenced the absorbed and reflected light by the blood and finger tissues, and in turn increased the measurement error [26]. Thus, using our app in relatively controlled light and temperature environments could achieve more valid results. Although 2kmFIT-App was proven to be a portable and cost-effective tool for monitoring HR in resting and postexercise conditions, such influencing factors should be considered when using the app.

The most salient finding of this study is that the Bland-Altman graph showed valuable information about the high level of agreement for app VO_2max at P2 against VO_2max estimated by the UKK test, as revealed by the ICC (0.94) and MAPE (4.93%). Furthermore, the accuracy of the app VO_2max at P2 was the same regardless of the VO_2max obtained at the end of the test. In addition, the criterion-related validity analyses suggested that app VO_2max at P2 provides higher values than the criterion, which was interpreted as an overestimation. According to the UKK test guide, walking time is the most important factor affecting the results of the test. In this sense, 2kmFIT-App underestimated the time by an average of 0.58 minutes, which provides one possible explanation for the overestimation. Another important aspect that likely contributed to the measurement error of app VO_2max is HR. 2kmFIT-App took an average of 25.85 seconds to measure HR from the end of the test (time at P0 – time at P2), and during that time the HR fell 16.18 beats/minute on average with 59.46% variation (between participants). Thus, the time needed for measuring HR through 2kmFIT-App is the major drawback to be recognized in our investigation. Nevertheless, this limitation could be improved with two actions: decreasing the measurement time of HR by 2kmFIT-App (ie, improving smartphone technology) and applying a correction factor to the equation for estimating app VO_2max. In this context, the equation VO_2max corrected=2.24 + 0.89 × X can be used to solve the overestimation of 2kmFIT-App in determining the CRF, in which X is the estimated VO_2max when using the criterion methods recommended in the original UKK test.

Moreover, reliability analyses revealed excellent agreement (ICC=0.93) within trials (retest minus test) of app VO_2max. These results suggest the suitability of 2kmFIT-App for monitoring changes in intervention studies with a smaller margin of error, in which precise measurements are needed. Similarly, Brooks et al [27] demonstrated the reliability of the SA-6MWT app to measure exertional capacity using the 6-minute walk test. However, 2kmFIT-App is the only app that is currently able to estimate VO_2max considering a single physiological measure (HR).

There is no doubt that apps have great potential in clinical and research settings [28]; however, they may not always have an acceptable margin of error. Therefore, researchers, clinicians, and sports specialists should demand scientific validation of apps. In this context, the risk of CRF testing should also be considered, especially in maximal tests. Thus, an app using a submaximal test to estimate CRF is a safer choice, especially when testing is not supervised. 2kmFIT-App is the first validated app for this purpose that is capable of being self-administered and to remotely monitor CRF, making it suitable for most people. Although 2kmFIT-App is suitable for the general population, it can be an especially powerful tool for the prevention and management of cardiovascular disease risk factors [29].

Limitations and Strengths

A limitation of our study is the time required by 2kmFIT-App to calculate HR immediately at the end of the 2-km walk. This fact leads to a decrease in HR (recovery) and results in a slight overestimation of VO_2max. Moreover, there were some occasional technical issues that should be recognized, such as app freezing or reduced blood circulation in the fingers when a colder temperature could influence data collection. In addition, as a common feature of device-based experiments, the participants were aware of the device worn and the study design, and therefore were not blinded. Additionally, 2kmFIT-App was tested with specific Android hardware, whereas other smartphones with different hardware but the same operating system might potentially present different results. The gold-standard methods to measure HR and VO_2max were not used in this investigation, since our goal was to test whether we could translate the original UKK test protocol to a smartphone version, and therefore we followed the instructions of the original protocol. The validity of the UKK test to estimate VO_2max against gas analyzed in laboratory conditions has been proven elsewhere [11,18,30,31]. However, to our knowledge, 2kmFIT-App is the first native app capable of estimating CRF, including a physiological measure (HR), after exercise using PPG technology. The high reliability shown by 2kmFIT-App to estimate CRF, HR, and distance should be recognized as a

strength. The possibility of self-administering this test anywhere in the world and at any time makes 2kmFIT-App a powerful tool for public health.

Conclusion

The results of this study provide important messages for sports specialists and health care professionals. 2kmFIT-App is a new and scientifically validated tool that is capable of objectively and remotely estimating CRF, HR, and distance with a low margin of error in the Android, but not in the iOS, mobile operating system. Given the high reliability achieved, 2kmFIT-App can be used for measuring and monitoring changes precisely with an Android phone. The utility of this app would not only be for the scientific field but also for the millions of people who currently perform physical exercise at a recreational level (not professionally) and want to track their level of CRF.

Acknowledgments

FBO research activity was supported by the Spanish Ministry of Economy and Competitiveness—MINECO/FEDER DEP2016-79512-R; the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 667302; the University of Granada, Plan Propio de Investigación 2016, Unit of Excellence on Exercise and Health (UCEES); the Junta de Andalucía, Consejería de Conocimiento, Investigación y Universidades, and European Regional Development Fund (ERDF), ref. SOMM17/6107/ UGR; the EXERNET Research Network on Exercise and Health in Special Populations (DEP2005-00046/ACTI); and the SAMID III network, RETICS, funded by the PN I+D+I 2017-2021 (Spain), ISCIII Sub-Directorate-General for Research Assessment and Promotion, the European Regional Development Fund (ERDF) (Ref. RD16/002). AN was supported by the Ministry of Economy and Competitiveness and the Instituto de Salud Carlos III through the CIBERFES (CB16/10/00239), the Seneca Foundation through the unit of excellence (Grant 19899/GERM/15), and the Ministry of Science Innovation and Universities RTI2018-093528-B-I00 (all of which are cofinanced by FEDER). CCS is supported by the Spanish Ministry of Science, Innovation and Universities (FJC2018-037925-I). The views expressed are those of the authors and do not reflect the official policy or position of the institutions they belong to.

Conflicts of Interest

None declared.

Multimedia Appendix 1
In-lab validation methods and results, detailed description of 2kmFIT-App structure and content, heart rate protocol, and supplementary results (Figures S1 and S2; Tables S1-S6).
[DOCX File , 1641 KB - mhealth_v9i1e14864_app1.docx]

References

1. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008 Jan;32(1):1-11. [doi: 10.1038/sj.ijo.0803774] [Medline: 18043605]
2. Lee D, Artero E, Sui X, Blair S. Mortality trends in the general population: the importance of cardiorespiratory fitness. J Psychopharmacol 2010 Nov;24(4 Suppl):27-35 [FREE Full text] [doi: 10.1177/1359786810382057] [Medline: 20923918]
3. Ortega FB, Ruiz JR, Labayen I, Lavie CJ, Blair SN. The Fat but Fit paradox: what we know and don't know about it. Br J Sports Med 2018 Feb;52(3):151-153. [doi: 10.1136/bjsports-2016-097400] [Medline: 28583992]
4. Ortega FB, Lavie CJ, Blair SN. Obesity and Cardiovascular Disease. Circ Res 2016 May 27;118(11):1752-1770. [doi: 10.1161/CIRCRESAHA.115.306883] [Medline: 27230640]
5. Harber MP, Kaminsky LA, Arena R, Blair SN, Franklin BA, Myers J, et al. Impact of Cardiorespiratory Fitness on All-Cause and Disease-Specific Mortality: Advances Since 2009. Prog Cardiovasc Dis 2017;60(1):11-20. [doi: 10.1016/j.pcad.2017.03.001] [Medline: 28286137]
6. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA 2009 May 20;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]
7. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, American Heart Association Physical Activity Committee of the Council on LifestyleCardiometabolic Health, Council on Clinical Cardiology, Council on Epidemiology Prevention, Council on Cardiovascular Stroke Nursing, Council on Functional Genomics Translational Biology, Stroke Council. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. Circulation 2016 Dec 13;134(24):e653-e699. [doi: 10.1161/CIR.0000000000000461] [Medline: 27881567]
8. Shephard RJ, Allen C, Benade AJ, Davies CT, Di Prampero PE, Hedman R, et al. The maximum oxygen intake. An international reference standard of cardiorespiratory fitness. Bull World Health Organ 1968;38(5):757-764 [FREE Full text] [Medline: 5303329]

9. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription, 9th edition. Riverwoods, IL: Wolters Kluwer; Feb 09, 2013.
10. Ortega FB, Cadenas-Sanchez C, Lee DC, Ruiz JR, Blair SN, Sui X. Fitness and Fatness as Health Markers through the Lifespan: An Overview of Current Knowledge. Prog Prev Med (NY) 2018 Apr;3(2):e0013 [FREE Full text] [doi: 10.1097/pp9.0000000000000013] [Medline: 32671316]
11. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. Int J Sports Med 1991 Aug;12(4):356-362. [doi: 10.1055/s-2007-1024694] [Medline: 1917218]
12. Benson AC, Bruce L, Gordon BA. Reliability and validity of a GPS-enabled iPhone "app" to measure physical activity. J Sports Sci 2015;33(14):1421-1428. [doi: 10.1080/02640414.2014.994659] [Medline: 25555093]
13. Gordon BA, Bruce L, Benson AC. Physical activity intensity can be accurately monitored by smartphone global positioning system 'app'. Eur J Sport Sci 2016 Aug;16(5):624-631. [doi: 10.1080/17461391.2015.1105299] [Medline: 26505223]
14. Martinez-Nicolas A, Muntaner-Mas A, Ortega FB. Runkeeper: a complete app for monitoring outdoor sports. Br J Sports Med 2016 Aug 18;51(21):1560-1561. [doi: 10.1136/bjsports-2016-096678]
15. Jonathan E, Leahy M. Investigating a smartphone imaging unit for photoplethysmography. Physiol Meas 2010 Nov;31(11):N79-N83. [doi: 10.1088/0967-3334/31/11/N01] [Medline: 20871134]
16. Mitchell K, Graff M, Hedt C, Simmons J. Reliability and validity of a smartphone pulse rate application for the assessment of resting and elevated pulse rate. Physiother Theory Pract 2016 Aug;32(6):494-499. [doi: 10.1080/09593985.2016.1203046] [Medline: 27459148]
17. Poh MZ, Poh YC. Validation of a Standalone Smartphone Application for Measuring Heart Rate Using Imaging Photoplethysmography. Telemed J E Health 2017 Aug;23(8):678-683. [doi: 10.1089/tmj.2016.0230] [Medline: 28140834]
18. Oja K, editor. UKK Walk Test Tester's Guide. Finland: UKK Institute for Health Promotion Research; 2011.
19. Rodríguez F, Valenzuela A, Gusi N, Nàcher S, Gallardo I. Valoración de la condición física saludable en adultos (y II): fiabilidad, aplicabilidad y valores normativos de la batería AFISAL-INEFC. Apunts Educacion Fisica y Deportes 1998;4(54):54-65 [FREE Full text]
20. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. Statistician 1983 Sep;32(3):307. [doi: 10.2307/2987937]
21. Bouts AM, Brackman L, Martin E, Subasic AM, Potkanowicz ES. The Accuracy and Validity of iOS-Based Heart Rate Apps During Moderate to High Intensity Exercise. Int J Exerc Sci 2018;11(7):533-540 [FREE Full text] [Medline: 29541341]
22. Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med 2011 Nov;41(5):508-515 [FREE Full text] [doi: 10.1016/j.amepre.2011.06.046] [Medline: 22011423]
23. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W, et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. JMIR Mhealth Uhealth 2015 Apr 15;3(2):e36 [FREE Full text] [doi: 10.2196/mhealth.3505] [Medline: 25881662]
24. Coppetti T, Brauchlin A, Müggler S, Attinger-Toller A, Templin C, Schönrath F, et al. Accuracy of smartphone apps for heart rate measurement. Eur J Prev Cardiol 2017 Aug 02;24(12):1287-1293. [doi: 10.1177/2047487317702044] [Medline: 28464700]
25. Yan BP, Chan CK, Li CK, To OT, Lai WH, Tse G, et al. Resting and Postexercise Heart Rate Detection From Fingertip and Facial Photoplethysmography Using a Smartphone Camera: A Validation Study. JMIR Mhealth Uhealth 2017 Mar 13;5(3):e33 [FREE Full text] [doi: 10.2196/mhealth.7275] [Medline: 28288955]
26. Siddiqui SA, Zhang Y, Feng Z, Kos A. A Pulse Rate Estimation Algorithm Using PPG and Smartphone Camera. J Med Syst 2016 May 11;40(5):126. [doi: 10.1007/s10916-016-0485-6] [Medline: 27067432]
27. Brooks GC, Vittinghoff E, Iyer S, Tandon D, Kuhar P, Madsen KA, et al. Accuracy and Usability of a Self-Administered 6-Minute Walk Test Smartphone Application. Circ Heart Fail 2015 Sep;8(5):905-913 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.115.002062] [Medline: 26283292]
28. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
29. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution--using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015 Jun;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]
30. Suni J, Husu P, Rinne M. Fitness for health: the ALPHA-FIT test battery for adults aged 18–69. Tester's Manual. Finland: European Union DS and the UKK Institute for Health Promotion Research; 2011.
31. Laukkanen RMT, Oja P, Ojala KH, Pasanen ME, Vuori IM. Feasibility of a 2-km walking test for fitness assessment in a population study. Scand J Soc Med 1992 Jun 05;20(2):119-126. [doi: 10.1177/140349489202000210] [Medline: 1496331]

Abbreviations

CRF: cardiorespiratory fitness
HR: heart rate

ICC: intraclass correlation coefficient
MAPE: mean absolute percentage error
PPG: photoplethysmography
UKK test: 2-km walk test
VO2max: maximal oxygen consumption

Edited by L Buis; submitted 29.05.19; peer-reviewed by M Gaman, B Worm, R Islam, D Rubin; comments to author 14.05.20; revised version received 02.06.20; accepted 27.10.20; published 08.01.21.

Please cite as:
Muntaner-Mas A, Martinez-Nicolas A, Quesada A, Cadenas-Sanchez C, Ortega FB
Smartphone App (2kmFIT-App) for Measuring Cardiorespiratory Fitness: Validity and Reliability Study
JMIR Mhealth Uhealth 2021;9(1):e14864
URL: http://mhealth.jmir.org/2021/1/e14864/
doi:10.2196/14864
PMID:33416503

Original Paper

Deep Learning–Based Multimodal Data Fusion: Case Study in Food Intake Episodes Detection Using Wearable Sensors

Nooshin Bahador[1], PhD; Denzil Ferreira[1], PhD; Satu Tamminen[1], DSc; Jukka Kortelainen[1], MD, PhD

Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

Corresponding Author:
Nooshin Bahador, PhD
Faculty of Information Technology and Electrical Engineering
University of Oulu
Pentti Kaiteran katu 1
Oulu, FIN-90570
Finland
Phone: 358 417246274
Email: nooshin.bahador@oulu.fi

Abstract

Background: Multimodal wearable technologies have brought forward wide possibilities in human activity recognition, and more specifically personalized monitoring of eating habits. The emerging challenge now is the selection of most discriminative information from high-dimensional data collected from multiple sources. The available fusion algorithms with their complex structure are poorly adopted to the computationally constrained environment which requires integrating information directly at the source. As a result, more simple low-level fusion methods are needed.

Objective: In the absence of a data combining process, the cost of directly applying high-dimensional raw data to a deep classifier would be computationally expensive with regard to the response time, energy consumption, and memory requirement. Taking this into account, we aimed to develop a data fusion technique in a computationally efficient way to achieve a more comprehensive insight of human activity dynamics in a lower dimension. The major objective was considering statistical dependency of multisensory data and exploring intermodality correlation patterns for different activities.

Methods: In this technique, the information in time (regardless of the number of sources) is transformed into a 2D space that facilitates classification of eating episodes from others. This is based on a hypothesis that data captured by various sensors are statistically associated with each other and the covariance matrix of all these signals has a unique distribution correlated with each activity which can be encoded on a contour representation. These representations are then used as input of a deep model to learn specific patterns associated with specific activity.

Results: In order to show the generalizability of the proposed fusion algorithm, 2 different scenarios were taken into account. These scenarios were different in terms of temporal segment size, type of activity, wearable device, subjects, and deep learning architecture. The first scenario used a data set in which a single participant performed a limited number of activities while wearing the Empatica E4 wristband. In the second scenario, a data set related to the activities of daily living was used where 10 different participants wore inertial measurement units while performing a more complex set of activities. The precision metric obtained from leave-one-subject-out cross-validation for the second scenario reached 0.803. The impact of missing data on performance degradation was also evaluated.

Conclusions: To conclude, the proposed fusion technique provides the possibility of embedding joint variability information over different modalities in just a single 2D representation which results in obtaining a more global view of different aspects of daily human activities at hand, and yet preserving the desired performance level in activity recognition.

(JMIR Mhealth Uhealth 2021;9(1):e21926) doi:10.2196/21926

KEYWORDS

deep learning; image processing; data fusion; covariance distribution; food intake episode; wearable sensors

Introduction

It is a proven fact that chronic diseases including obesity, diabetes, and metabolic disorders are highly correlated with eating behavior and regarding the importance of this issue, application of wearable sensors for capturing eating-related activities has been widely studied in the literature [1-6]. These studies can be categorized into 3 groups, including food intake detection, food type classification, and food content estimation. Among these groups, food intake detection has been considered as the first phase in food intake monitoring, and studies around it mainly focused on detecting chewing activity (acoustic-based assessment) [7-10] or hand gestures movement (motion-based assessment) [11-13] during eating episodes. The majority of the proposed methods rely on single sensing approaches, for example, using electromyography sensor, accelerometer sensor, or microphone [14-17]. However, it is believed that precisely discriminating eating episodes from other confounding activities requires processing multiple parameters from several sources. For this reason, multimodal assessment is a common target of interest today. Taking as an example, using data from both accelerometer and gyroscope sensors proposed in [18] reached an accuracy of 75% in detecting eating activity. These kinds of sensors quantify specific features of hand-to-mouth gestures as well as jaw motion associated with dietary intake. Later on, adding images of food into data captured by accelerometer and gyroscope was suggested for eating episodes detection [19]. Analysis of these meal images can also extract information of food content and estimate dietary energy intake. Data taken by GPS were also added as input parameters to correctly predict eating activity [20]. Audio signals of chewing sound was a further option added to a data set of both motion data and meal images which improve the accuracy of eating periods detection up to 85% [21]. Therefore, according to the aforementioned studies, it is valuable to develop algorithms that can take advantage of multiple data sources for monitoring applications rather than focusing on a single sensor. The sources of different modalities will provide richer information in comparison to a single source.

However, although using different modalities provides further opportunities to explore more complementary information, the growing number of different modalities has brought new challenges due to increase in the volume and complexity of the data. For dealing with these high-dimensional data sets and lowering the computational time, some studies implement feature selection process including forward features selection [22], random forest [23], and principal component analysis [24] to reduce the data size and select important parameters/features. For combining the information captured by different sensors, the classification score fusion has been introduced in literature as an option. Papapanagiotou et al [25] fused support vector machine scores from both photoplethysmography and audio signals. Regarding discriminating eating episodes from other activities, researchers applied different classification tools from support vector machine [26] to artificial neural network [27]. They also found that appropriate epoch size ranges from 10 to 30 seconds [28]. However, recent advances in machine learning methods have increasingly captured the attention of many research groups for distinguishing food intake intervals from others based on deep learning techniques. Convolutional neural network has been used for automatically detecting intake gestures from raw video frames [29]. Convolutional neural network was also proposed by Ciocca et al [30] for image-based food recognition.

The aforementioned studies focused on data represented by features set and its corresponding fusion as well as decision fusion of classifiers. What remains to be addressed is investigating the sensor fusion for quantitatively integrating heterogeneous sources of information. Taking this into account, this study aimed to combine data derived from disparate sensors such that the resulting information has lower dimension and yet maintains the important aspects of original data.

To the best of our knowledge, there is no research focused on sensor fusion for personalized activity identification using different data sets collected by wearable devices. The proposed fusion here is based on a hypothesis that data captured by multiple sources are statistically correlated with each other and their 2D covariance representation has a unique distribution associated with the type of activity.

Methods

Implementation of Sensor Fusion Algorithm

The proposed algorithm automatically transformed data from different sensors in time into a single 2D color representation that provides fast effective support for discriminating eating episodes from others. The idea of this method was on the hypothesis that data driven by different sensors have correlation with each other and a covariance matrix of all these measurements has a unique distribution associated with each type of activity which can be visualized as a contour plot. With 2D covariance contour as input data sets, deep model–learned specific patterns in 2D correlation representation related to specific activity.

The following is a summary of the steps followed in the proposed method to detect eating episodes:

Step 1

Forming the observation matrix derived from all sensors; the corresponding covariance matrix can then be formed in 2 ways. The first way is to calculate pairwise covariance between each sample across all signals. The second way is to calculate pairwise covariance between each signal across all samples. The algorithm steps based on the second way are as follows:

The pairwise covariance calculation between each column combination:

$$C_{ij} = \text{cov}(H(:, i), H(:, j)) \quad \textbf{(1)}$$

where H is observation matrix.

The covariance coefficient of 2 columns of i and j can be calculated as follows:

$$\text{cov}(S_i, S_j)=1/(n-1)\Sigma^{m}_{k=1}(S_{ik}-\mu_i)(S_{jk}-\mu_j) \quad \textbf{(2)}$$

$$S_i = M(:, i),\ S_j = M(:, j)$$

where μ_i is the mean of S_i, μ_j is the mean of S_j, and m is the number of samples within the window.

Step 2

Obtaining the covariance coefficient matrix of all columns according to the following equation:

where n is the number of observations.

Step 3

Creating a filled contour plot containing the isolines of obtained matrix C so that given a value for covariance, lines are drawn for connecting the (x, y) coordinates where that covariance value occurs. The areas between the lines were then filled in solid color associated with the corresponding covariance value.

Step 4

Feeding contour plot to the deep network to classify the sequences related to each activity. Two different scenarios were considered for this study. These scenarios were different in terms of covariance matrix calculation, temporal segment size, type of activity, wearable device, subjects, and deep learning architecture. In the first scenario, calculating pairwise covariance between each sample across all signals was considered. In the second scenario, calculating pairwise covariance between each signal across all samples was taken into consideration.

First Scenario for Evaluation of Algorithm

In the first scenario, data were recorded from a single participant wearing the Empatica E4 wristband on the right hand for 3 days. The data set includes the following data: (1) ACC—data from 3-axis accelerometer sensor in the range [–2g, 2g] (sampled at 32 Hz); (2) BVP—data from photoplethysmograph (sampled at 64 Hz); (3) EDA—data from the electrodermal activity sensor in microsiemens (sampled at 4 Hz); (4) IBI—interbeat intervals, which represent the time interval between individual beats of the heart (intermittent output with 1/64-second resolution); (5) TEMP—data from temperature sensor expressed in the °C scale (sampled at 4 Hz); and (6) HR—These data contain the average heart rate values, computed in spans of 10 seconds.

The window length for this scenario was selected to be 500 samples. This analysis was performed on 2954 500-sample segments after making all signals in the same size with sample frequency of 64 Hz. Segments were picked so that 1000 of them contained sleeping intervals, and 1000 of them captured during working with computer and others were during eating episodes.

The deep learning architecture used in this scenario was a deep residual network. The proposed deep learning architecture for image-to-label classification is presented in Figure 1 and consisted of a deep residual network with 3 2D convolution layers, followed by batch normalization, ReLU, max pooling, and fully connected layers. The 2D convolutional layer applied sliding convolutional filters to the input contour image. The output of this network is a categorical response, and therefore a softmax and classification layers were also added as last layers. There is also a shortcut to jump over some layers.

Figure 1. Deep learning network architecture.

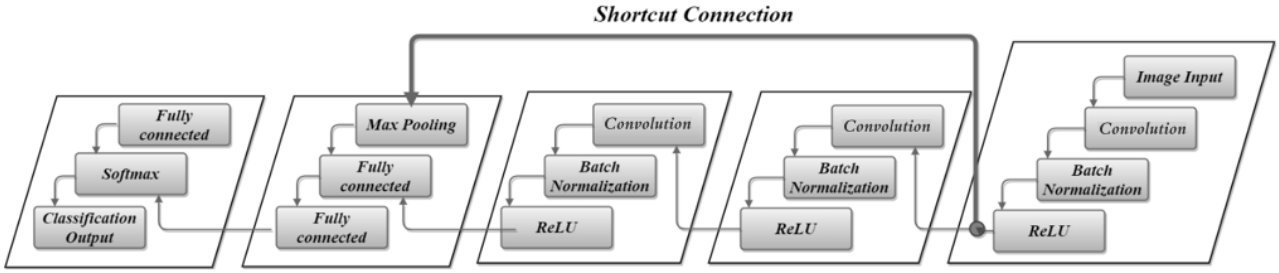


Table 1 provides detailed information about the proposed network layers. This information includes the sizes of layer activations. The training parameters of the deep learning model are given in Table 2. The mini-batch size and the maximum number of epochs were set to 100 and 10, respectively. Fivefold cross-validation was also used to check the performance of the model.

Table 1. Detailed information about the layers of proposed network.

Number	Layer type	Activations
1	Image input	300 × 300 × 3
2	Convolution	300 × 300 × 32
3	Batch normalization	300 × 300 × 32
4	ReLU	300 × 300 × 32
5	Convolution	150 × 150 × 64
6	Batch normalization	150 × 150 × 64
7	ReLU	150 × 150 × 64
8	Convolution	150 × 150 × 128
9	Batch normalization	150 × 150 × 128
10	ReLU	150 × 150 × 128
11	Convolution	150 × 150 × 128
12	Addition	150 × 150 × 128
13	Max pooling	75 × 75 × 128
14	Fully connected	1 × 1 × 500
15	Fully connected	1 × 1 × 10
16	Fully connected	1 × 1 × 3
17	Softmax	1 × 1 × 3
18	Classification output	—[a]

[a]—: Not available

Table 2. The model training parameters.

Parameter	Value
Initial learn rate	0.001
Learn rate drop factor	0.1
Learn rate drop period	2
Mini batch size	100
Max epochs	10
Learn rate schedule	Piecewise

Second Scenario for Evaluation of Algorithm

In the second scenario, an open data set associated with the activities of daily living was considered. The data set was collected from 10 healthy participants, performing 186 activities of daily living while wearing 9-axis inertial measurement units on both left and right arms [31].

The considered activities can be grouped into 4 separate categories: (1) mobility, including walking, sitting down and standing up, and opening and closing the door; (2) eating, including pouring water and drinking from glass; (3) personal hygiene, including brushing teeth; and (4) housework, including cleaning the table [31].

The recorded data include quaternions (with resolution of 0.0001), accelerations along the x, y, and z axes (with resolution of 0.1 mG), and angular velocity along the x, y, and z axes (with resolution of 0.01 degrees per second) [31].

Data annotation for all the experiments was manually performed based on videos recorded by an RGB camera [31].

The window length for this scenario was selected to be 50 samples. This analysis was performed on 4478 50-sample segments. Segments were picked so that 1132 segments contained walking episodes, 20 segments contained sitting down episodes, 16 segments contained standing up episodes, 366 segments contained opening the door episodes, 400 segments contained closing the door episodes, 1208 segments contained pouring water and drinking from glass episodes, 704 segments contained brushing teeth episodes, and 632 segments contained cleaning the table episodes.

As training from the scratch on relatively small-scale data sets is susceptible to overfitting, a pretrained model for extracting deep features was suggested in this scenario. The deep learning architecture used in this section was the InceptionResNetV2 pretrained model. This pretrained classification network has

already learned on more than 1 million images. As this network was trained on extremely large data sets, it is capable of being served as a generic model. Therefore, this section used layer activation of the pretrained InceptionResNetV2 architecture as features to train a support vector machine for classifying different activities. The parallel computing platform of Tesla P100 PCIe 16 GB was used for implementing this deep structure. The depth, size, and number of parameters in the pretrained InceptionResNetV2 network were 164, 209 MB, and 55.9 M, respectively. Leave-one-subject-out cross-validation was considered for performance evaluation of classification.

Results

Applying the Proposed Algorithm on the First Scenario

Signals captured from different sensors during eating, sleeping, and working with computer are plotted in Figure 2. This figure illustrates how characteristics of data coming from different sources vary a lot. Figure 3 shows how the values in the eating-related data captured by different sensors are spread out in their boxplots, and how their distributions differ from each other. As these signals cannot be described by the same distribution, they are said to be heterogeneous. This heterogeneity brings up the issue of how to integrate the information from such diverse modalities. This spread in the range of scales across the various modalities causes a simple approach to be not enough for reliable activity detection and a therefore a more sophisticated technique is required.

Figure 2. Time series amplitude of each modality captured during 1 episode of 3 different activities of eating, sleeping, and working with computer (the number of samples per second is 64).

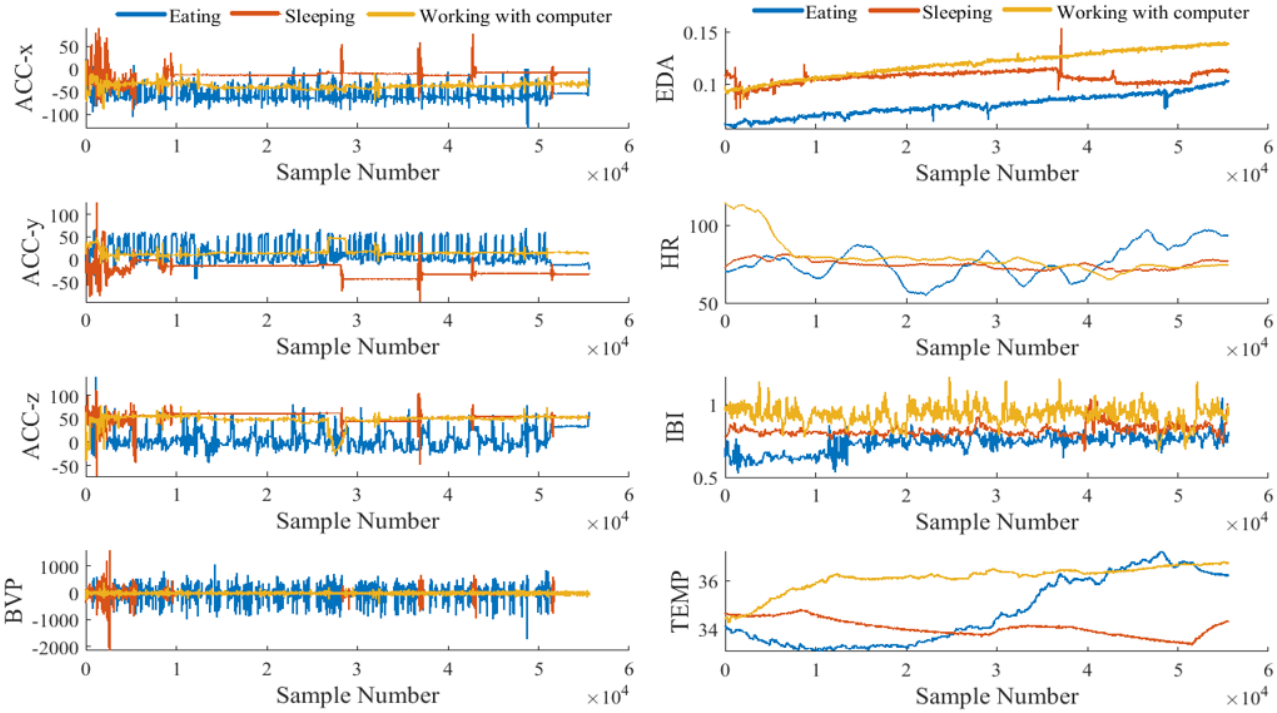


Figure 3. Distribution of data captured by different sensors for a portion of eating episode.

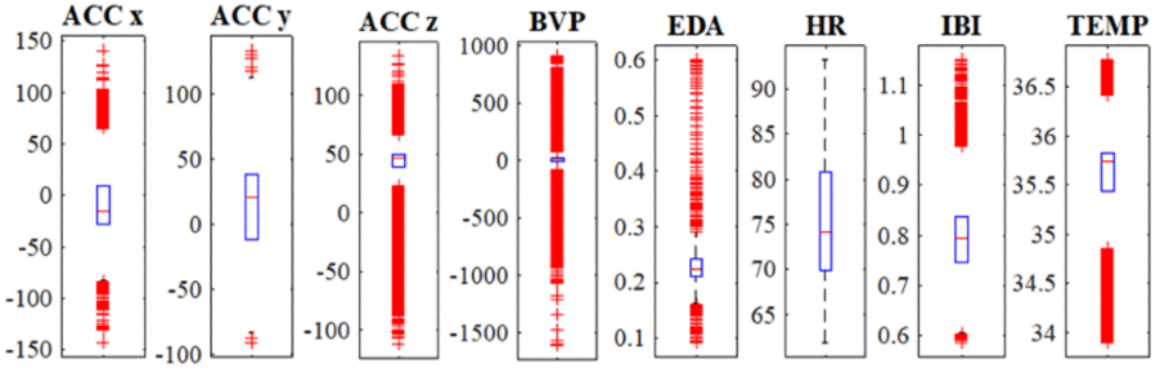


Figure 4 shows the preprocessing steps performed on the raw data to prepare input data for a 2D deep residual network. As shown in Figure 5, covariance coefficients between channels were first derived and presented in the form of contour map. The obtained image was fed to a deep network as input image.

Figure 5 shows examples of covariance coefficients contour generated from different modalities. The horizontal and vertical axes represent the sample number. The value of correlation coefficients is represented by the color, with dark colors corresponding to low values and bright colors corresponding to high values. As seen in figures, there is a visible difference in the color patterns of correlation coefficients contour for different activities.

Figure 4. Generating covariance coefficients contour for a 500-sample eating episode (pairwise covariance was calculated between each sample across all signals).

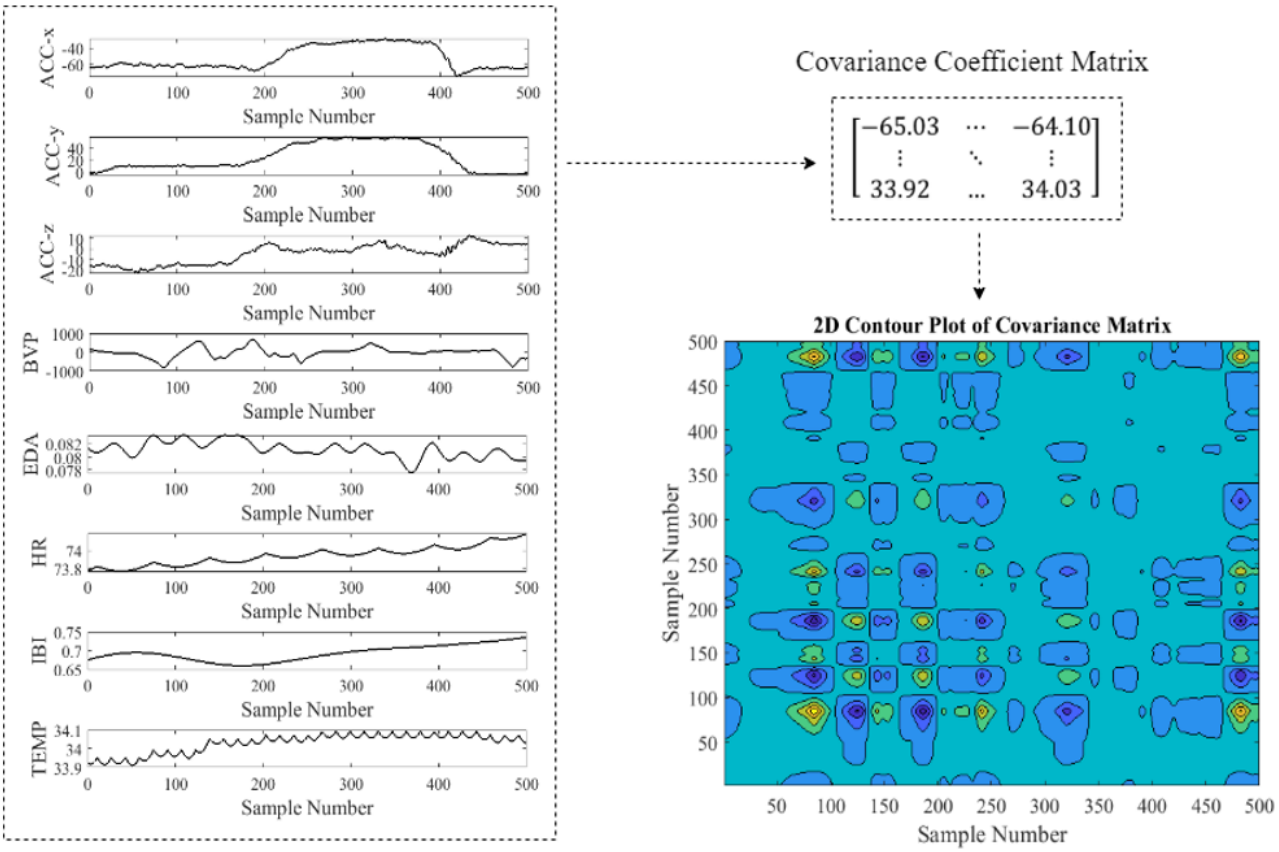
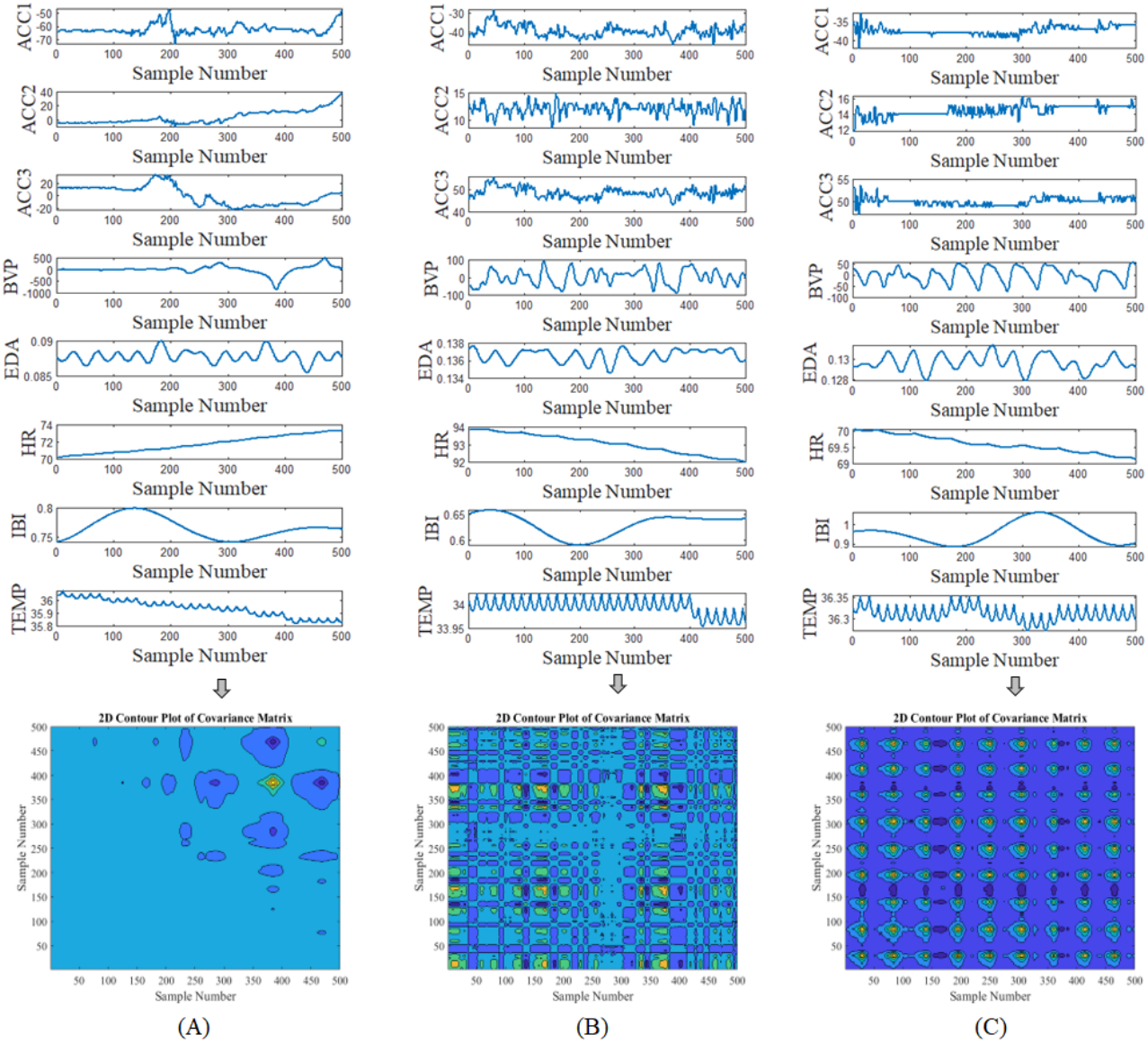
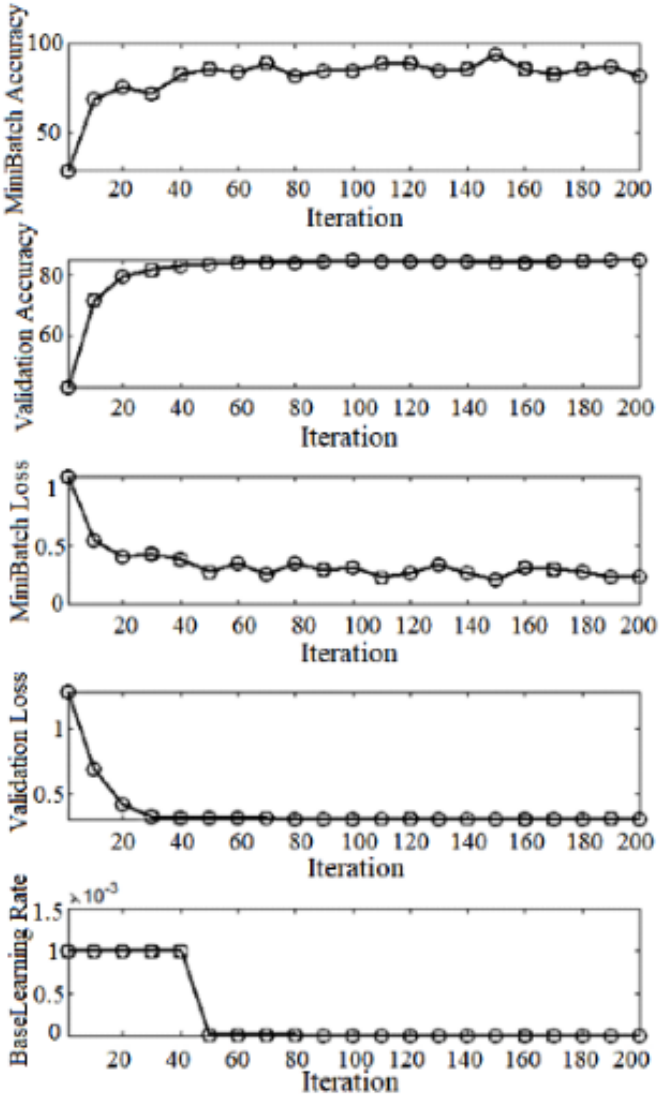


Figure 5. Examples of covariance coefficients contour for different activities. (A) Eating; (B) Working with computer; (C) Sleeping.



The performance metrics variations including epoch number, iteration number, time elapsed, mini-batch accuracy, validation accuracy, and loss function value for the validation data are plotted in Figure 6. The number of epochs was chosen to be 10 over 200 iterations. The training and testing proportions, being considered as 70% and 30%, respectively, were randomly assigned from each label. The training data were also shuffled before every epoch. Learning rate was reduced over epochs and its speed was updated by decreasing the learning rate, and multiplying it by a fractional learn rate drop factor over a specific number of epochs. According to Figure 6, the small validation loss allows to conclude that this method has generalization capability.

Figure 6. Deep learning model performance over observations in the mini-batch.



Convergence of average accuracy and loss function during training and validation for 10 epochs was plotted in Figures 7 and 8. The result of the covariance-based model has rapidly converged to a stable value with no sign of overfitting.

Confusion matrix in Figure 9 shows the results obtained from validation data sets of covariance coefficients contour.

Figure 7. Accuracy variation in each epoch of deep residual network with input images of covariance coefficients contour.

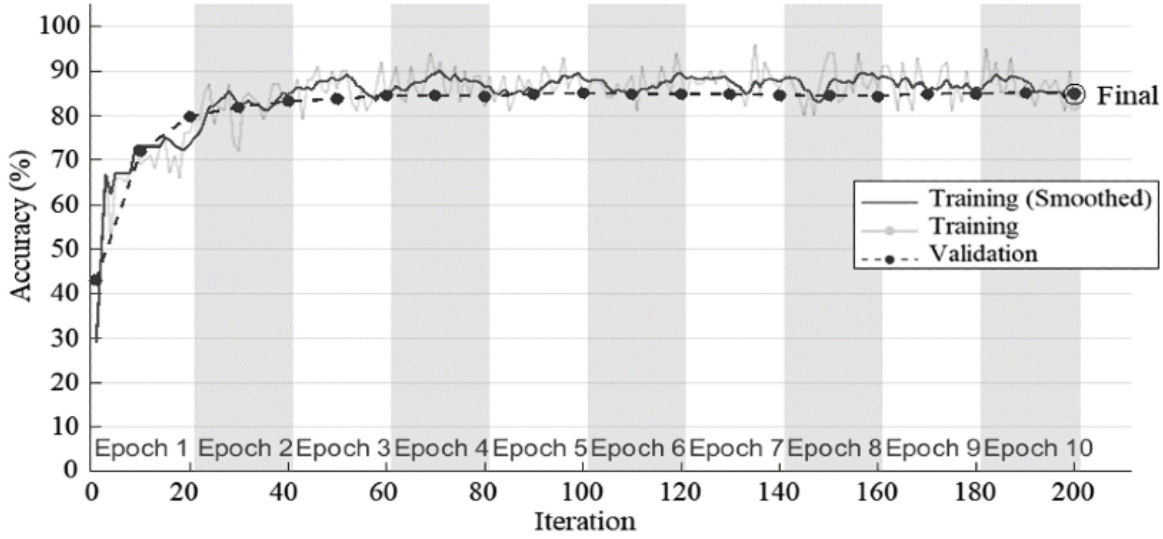


Figure 8. Loss function variation in each epoch of deep residual network with input images of covariance coefficients contour.

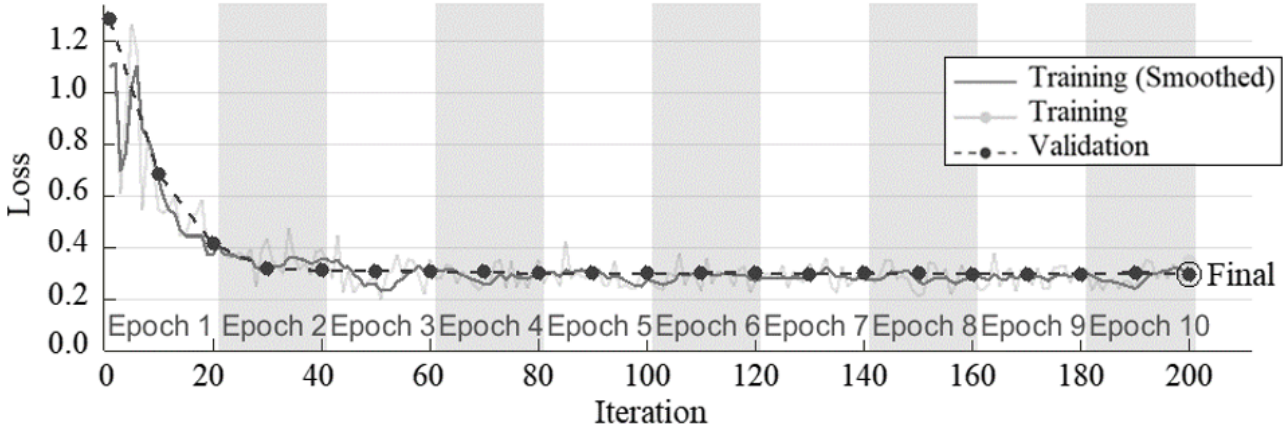
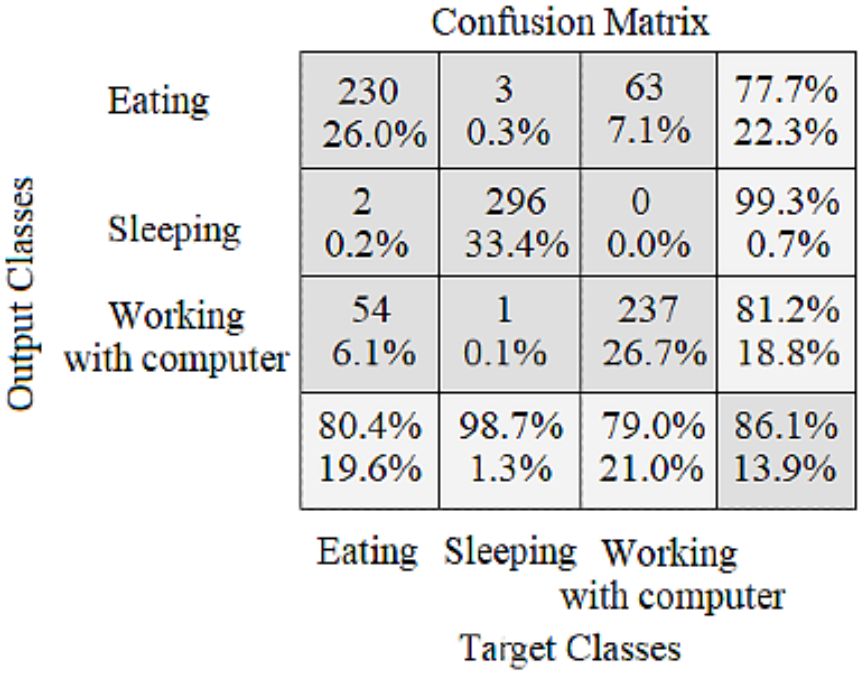


Figure 9. Confusion matrix for the validation set of deep residual network with input images of covariance coefficients contour.



Applying the Proposed Algorithm on the Second Scenario

Figure 10 shows an example of covariance map generated from different modalities in the second scenario.

The visualization of performance of the fusion method applied on the second scenario is plotted in Figure 11 and shows how the algorithm is confusing 2 classes.

The elapsed time for the training process of the model was 39.2714 seconds. The latency for making decision on the new input was 0.1459 seconds. Various statistics calculated from the confusion matrix for a comprehensive study are listed in Table 3. Based on classification results, it is possible to find how well classification of different activities was done by taking advantage of the proposed sensor fusion.

Figure 10. Covariance map for a 50-sample walking episode (pairwise covariance was calculated between each signal across all samples was taken into consideration).

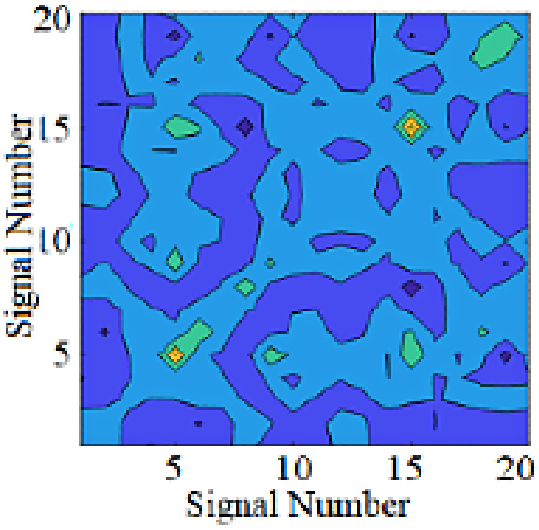


Figure 11. The results of applying pretrained deep learning architecture on the fused data.

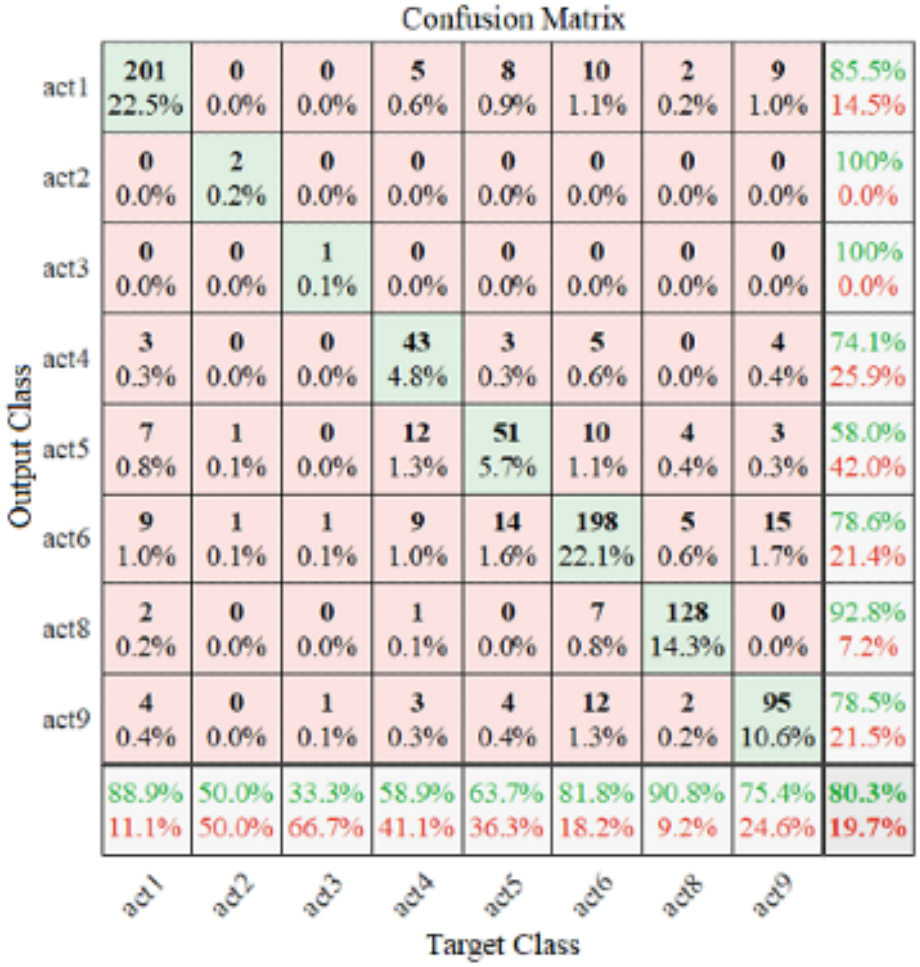


Table 3. Comprehensive study of pretrained deep learning classifier's performance.

Metrics of classifier's performance	Values
Accuracy	0.95083
Precision (positive predictive value)	0.80335
False discovery rate	0.19664
False omission rate	0.02809
Negative predictive value	0.97190
Prevalence	0.12500
Recall (hit rate, sensitivity, true positive rate)	0.80335
False-positive rate (fall-out)	0.02809
Positive likelihood ratio	28.5965
False-negative rate (miss rate)	0.19664
True-negative rate (specificity)	0.97190
Negative likelihood ratio	0.20233
Diagnostic odds ratio	141.334
Informedness	0.77525
Markedness	0.77525
F-score	0.80335
G-measure	0.80335
Matthews correlation coefficient	0.77525

Comparison With Similar Studies

Data captured by different sensors in order to detect eating intervals have been explored by many studies. These studies focused on analyzing several types of data captured by different sensors from accelerometers and gyroscopes to respiratory inductance plethysmography and oral cavity. However, as seen in Table 3, the number of modalities involved in detecting food intake intervals has been up to 2. Therefore, this study investigated whether eating event detection by simultaneous processing of 8 different modalities (eg, 3-axis accelerometer sensor, photoplethysmography, electrodermal activity sensor, interbeat intervals, temperature sensor and heart rate) is feasible. The obtained results showed an overall validation accuracy comparable to the approaches proposed earlier in the literature (Table 4). Furthermore, the proposed data fusion framework in this research provided a simple way of integrating multiple data sources applicable for deep learning methods in human activity monitoring, while previous studies focused on applying raw data. When it comes to cloud computing as well as big data for the purpose of human activity monitoring using wearable sensor-based technologies, the cost of directly applying high-dimensional raw data to a deep classifier would be computationally expensive.

Table 4. Comparison of previous studies on food intake episodes detection.

Study	Modalities	Method	Accuracy, %
[32]	Acoustic signal	Correlation matching	85
[33]	Food image and speech recording	Support vector machine classification	90.6
[34]	Electroglottograph	Artificial neural network	86.6
[35]	Piezoelectricity	Time and amplitude thresholding	86
[36]	Accelerometer and gyroscope	Decision tree classifier	85.5
[37]	Chewing sound	(1) Deep Boltzmann and (2) Machine with deep neural network classifier	77
[38]	Piezoelectricity	Convolutional neural network	91.9
[39]	Acceleration and orientation velocity	Convolutional-recurrent neural network	82.5-96.4

Investigating the Effect of Missing Values

The proposed data fusion technique also has the challenge of data imperfection, which could be overcome by using data imputation methods. Missing samples can affect the contour representation of covariance matrix to a great extent which makes it necessary to be recovered using missing value filling methods.

Figure 12 demonstrates how the moving median method (as a sample of interpolation methods) can recover the contour representation suffered from the missing data. The moving median was done over a window of length 20. However, there is still a huge gap for addressing the issue of data inconsistencies, which will let the gates open for future studies.

Table 5 shows the impact of missing data without using any imputation on performance degradation in the second scenario.

Figure 12. Contour representation (A and C) before and (B and D) after applying moving median for missing data.

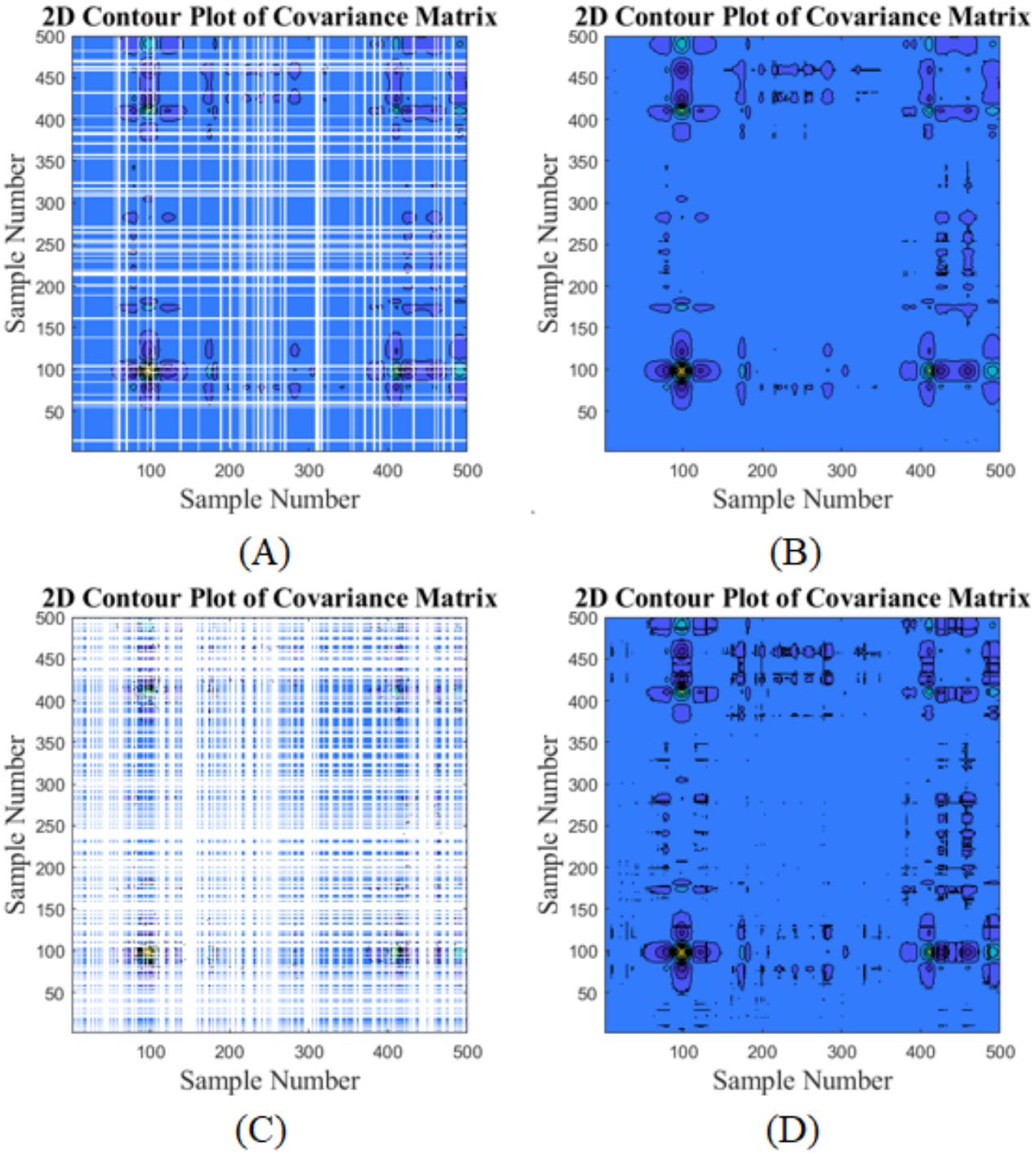


Table 5. Impact of missing data on performance degradation.

Metrics of classifier's performance	Missing value percentage					
	0	1.11	2.22	8.33	15.5	25
Accuracy	0.95083	0.91400	0.89591	0.89533	0.88501	0.88326
Precision (positive predictive value)	0.80335	0.65603	0.58365	0.58132	0.54007	0.53307
False discovery rate	0.19664	0.34396	0.41634	0.41867	0.45992	0.46692
False omission rate	0.02809	0.04913	0.05947	0.05981	0.06570	0.06670
Negative predictive value	0.97190	0.95086	0.94052	0.94018	0.93429	0.93329
Prevalence	0.12500	0.12500	0.12500	0.12500	0.12500	0.12500
Recall (hit rate, sensitivity, true-positive rate)	0.80335	0.65603	0.58365	0.58132	0.54007	0.53307
False positive rate (fall-out)	0.02809	0.04913	0.05947	0.05981	0.06570	0.06670
Positive likelihood ratio	28.5965	13.3506	9.81308	9.71933	8.21996	7.99166
False-negative rate (miss rate)	0.19664	0.34396	0.41634	0.41867	0.45992	0.46692
True-negative rate (specificity)	0.97190	0.95086	0.94052	0.94018	0.93429	0.93329
Negative likelihood ratio	0.20233	0.36174	0.44267	0.44531	0.49226	0.50029
Diagnostic odds ratio	141.334	36.9063	22.1678	21.8259	16.6982	15.9738
Informedness	0.77525	0.60689	0.52418	0.52151	0.47437	0.46637
Markedness	0.77525	0.60689	0.52418	0.52151	0.47437	0.46637
F-score	0.80335	0.65603	0.58365	0.58132	0.54007	0.53307
G-measure	0.80335	0.65603	0.58365	0.58132	0.54007	0.53307
Matthews correlation coefficient	0.77525	0.60689	0.52418	0.52151	0.47437	0.46637

Discussion

Principal Findings

Recent advances in biosensor technologies [40,41] and consumer electronics have led to precise physiological monitoring and more specifically accurate activity recognition. Activity recognition based on wearable device is one of the most rapidly growing research areas in personalization of analyses. Physical characteristics, health state, lifestyle, moving style, and gender are parameters that can be highly personalized. Therefore, in order to consider generalization of prediction or classification models, the data should be labeled personally, and the focus of research should be more on personalized analysis [42,43]. One way to personalize data is automatic identification of human activities and consequently labeling data based on different activities.

Regarding human activity recognition, we are facing upcoming transition from analyzing single modality to processing data collected from multiple sources with enormous diversity in terms of information, size, and behavior. This increases the complexity of classification problems and requires low-level data fusion to simultaneously integrate significant information in all modalities, and yet compressing the data directly at the source. This fusion process is important in a sense that reduction of communication load to other device or to the cloud requires local extraction of information from raw data stream in the sensor level. Therefore, fused raw data in a compressed form are super important in terms of minimizing the amount of data needed to be stored or needed to be sent. It is also important in saving battery power and reducing transmission time.

Regarding the importance of implementing a low-level fusion method with simple structure, this study presented a general framework for implementing efficient fusion based on covariance map. The promising classification results reached precision of 80.3% and showed that global 2D covariance representation can reliably quantify the difference between activities, as it provides a simple abstract representation of correlation over modalities.

For performance assessment of the proposed algorithm, the method was implemented on 2 separate scenarios. These scenarios were different regarding the temporal segment size, type of activity, wearable device, subjects, and deep learning architecture. The obtained results showed the ability of the proposed fusion technique to generalize to other data sets with different modalities, participants, and tasks.

Limitations in Existing Literature

There are many ways of integrating modalities for activity recognition task [44-57], with the 3 major groups being sensor-level [45,49,50], feature-level [44,57], and decision-level [51-56] fusion. Fusion at the decision level is the most frequently used method which takes advantage of training machine learning and deep learning models for each modality. When it comes to merging scores of these networks for the purpose of data fusion, the practical applications are limited by their complex process which lead to more computationally heavy processing and make them inapplicable for implementing on low-power systems.

Therefore, such techniques cannot be considered as low-level data fusion.

There are only a few methods for low-level data fusion, with 2 focused on using Bayesian network for sensor-level fusion [45,50]. These Bayesian networks usually involve a time-consuming process of hyperparameter tuning. Correctly implementing hyperparameter optimization is usually complex and computationally expensive. A small change in the values of hyperparameters can highly affect the performance of model. In this sense, it could be said that the low-level fusion methods with simple structure can outperform sophisticated ones.

Strengths of the Proposed Method

Unlike complex fusion techniques based on evolutionary computation [48], machine learning approaches [47,51,53,54], Bayesian models [44,45,50,55], Kalman filtering [49], and neural networks [46], the proposed fusion method had very simple implementing procedure, and yet capable of revealing the common trends and similarities among recorded modalities. It was also free of the number and type of sensors used for collecting data. This could be an important benefit as there is high diversity in sensor technology deployed for activity recognition and the choice of sensors vary a lot from one case to another. The sensors used in activity recognition studies include vibration and contact sensor [44], tap sensor [45], motion sensor [46], ventilation sensor [47], heart rate sensor [48], magnetometer sensor [49], temperature sensor [50], electrocardiogram sensor [51], accelerometer sensor [52], and gyroscope [54]. Furthermore, this method is universal in a sense that it can cover a wide spectrum of problems including tracking activity of daily living, elderly monitoring, fall detection, smart home, ambient assisted living, behavior analysis, among others. It could help in decreasing the final cost of the monitoring framework by deploying fusion in the first step of classification process (applying fusion algorithm in the final step needs independently analyzing data for every single component and combining the final results which make the implementation computationally expensive). Providing the possibility of visually representing the correlation among modalities and reducing dimension by embedding the sensory data in just a single 2D representation can be considered as other strengths of the studied technique.

Future Work

A limitation to this research is that both tested scenarios were performed on healthy volunteers, which may be far from the cases including actual patients who suffer from movement disability or major health problem. This could be included in future work.

Future research direction will also include implementing the fusion algorithms for the scenarios in which one or more modalities are missing. The applicability of the findings will be also tested for other problems rather than activity recognition.

Acknowledgments

This work was supported by a grant (No. 308935) from the Academy of Finland and Infotech, and Orion Research Foundation sr, Oulu, Finland.

Conflicts of Interest

None declared.

References

1. van den Boer J, van der Lee A, Zhou L, Papapanagiotou V, Diou C, Delopoulos A, et al. The SPLENDID Eating Detection Sensor: Development and Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e170. [doi: 10.2196/mhealth.9781] [Medline: 30181111]
2. Selamat NA, Ali SHM. Automatic Food Intake Monitoring Based on Chewing Activity: A Survey. IEEE Access 2020;8:48846-48869. [doi: 10.1109/access.2020.2978260]
3. Bell BM, Alam R, Alshurafa N, Thomaz E, Mondol AS, de la Haye K, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. NPJ Digit Med 2020 Mar 13;3(1):38. [doi: 10.1038/s41746-020-0246-2] [Medline: 32195373]
4. Chun KS, Bhattacharya S, Thomaz E. Detecting Eating Episodes by Tracking Jawbone Movements with a Non-Contact Wearable Sensor. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-21. [doi: 10.1145/3191736]
5. Huang Q, Wang W, Zhang Q. Your Glasses Know Your Diet: Dietary Monitoring Using Electromyography Sensors. IEEE Internet Things J 2017 Jun;4(3):705-712. [doi: 10.1109/jiot.2017.2656151]
6. Prioleau T, Moore E, Ghovanloo M. Unobtrusive and Wearable Systems for Automatic Dietary Monitoring. IEEE Trans. Biomed. Eng 2017 Sep;64(9):2075-2089. [doi: 10.1109/tbme.2016.2631246]
7. Pasler S, Fischer W. Food Intake Monitoring: Automated Chew Event Detection in Chewing Sounds. IEEE J. Biomed. Health Inform 2014 Jan;18(1):278-289. [doi: 10.1109/jbhi.2013.2268663]
8. Amft O. A wearable earpad sensor for chewing monitoring. New York, NY: IEEE; 2010 Nov 1 Presented at: SENSORS, 2010 IEEE; November 1-4, 2010; Waikoloa, HI. [doi: 10.1109/jsen.2011.2172411]

9. Farooq M, Sazonov E. Segmentation and Characterization of Chewing Bouts by Monitoring Temporalis Muscle Using Smart Glasses With Piezoelectric Sensor. IEEE J. Biomed. Health Inform 2017 Nov;21(6):1495-1503. [doi: 10.1109/jbhi.2016.2640142]
10. Zhang R, Amft O. Monitoring Chewing and Eating in Free-Living Using Smart Eyeglasses. IEEE J. Biomed. Health Inform 2018 Jan;22(1):23-32. [doi: 10.1109/jbhi.2017.2698523]
11. Fan D, Gong J, Lach J. Eating gestures detection by tracking finger motion. New York, NY: IEEE; 2016 Presented at: 2016 IEEE Wireless Health (WH); October 25-27, 2016; Bethesda, MD p. 1-6. [doi: 10.1109/WH.2016.7764549]
12. Parate A, Ganesan D. Detecting eating and smoking behaviors using smartwatches. In: Rehg M, Murphy SA, Kumar S, editors. Mobile Health. Cham, Switzerland: Springer; 2017:175-201.
13. Heydarian H, Adam M, Burrows T, Collins C, Rollo ME. Assessing Eating Behaviour Using Upper Limb Mounted Motion Sensors: A Systematic Review. Nutrients 2019 May 24;11(5):1168 [FREE Full text] [doi: 10.3390/nu11051168] [Medline: 31137677]
14. Zhang R, Amft O. Free-living eating event spotting using EMG-monitoring eyeglasses. New York, NY: IEEE; 2018 Presented at: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); March 4-7, 2018; Las Vegas, NV p. 128-132. [doi: 10.1109/bhi.2018.8333386]
15. Ye X, Chen G, Gao Y, Wang H, Cao Y. Assisting food journaling with automatic eating detection, in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 3255?3262, 2016. New York, NY: ACM; 2016 Presented at: CHI EA '16: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 2016; San Jose, CA p. 3255-3262. [doi: 10.1145/2851581.2892426]
16. Thomaz E, Essa I, Abowd G. A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:1029-1040 [FREE Full text] [doi: 10.1145/2750858.2807545] [Medline: 29520397]
17. Schiboni G, Wasner F, Amft O. A Privacy-Preserving Wearable Camera Setup for Dietary Event Spotting in Free-Living. New York, NY: IEEE; 2018 Mar 19 Presented at: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 19-23, 2018; Athens, Greece. [doi: 10.1109/percomw.2018.8480222]
18. Sharma S, Jasper P, Muth E, Hoover A. Automatic Detection of Periods of Eating Using Wrist Motion Tracking. New York, NY: IEEE; 2016 Jun 27 Presented at: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Washington, DC, USA p. A. [doi: 10.1109/chase.2016.15]
19. Sen S, Subbaraju V, Misra A, Balan R, Lee Y. Annapurna: Building a Real-World Smartwatch-Based Automated Food Journal. New York, NY: IEEE; 2018 Jun 12 Presented at: 2018 IEEE 19th International Symposium on; June 12-15, 2018; Chania, Greece. [doi: 10.1109/wowmom.2018.8449755]
20. Navarathna P, Bequette B. Wearable Device Based Activity Recognition and Prediction for Improved Feedforward Control. 2018 Jun 27 Presented at: 2018 Annual American Control Conference (ACC); June 27-29, 2018; Milwaukee, WI. [doi: 10.23919/acc.2018.8430775]
21. Mirtchouk M, Lustig D, Smith A, Ching I, Zheng M, Kleinberg S. Recognizing Eating from Body-Worn Sensors. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-20. [doi: 10.1145/3131894]
22. Sazonov ES, Fontana JM. A Sensor System for Automatic Detection of Food Intake Through Non-Invasive Monitoring of Chewing. IEEE Sensors J 2012 May;12(5):1340-1348. [doi: 10.1109/jsen.2011.2172411]
23. Lopez-Meyer P, Schuckers S, Makeyev O, Sazonov E. Detection of periods of food intake using Support Vector Machines. New York, NY: IEEE; 2010 Aug 31 Presented at: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology; August 31 to September 4, 2010; Buenos Aires, Argentina. [doi: 10.1109/chase.2016.14]
24. Bedri A, Verlekar A, Thomaz E, Avva V, Starner T. Detecting Mastication: A Wearable Approach. New York, NY: ACM; 2015 Nov 13 Presented at: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction; November 13, 2015; Seattle, WA p. 247. [doi: 10.1145/2818346.2820767]
25. Papapanagiotou V, Diou C, Zhou L, van den Boer J, Mars M, Delopoulos A. A Novel Chewing Detection System Based on PPG, Audio, and Accelerometry. IEEE J. Biomed. Health Inform 2017 May;21(3):607-618. [doi: 10.1109/jbhi.2016.2625271]
26. Chung J, Chung J, Oh W, Yoo Y, Lee WG, Bang H. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. Sci Rep 2017 Jan 30;7(1):41690 [FREE Full text] [doi: 10.1038/srep41690] [Medline: 28134303]
27. Farooq M, Fontana J. A Comparative Study of Food Intake Detection Using Artificial Neural Network and Support Vector Machine. 2013 Dec 04 Presented at: 12th International Conference on Machine Learning and Applications; December 4-7, 2013; Miami, FL. [doi: 10.1109/icmla.2013.33]
28. Bedri A, Li R, Haynes M, Kosaraju RP, Grover I, Prioleau T, et al. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-20 [FREE Full text] [doi: 10.1145/3130902] [Medline: 30135957]
29. Rouast PV, Adam MTP. Learning Deep Representations for Video-Based Intake Gesture Detection. IEEE J. Biomed. Health Inform 2020 Jun;24(6):1727-1737. [doi: 10.1109/jbhi.2019.2942845]

30. Ciocca CG, Napoletano P, Schettini R. Learning CNN-based Features for Retrieval of Food Images. In: Battiato S, Farinella G, Leo M, Gallo G, editors. New Trends in Image Analysis and Processing – ICIAP 2017. Cham, Switzerland: Springer; 2017:426-434.
31. Ruzzon M, Carfì A, Ishikawa T, Mastrogiovanni F, Murakami T. A multi-sensory dataset for the activities of daily living. Data Brief 2020 Oct;32:106122 [FREE Full text] [doi: 10.1016/j.dib.2020.106122] [Medline: 32904359]
32. Olubanjo T, Moore E, Ghovanloo M. Detecting food intake acoustic events in noisy recordings using template matching. 2016 Feb 24 Presented at: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); February 24-27, 2016; Las Vegas, NV. [doi: 10.1109/bhi.2016.7455916]
33. Puri M, Zhu Z, Yu Q, Divakaran A, Sawhney H. Recognition and volume estimation of food intake using a mobile device. 2009 Dec 07 Presented at: 2009 Workshop on Applications of Computer Vision (WACV); December 7-8, 2009; Snowbird, UT. [doi: 10.1109/wacv.2009.5403087]
34. Farooq M, Fontana JM, Sazonov E. A novel approach for food intake detection using electroglottography. Physiol Meas 2014 May 26;35(5):739-751 [FREE Full text] [doi: 10.1088/0967-3334/35/5/739] [Medline: 24671094]
35. Sen S, Subbaraju V, Misra R, Balan K, Lee Y. The case for smartwatch-based diet monitoring. New York, NY: IEEE; 2015 Mar 23 Presented at: 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops); March 23-27, 2015; St. Louis, MO p. 585-590. [doi: 10.1109/percomw.2015.7134103]
36. Gao Y, Zhang N, Wang H, Ding X, Ye X, Chen G, et al. iHear foodating detection using commodity Bluetooth headsets. New York, NY: IEEE; 2016 Presented at: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Washington, DC p. 172. [doi: 10.1109/CHASE.2016.14]
37. Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems (Volume 2). Red Hook, NY: Curran Associates Inc; Dec 2012:2222-2230.
38. Hussain G, Maheshwari MK, Memon ML, Jabbar MS, Javed K. A CNN Based Automated Activity and Food Recognition Using Wearable Sensor for Preventive Healthcare. Electronics 2019 Nov 29;8(12):1425. [doi: 10.3390/electronics8121425]
39. Kyritsis K, Diou C, Delopoulos A. A Data Driven End-to-End Approach for In-the-Wild Monitoring of Eating Behavior Using Smartwatches. IEEE J. Biomed. Health Inform 2021 Jan;25(1):22-34. [doi: 10.1109/jbhi.2020.2984907]
40. Asgari S, Fabritius T. Tunable Mid-Infrared Graphene Plasmonic Cross-Shaped Resonator for Demultiplexing Application. Applied Sciences 2020 Feb 10;10(3):1193. [doi: 10.3390/app10031193]
41. Asgari S, Granpayeh N, Fabritius T. Controllable terahertz cross-shaped three-dimensional graphene intrinsically chiral metastructure and its biosensing application. Optics Communications 2020 Nov;474:126080. [doi: 10.1016/j.optcom.2020.126080]
42. Li K, Habre R, Deng H, Urman R, Morrison J, Gilliland FD, et al. Applying Multivariate Segmentation Methods to Human Activity Recognition From Wearable Sensors' Data. JMIR Mhealth Uhealth 2019 Feb 07;7(2):e11201 [FREE Full text] [doi: 10.2196/11201] [Medline: 30730297]
43. Ferrari A, Micucci D, Mobilio M, Napoletano P. On the Personalization of Classification Models for Human Activity Recognition. IEEE Access 2020;8:32066-32079. [doi: 10.1109/access.2020.2973425]
44. Lu CH, Fu LC. Robust Location-Aware Activity Recognition Using Wireless Sensor Network in an Attentive Home. IEEE Trans. Automat. Sci. Eng 2009 Oct;6(4):598-609. [doi: 10.1109/tase.2009.2021981]
45. Liao J, Bi Y, Nugent C. Using the Dempster–Shafer Theory of Evidence With a Revised Lattice Structure for Activity Recognition. IEEE Trans. Inform. Technol. Biomed 2011 Jan;15(1):74-82. [doi: 10.1109/titb.2010.2091684]
46. Zhu C, Sheng W. Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living. IEEE Trans. Syst., Man, Cybern. A 2011 May;41(3):569-573. [doi: 10.1109/tsmca.2010.2093883]
47. Liu S, Gao RX, John D, Staudenmayer JW, Freedson PS. Multisensor Data Fusion for Physical Activity Assessment. IEEE Trans. Biomed. Eng 2012 Mar;59(3):687-696. [doi: 10.1109/tbme.2011.2178070]
48. Chernbumroong S, Cang S, Yu H. Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People. IEEE J. Biomed. Health Inform 2015 Jan;19(1):282-289. [doi: 10.1109/jbhi.2014.2313473]
49. Zhao H, Wang Z. Motion Measurement Using Inertial Sensors, Ultrasonic Sensors, and Magnetometers With Extended Kalman Filter for Data Fusion. IEEE Sensors J 2012 May;12(5):943-953. [doi: 10.1109/jsen.2011.2166066]
50. De Paola A, Ferraro P, Gaglio S, Re GL, Das SK. An Adaptive Bayesian System for Context-Aware Data Fusion in Smart Environments. IEEE Trans Mobile Comput 2017 Jun 1;16(6):1502-1515. [doi: 10.1109/tmc.2016.2599158]
51. Peng L, Chen L, Wu X, Guo H, Chen G. Hierarchical Complex Activity Representation and Recognition Using Topic Model and Classifier Level Fusion. IEEE Trans Biomed Eng 2017 Jun;64(6):1369-1379. [doi: 10.1109/tbme.2016.2604856]
52. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG. Physical Activity Recognition Using Posterior-Adapted Class-Based Fusion of Multiaccelerometer Data. IEEE J. Biomed. Health Inform 2018 May;22(3):678-685. [doi: 10.1109/jbhi.2017.2705036]
53. Jain A, Kanhangad V. Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors. IEEE Sensors J 2018 Feb 1;18(3):1169-1177. [doi: 10.1109/jsen.2017.2782492]
54. Wang Y, Cang S, Yu H. A Data Fusion-Based Hybrid Sensory System for Older People's Daily Activity and Daily Routine Recognition. IEEE Sensors J 2018 Aug 15;18(16):6874-6888. [doi: 10.1109/jsen.2018.2833745]

55. Acharya S, Mongan WM, Rasheed I, Liu Y, Anday E, Dion G, et al. Ensemble Learning Approach via Kalman Filtering for a Passive Wearable Respiratory Monitor. IEEE J. Biomed. Health Inform 2019 May;23(3):1022-1031. [doi: 10.1109/jbhi.2018.2857924]
56. Guo M, Wang Z, Yang N, Li Z, An T. A Multisensor Multiclassifier Hierarchical Fusion Model Based on Entropy Weight for Human Activity Recognition Using Wearable Inertial Sensors. IEEE Trans. Human-Mach. Syst 2019 Feb;49(1):105-111. [doi: 10.1109/thms.2018.2884717]
57. Muaaz M, Chelli A, Abdelgawwad AA, Mallofre AC, Patzold M. WiWeHAR: Multimodal Human Activity Recognition Using Wi-Fi and Wearable Sensing Modalities. IEEE Access 2020;8:164453-164470. [doi: 10.1109/access.2020.3022287]

Edited by L Buis; submitted 29.06.20; peer-reviewed by C Diou, D Goyal; comments to author 01.10.20; revised version received 10.11.20; accepted 18.12.20; published 28.01.21.

Please cite as:
Bahador N, Ferreira D, Tamminen S, Kortelainen J
Deep Learning–Based Multimodal Data Fusion: Case Study in Food Intake Episodes Detection Using Wearable Sensors
JMIR Mhealth Uhealth 2021;9(1):e21926
URL: http://mhealth.jmir.org/2021/1/e21926/
doi:10.2196/21926
PMID:33507156

Original Paper

Effects of Activity Tracker Use With Health Professional Support or Telephone Counseling on Maintenance of Physical Activity and Health Outcomes in Older Adults: Randomized Controlled Trial

Katie-Jane Brickwood[1], PhD; Kiran D K Ahuja[1], PhD; Greig Watson[1], PhD; Jane A O'Brien[2], PhD; Andrew D Williams[1], PhD

[1]School of Health Sciences, College of Health and Medicine, University of Tasmania, Launceston, Australia

[2]School of Nursing, College of Health and Medicine, University of Tasmania, Launceston, Australia

Corresponding Author:
Katie-Jane Brickwood, PhD
School of Health Sciences
College of Health and Medicine
University of Tasmania
Newnham Drive
Launceston, 7250
Australia
Phone: 61 0363245487
Email: katiejane.brickwood@utas.edu.au

Abstract

Background: Despite a range of efforts to increase physical activity participation in Australia, inactivity levels in older adults have remained high over recent decades, contributing to increased rates of chronic health conditions. Lifestyle interventions, including telephone counseling (TC), improve physical activity participation and associated health outcomes over the short term; however, ongoing feedback and support is required to maintain these changes. Newer technologies such as wearable activity trackers (ATs) may offer an alternative method for providing ongoing support.

Objective: This study aims to investigate whether newer technologies such as wearable ATs assist in providing ongoing support to maintain physical activity levels and health outcomes.

Methods: Older adults aged >60 years who had just completed a 12-week face-to-face individualized community exercise program in Tasmania, Australia, participated in the study. They were randomized to receive AT, TC, or usual care (UC). All groups received a home exercise program and an optional referral to a community-based exercise program. The AT group also received an AT and text message feedback from an accredited exercise physiologist (AEP). The TC group received phone calls from an AEP throughout the 12-month intervention. The primary outcome was daily steps measured by an ActivPAL (TM) accelerometer at baseline and at 3, 6, and 12 months. Secondary outcome measures included body composition, blood pressure, 10-time sit-to-stand (TTSTS) test, timed up and go test, and cardiorespiratory fitness. This trial was approved by the Tasmanian Health and Medical Human Research Ethics Committee (H0014713).

Results: A total of 117 participants were randomized to the study (AT, n=37; TC, n=38; UC, n=42). At baseline, the participants (75/117, 64.1% female; mean age 72.4 years, SD 6.4) completed an average of 6136 steps (SD 2985) per day. Although there were no significant differences between groups, the TC and AT groups maintained daily step counts (mean difference [MD] −79 steps, 95% CI −823 to 663 steps; *P*=.81; and MD −588 steps, 95% CI −1359 to 182 steps; *P*=.09), and UC showed a reduction in daily steps (MD 981 steps, 95% CI −1668 to −294 steps; *P*=.003) during the 12-month period. Diastolic blood pressure was significantly higher after AT than after UC (MD 5.62 mm Hg, 95% CI 1.30 to 9.94 mm Hg; *P*=.01), and TTSTS was significantly slower on TC compared with UC (MD 2.36 seconds, 95% CI −0.14 to 4.87 seconds; *P*=.03).

Conclusions: The use of an AT with AEP support or TC is effective at maintaining daily step count in older adults over a 12-month period, suggesting that wearable ATs are as effective as TC. Further research to investigate which option is more cost-effective would be beneficial.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12615001104549; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=369118

(JMIR Mhealth Uhealth 2021;9(1):e18686) doi:10.2196/18686

KEYWORDS

physical activity; fitness trackers; telemedicine; feedback; older adults; eHealth; mobile phone

Introduction

Background

Appropriate and ongoing support is needed to assist older adults to engage in regular physical activity to minimize the functional decline and loss of independence associated with aging [1]. Structured lifestyle interventions are effective at increasing physical activity participation and improving strength and functional capacity in older adults [2,3]. Traditionally, structured lifestyle interventions use education sessions, behavior change techniques (BCTs), and self-monitoring [4]. Although these methods provide an initial increase in physical activity participation, physical activity levels tend to revert to preintervention levels once the structured intervention finishes [5,6]. As physical activity participation needs to be maintained to preserve the associated health benefits, effective strategies to assist older adults to continue to be physically active following a lifestyle intervention are required.

Several systematic reviews have assessed methods of maintaining physical activity participation [7,8]. Telephone counseling (TC) has been shown to be effective as both a booster strategy [7] and an intervention to provide ongoing support to promote habitual behavior change [8]. Evidence suggests that longer duration interventions with more regular phone calls demonstrate greater effectiveness [8]. Despite the established benefits of TC, there are significant time and resource barriers to its implementation in standard practice [9].

Activity trackers (ATs) provide consumers with the ability to objectively monitor and receive feedback relating to daily physical activity and can provide health professionals with an objective measure, allowing for the provision of targeted feedback and ongoing support [10]. Furthermore, ATs may offer a less resource-intensive alternative to TC. The use of an AT, particularly when included as part of a broader behavioral intervention, has been shown to be effective at increasing physical activity levels in different populations [11]. There is a paucity of long-term interventions (>6 months) in relation to older adults. In addition, few studies have compared the use of an AT as a stand-alone intervention compared with other established behavioral or lifestyle interventions [12,13].

Objectives

Consequently, this study aimed to examine the effects of a commercially available AT with usual care (UC) and an established method of postintervention follow-up (TC) to assist older adults in maintaining their daily step count over a 12-month period. It was hypothesized that older adults in both the AT and TC groups would maintain physical activity levels over the 12-month intervention compared with those in the UC group.

Methods

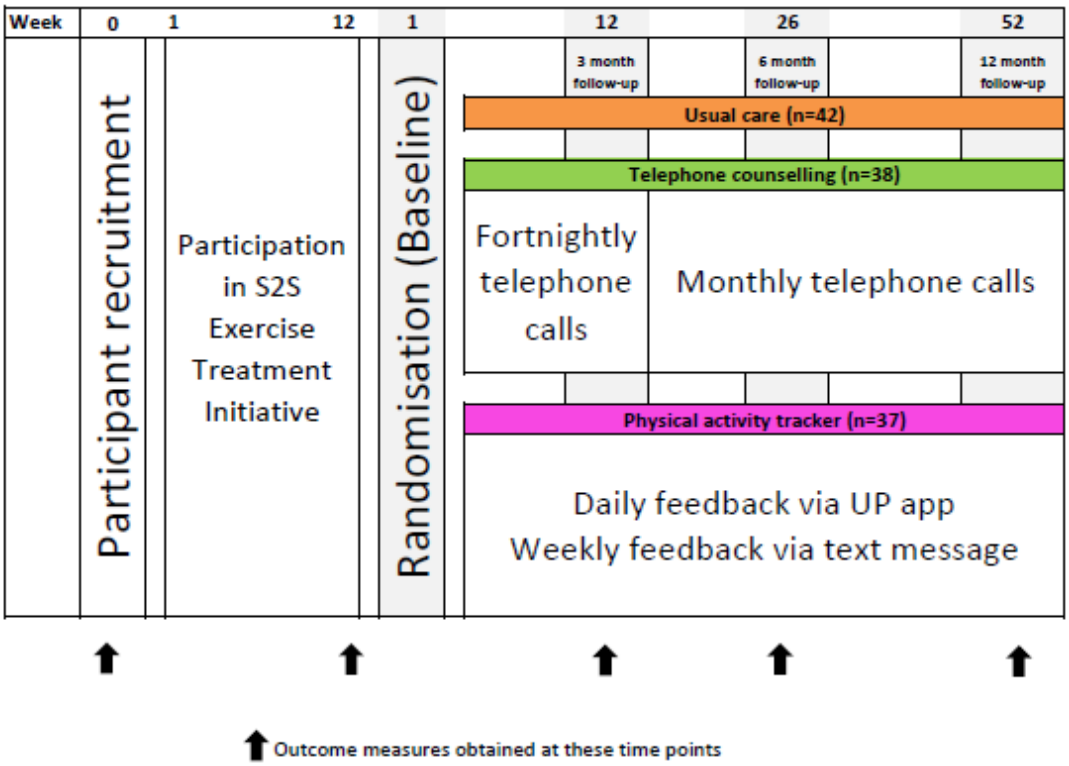
The study was a three-arm, 12-month randomized controlled trial (RCT) investigating the use of an AT, TC, and UC in the maintenance of physical activity levels and health outcomes in older adults. The study was approved by the Tasmanian Health and Medical Human Research Ethics Committee (H0014713) and was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12615001104549). This study was reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) checklist [14].

Recruitment and Randomization

To be eligible, participants had to have completed the 12-week Strength2Strength (S2S) Tasmania Exercise Treatment Initiative, be above the age of 60 years, and have or be at risk of developing a chronic medical condition. The S2S program was a community-based 12-week exercise and education program led by accredited exercise physiologists (AEPs) in Tasmania, Australia. The S2S program provided participants with individually tailored exercise programs to suit their health conditions and goals. The primary focus of the exercise program was strength-based exercises; however, a range of cardiovascular, strength, and balance exercises were included. Participants' exercise prescriptions were reviewed weekly and progressed as required. In addition to attending the S2S program, participants were encouraged to complete an adapted version of their exercises for the home environment at least twice a week.

The RCT was a community intervention, with assessments conducted at the University of Tasmania exercise clinic. Participants were excluded from the study if they chose not to participate in the S2S program, had an unstable medical condition that prevented them from participating in regular physical activity, had a neurological condition, or had a limited understanding of English, which prevented them from meeting the self-reporting requirements of the study. All outcome measures were collected at the start and the end of the S2S program and at the 3-, 6-, and 12-month follow-up (Figure 1). Data collected at the start of the S2S program were not included in this analysis, as they were collected to inform a health economic analysis that will be reported elsewhere. Additional details regarding the S2S initiative and the design of the larger study have been published elsewhere as a protocol paper [15].

Figure 1. Study design showing time points for recruitment, randomization, and all data collection. All intervention groups included usual care. UC: usual care.



All participants were provided with information regarding the study at the beginning of the S2S program. The lead researcher answered questions and completed the consent process with interested participants. Randomization was performed using computer-generated blocks of 15 by a third person not directly involved in the study and was recorded in sealed opaque envelopes with envelopes opened sequentially at the end of the S2S program to reveal the intervention allocation. Participants from the same household were randomized to the same intervention group to maintain intervention integrity (a total of 16 participants). Participants were asked to avoid using physical activity monitoring devices other than those directly provided by researchers for the duration of the study.

Intervention Groups

The UC group received standard care, which included the provision of an individualized home-based exercise program. The home program included similar exercises to those prescribed during the S2S program but modified for the home environment and for any equipment that was available to the participant. An optional referral to a range of community-based physical activity programs was also offered.

In addition to receiving UC, participants randomized to AT were provided with a Jawbone UP24 (TM; Jawbone, Inc) AT and ZTE (TM) mobile device and data plan. Participants who already had a compatible smartphone could choose to use the mobile device provided or their own device. No reimbursement for data costs was provided to those who chose to use their own smartphone. The device was worn on the nondominant wrist for the duration of the 12-month intervention. Participants were provided with an individual information session on how to use, pair, and charge the device at the time of randomization to the AT group. Written instructions and telephone support were provided to troubleshoot any technical issues. A separate Jawbone account was created for each participant, allowing the AEP to remotely access each participant's daily step data. The Jawbone UP24 (TM) was paired to synchronize with the Jawbone UP (TM) app on the mobile device. A daily step goal was individually prescribed for each participant based on their physical function and current level of physical activity. This daily step goal was programmed into the UP (TM) app at the time of randomization, with automated feedback provided by the tracker and app based on this daily step goal. Participants were asked to synchronize the tracker with the UP (TM) app at the end of each day but could check the progress toward their daily step goal as desired. In addition to daily feedback available through the UP (TM) app, participants received weekly, personalized text messages from an AEP. The text message contained feedback related to average daily steps and a comparison of their daily step goal with that of the previous week. The daily step goal was slightly adjusted (±200-500 steps) during the weekly text message feedback from the AEP based on the previous week's step data. The daily step goal in the UP (TM) app however remained the same, as this could not be adjusted remotely. If participants continued to significantly underachieve or exceed their initial daily step goal, the UP (TM) app was adjusted during their 3-, 6-, or 12-month assessment. An example of the text messages sent to participants by the AEP is provided in Multimedia Appendix 1.

Participants randomized to the TC group received UC and a physical activity counseling phone call once a fortnight for the first 3 months and once a month for the remaining 9 months of the intervention. Phone calls were delivered by an AEP experienced in motivational interviewing techniques [16],

following set protocols to determine the participants' self-reported physical activity levels. The protocol allowed the AEP to offer tailored support and advice regarding exercise prescription and modification and assist participants to identify and address any barriers limiting their physical activity participation [15]. Participants were asked to self-report activity levels and compliance with their home-based exercise program and to identify any issues preventing them from being physically active. Constructs from the Social Cognitive Theory [17], including the use of goal setting and self-monitoring, the provision of feedback, and motivational interviewing techniques (eg, affirming, reflective listening, summarizing, and informing and advising) to improve self-efficacy were used during phone calls. Participants in the TC group were asked to refrain from using a wearable AT for the duration of the intervention.

Outcome Measures

The primary outcome measure was physical activity participation in the form of a daily step count measured by an ActivPAL (TM) accelerometer (PAL Technologies Ltd), which has been shown to be a valid and reliable device for monitoring physical activity in older adults and individuals with an altered walking gait. The ActivPAL (TM) was enclosed in a small flexible sleeve to cover the monitor and fitted to the front of the thigh using a Tegaderm (TM) film to allow participants to perform their usual daily activities (including showering or bathing) with the device in place. The ActivPAL (TM) was worn day and night for a 7-day period with a minimum of 5 days' worth of valid data required for inclusion at each assessment time. Data files from the ActivPAL (TM) were downloaded from the devices, and event files were created using a proprietary software (ActivPAL3, version 7.2.38; PAL Technologies). Event files were analyzed using a custom software (National Instruments Labview 2017) to determine daily step counts and total nonstepping time. Only days with full recordings were considered. Daily step counts were averaged for each assessment time point. Nonstepping time was calculated by summing the time spent lying, sitting, and standing over a 24-hour period. Self-reported physical activity (total minutes of activity and metabolic equivalent [MET] per minute per week) were obtained using the Active Australia Survey (AAS) [18]. The AAS assesses leisure time physical activity and includes the number of sessions and total time spent in planned walking, vigorous-intensity gardening or housework, and planned moderate- or vigorous-intensity physical activity. The AAS has been demonstrated to be valid in community-dwelling older adults [19]. The secondary outcomes included health risk factors, functional measures, and quality of life. Health risk factors were measured using standardized protocols and included body weight, BMI, and systolic blood pressure (SBP) and diastolic blood pressure (DBP) [15]. Body fat percentage (BF%) and lean mass (LM) were assessed using bioimpedance analysis scales [20]. LM was reported in kilograms rather than percentage to provide an absolute measure of LM, which may change relative to other body composition factors. The 10-time sit-to-stand (TTSTS) test was used to measure functional lower body strength [21] and the timed up and go (TUAG) test was used to assess dynamic balance and mobility [22], both of which relate to the ability to perform activities of daily living. The 6-min walk test (6MWT) [23] was used to assess cardiorespiratory fitness for participants who had mobility issues, including walking with aids. This is a self-paced test, requiring participants to walk as many laps of a 10-meter course as they can in 6 min. The Modified Shuttle Walk Test (MSWT) is an externally paced incremental walking test requiring participants to walk, jog, or run laps of a 10-meter course, keeping pace with an audio recording until they are unable to maintain the required pace [24], and was used for the remainder of the participants. Quality of life was assessed using the Short-Form 36-Item Health Survey (SF-36) [25]. Participants' health conditions were provided by the referring practitioner upon referral to the S2S program and checked with the participant during their initial assessment before commencing the S2S program. For ease of reporting, health conditions were categorized as follows: cardiovascular, metabolic, musculoskeletal, pulmonary, cancer, and mental health conditions. The wear time of the AT was calculated as the total number of days in which the tracker was worn divided by the total number of available days (365 days). Nonwear days were defined as days in which zero steps were recorded.

Power Calculation

During the 12-week S2S intervention, participants reported a mean increase of 450 MET minutes of physical activity per week, with an SD of 1.5 times the change. The sample size was calculated on a predicted maintenance of 100% of additional physical activity in the AT and TC groups and a decrease of 50% of additional physical activity in the UC group over the 12-month intervention. STATA 12 (Stata Corp) was used to calculate the sample size on the basis of a mean difference (MD) of 225 MET minutes per week with an SD of the change of 350 MET minutes per week, a power of 80%, and an α level of .05. This indicated a required sample size of 38 participants per group. To allow for withdrawals, 50 participants per group were recruited.

Statistical Analysis

All data were analyzed using STATA version 13.1 (StataCorp LLC) and graphically represented using GraphPad Prism (version 7.00 for Windows, GraphPad Software). Comparisons between the 3 interventions (as change from baseline) were made using mixed effects, repeated measures linear regression and replicated with ordered logistic regression adjusted for repeated measures because the assumptions of linear regression were not met for most variables. *P* values for comparison between groups were adjusted with Holm test for multiple comparisons. Statistical analysis was first conducted by completing the missing values through the last observation carried forward (LOCF) technique [15]. In addition, intention-to-treat (leaving missing values as blank) and per-protocol (only for people who completed all assessment time points) analyses were conducted. Intention-to-treat and LOCF included all 117 participants. The per-protocol analysis included 75 participants (UC=26, TC=25, and AT=24). The results are presented from mixed effects models in intention-to-treat analyses because it adjusts the maximum likelihood estimates based on the missing data and provides a more powerful analysis without ad hoc imputations with the

LOCF analysis. This technique also retains the sample size compared with the per-protocol analysis. For ease of understanding, results comparing the 3 interventions are shown as MD and 95% CI from mixed effects, repeated measures linear regression but with *P* values obtained from logistic regression analyses adjusted for multiple comparisons.

Results

Between September 2014 and June 2016, 152 people consented to participate when starting the S2S program, with 117 randomized to one of the study intervention groups on completion of the S2S program. The average age of the participants was 72.4 years (SD 6.5; range 60.3-88.7 years). Figure 2 shows the progression of the participants through the trial. Baseline characteristics are presented in Table 1.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram. ITT: Intention to Treat; S2S: Strength2Strength.

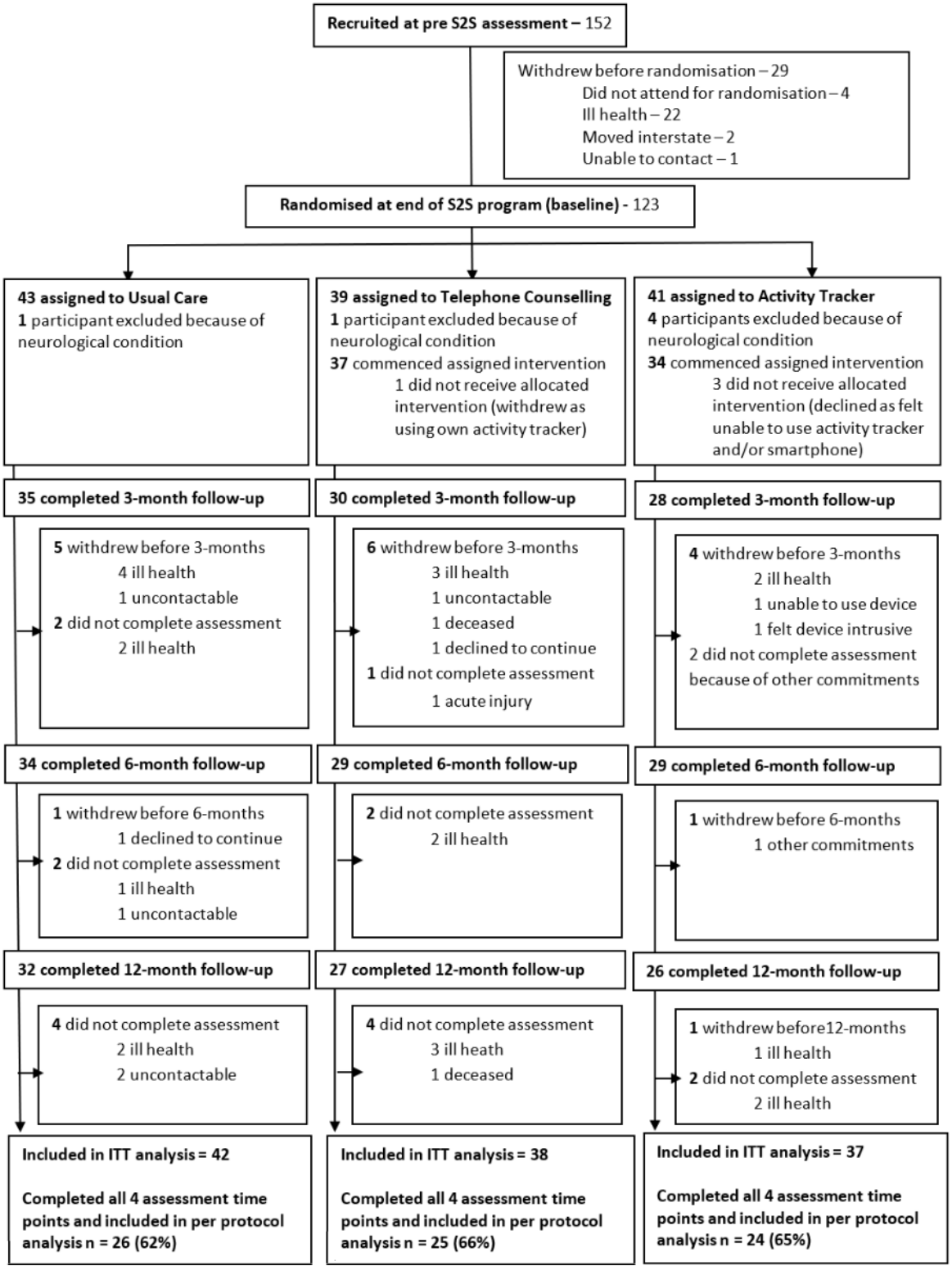


Table 1. Baseline (end of the Strength2Strength program) characteristics of randomized participants.

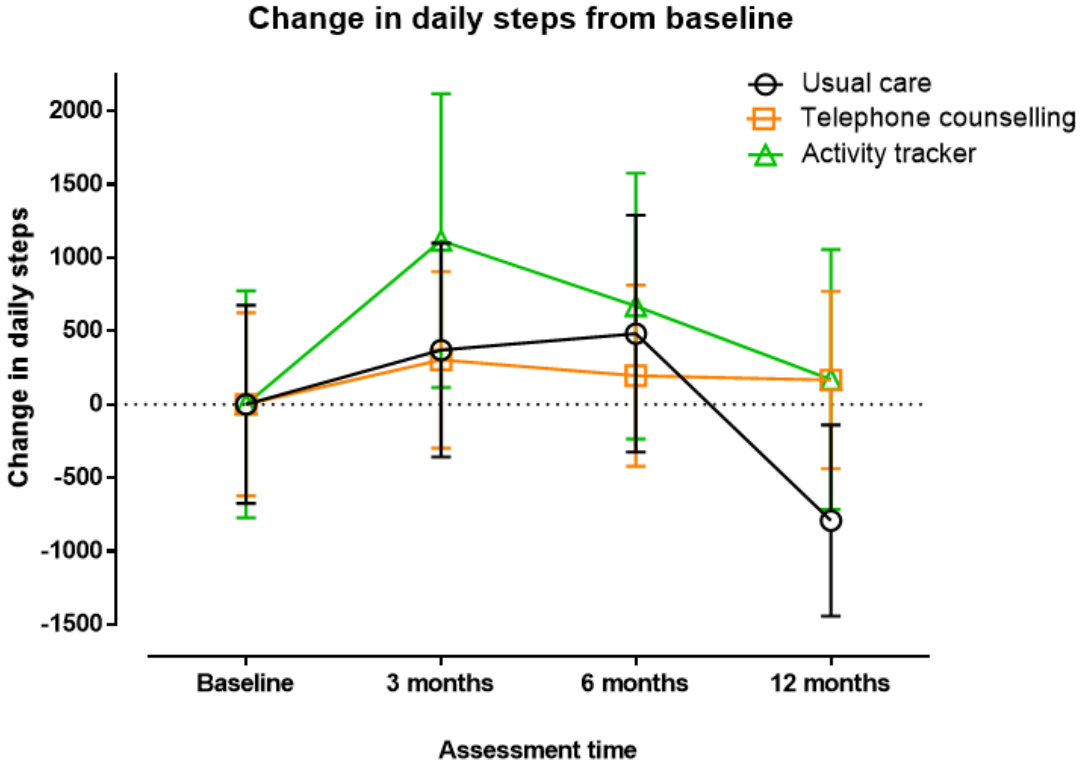
Outcome measure	Usual care group (n=42)	Telephone counseling group (n=38)	Activity tracker group (n=37)	Total (N=117)
Age (years), mean (SD)	71.9 (6.0)	72.8 (7)	72.3 (7)	72.4 (6.5)
Gender, n (%)				
Female	29 (69)	22 (58)	24 (65)	75 (64.1)
Male	13 (31)	16 (42)	13 (35)	42 (35.9)
Physical activity, mean (SD)				
Steps	6590 (2908)	4996 (2533)	6764 (3244)	6136 (2985)
Nonstepping time (minutes per 24 hours)	1073 (109)	1099 (138)	1093 (107)	1088 (118)
Self-reported activity (minutes)	200 (279)	176 (187)	174 (252)	183 (239)
Self-reported activity (metabolic equivalents per minute)	730 (994)	608 (680)	618 (892)	652 (855)
Body composition, mean (SD)				
Weight (kg)	84.5 (23.1)	84.6 (22.6)	83.6 (19)	84.2 (21.5)
BMI (kg/m^2)	31.5 (7.8)	31.2 (8.2)	30.2 (6.1)	31.0 (7.4)
Body fat (%)	37.9 (9.5)	35.3 (10.2)	36.9 (9.4)	36.7 (9.6)
Muscle mass (kg)	47.5 (10.6)	49.9 (11.1)	48.6 (9.1)	48.6 (10.3)
Blood pressure, mean (SD)				
Systolic blood pressure (mm Hg)	131 (16)	128 (20)	129 (13)	130 (17)
Diastolic blood pressure (mm Hg)	76 (9)	77 (10)	76 (8)	76 (9)
Physical function, mean (SD)				
Ten-time sit-to-stand (seconds)	21.7 (7.1)	21.5 (6.8)	23.7 (8.4)	22.2 (7.4)
Timed up and go (seconds)	7.2 (2.9)	7.5 (2.9)	7.8 (3.1)	7.5 (2.9)
Quality of life, mean (SD)				
SF-36[a] physical health summary score	52 (22)	55 (20)	55 (20)	54 (21)
SF-36 mental health summary score	64 (20)	68 (16)	68 (18)	67 (18)
Health conditions, n (%)				
Cardiovascular disease	25 (59)	21 (55)	21 (57)	67 (57)
Metabolic disease	26 (62)	14 (37)	16 (43)	56 (48)
Musculoskeletal conditions	25 (59)	23 (60)	30 (81)	79 (67)
Pulmonary conditions	7 (17)	9 (24)	6 (16)	22 (19)
Cancer	8 (19)	9 (24)	4 (11)	21 (18)
Mental health conditions	4 (9)	9 (24)	5 (13)	18 (15)
≥2 chronic conditions	29 (69)	28 (74)	27 (73)	84 (72)
Gait aid	4 (9)	4 (10)	6 (16)	14 (12)

[a]SF-36: Short-Form 36-Item Health Survey.

Participants in the AT group wore the AT for an average of 84% of the available days (306 out of 365 days). The number of nonwear days ranged from 1 to 164 days. The reasons for not wearing the band included forgetting to put the band on (all participants on at least one occasion), technical issues (n=20), illness (n=8), and holidays (n=5). Participants in the TC group received an average of 12 (SD 2) phone calls. Phone calls lasted an average of 11.1 min. The primary reason for missed phone calls was due to participants being away on holidays (n=5).

Over the 12-month intervention period, the UC group showed a significant reduction in the daily step count (MD 981 steps, 95% CI −1668 to −294 steps; *P*=.005). In contrast, step counts for the TC and AT groups did not change (MD −79 steps, 95% CI −823 to 663 steps; *P*=.81; and MD −588 steps, 95% CI −1359 to 182 steps; *P*=.09, respectively). There were no significant differences in changes in daily step counts between any of the groups over the 12-month intervention (*P*≥.14; Figure 3).

Figure 3. Change in daily steps with standard error of the mean between randomization and 12-month follow-up.



Objectively measured nonstepping time did not change over the 12-month period, and no differences were observed among the 3 groups between baseline and 12 months. Self-reported physical activity levels also did not change over the 12-month period, and there were no differences among the 3 groups at 12 months (Table 2).

No changes in body weight were observed over the 12-month period; however, there was a mean increase in BF% observed in both the TC and AT groups (MD 1.51%, 95% CI 0.43%-2.58%; *P*=.006; and MD 1.89%, 95% CI 0.82%-2.97%; *P*=.001, respectively). For LM, no change was observed between baseline and 12 months in the AT and TC groups, whereas the UC group showed a significant reduction in LM over the 12-month period (MD −1.13 kg, 95% CI −2.26 to −0.01 kg; *P*=.05).

DBP was significantly reduced over the 12-month period in the UC group (MD −4.10 mm Hg, 95% CI −7.02 to −1.18 mm Hg; *P*=.02), whereas no changes were observed in the TC and AT groups. SBP remained unchanged during the 12-month intervention period, and no differences were observed between the 3 groups. Therefore, there was a significant reduction in DBP in the UC group when compared with the AT group (MD 5.62 mm Hg, 95% CI 1.30 to 9.94 mm Hg; *P*=.01). No other between-group differences for DBP were observed between baseline and 12 months.

The UC group was the only group to demonstrate an improvement in TTSTS performance during the 12-month intervention period (MD −1.10, 95% CI −2.80 to 0.70; *P*=.03). The UC group performed the TTSTS significantly faster than the TC group between baseline and 12 months (MD 2.36 seconds, 95% CI −0.14 to 4.87 seconds; *P*=.03). The time taken to perform the TUAG did not change over the 12-month period for any of the intervention groups nor did the groups differ over the 12-month follow-up.

Of the 117 participants, a total of 58 participants (UC=19, TC=21, and AT=18) completed the MSWT, whereas 59 (UC=22, TC=17, and AT=18) completed the 6MWT. No differences in the distance walked during either the MSWT or 6MWT were observed between the 3 intervention groups or between baseline and 12 months within any of the intervention groups.

Self-reported physical function or mental health scores as measured by the SF-36 did not change during the 12-month intervention period for the 3 intervention groups, and there was no difference between groups at 12 months.

Table 2. Results of the included outcome measures at baseline and at 3, 6, and 12 months with within- and between-group *P* values.

Outcome measure	Within-group changes at 12 months					Between groups versus UC[a]	Between groups versus TC[b]
	Baseline	3 months	6 months	12 months	*P* value		
Physical activity participation							
Steps per day, mean (SD)							
UC	6590 (2908)	7050 (3083)	6968 (3551)	5836 (2422)	*<.01* [c]	N/A[d]	N/A
TC	4996 (2533)	5354 (2192)	5271 (2124)	5080 (2084)	.81	.14	N/A
AT[e]	6764 (3244)	7937 (4324)	7552 (3701)	7091 (3241)	.09	.45	.30
Nonstepping time (minutes per 24 hours), median (IQR)							
UC	1088 (797-1141)	1035 (942-1122)	1067 (961-1158)	1101 (1015-1204)	.07	N/A	N/A
TC	1119 (1021-1203)	1108 (965-1174)	1113 (1028-1180)	1106 (1071-1224)	.28	1.00	N/A
AT	1085 (1034-1151)	1064 (969-1139)	1046 (992-1149)	1100 (990-1158)	.29	.67	.99
Self-reported activity (minutes per week), median (IQR)							
UC	135 (50-265)	175 (106-330)	155 (75-255)	137 (30-230)	.27	N/A	N/A
TC	132.5 (50-400)	170 (120-260)	108.5 (40-202)	120 (60-290)	.28	.95	N/A
AT	270 (112.5-489)	210 (140-140)	180 (95-295)	207 (90-375)	.10	1.0	1.0
Self-reported activity (metabolic equivalents per minute per week), median (IQR)							
UC	460 (170-960)	625 (109-1215)	590 (300-930)	505 (105-871)	.34	N/A	N/A
TC	436.5(200-1330)	648 9480-965)	388 (145-765)	480 (210-1030)	.42	.98	N/A
AT	982.5 (415-1890)	840 (525-1550)	680 (632-1047)	760 (345-1425)	.11	1.0	1.0
Body composition							
Weight (kg), mean (SD)							
UC	84.5 (23.1)	82.8 (24.1)	79.3 (19.8)	77.7 (17.4)	.39	N/A	N/A
TC	84.6 (22.6)	86.2 (22.8)	86.1 (23.1)	86.6 (24.5)	.19	.24	N/A
AT	83.6 (19)	85.1 (16.6)	85.7 (17.6)	84.7 (17.2)	.61	.83	.19
BMI (kg/m^2), mean (SD)							
UC	31.5 (7.8)	31.2 (7.7)	30.1 (6.7)	29.9 (6.5)	.60	N/A	N/A
TC	31.2 (8.2)	31.7 (7.8)	31.8 (8.8)	31.5 (8.9)	.93	1.0	N/A
AT	30.2 (6.1)	30.9 (5.7)	30.8 (6.0)	30.5 (5.9)	.47	.87	1.0
Body fat (%), median (IQR)							
UC	40 (31.9-46.7)	39.9 (30-47.3)	39 (32.1-46.8)	41.8 (33.5-47.7)	.08	N/A	N/A
TC	37.3 (27.2-44.1)	40.3 (28.3-43)	40.4 (26.6-44.6)	40.7 (30-47.1)	*<.01* [c]	.39	N/A
AT	38 (30.4-44.4)	38.1 (31.8-44.3)	39.9 (29.4-44.4)	39.6 (32.5-44.2)	*<.01* [c]	.61	.95
Lean mass (kg), mean (SD)							
UC	47.5 (10.6)	45.9 (10.3)	44.9 (9.2)	43.1 (6.2)	*.05* [c]	N/A	N/A
TC	49.9 (11.1)	50.9 (11.1)	50.2 (10.3)	49.9 (10.8)	.36	.53	N/A
AT	48.6 (9.1)	49.6 (9.2)	50.7 (9.5)	49.3 (8.6)	.84	.49	.62
Blood pressure							
Systolic blood pressure (mm Hg), median (IQR)							
UC	130 (122-140)	137 (124-140)	129 (120-135)	130 (125-140)	.71	N/A	N/A
TC	130(120-140)	129 (120-137)	128 (123-144)	130 (122-135)	.95	.84	N/A

Outcome measure	Within-group changes at 12 months					Between groups versus UC[a]	Between groups versus TC[b]
	Baseline	3 months	6 months	12 months	*P* value		
AT	128 (120-138)	128 (120-139)	130 (122-135)	129 (125-140)	.80	.94	.90
Diastolic blood pressure (mm Hg), median (IQR)							
UC	74(70-80)	75 (70-80)	73 (65-79)	70 (70-75)	*.02* [c]	N/A	N/A
TC	76(70-83)	78 (70-84)	75 (70-80)	75 (70-80)	.35	.36	N/A
AT	78(70-80)	76 (73-80)	78 (70-80)	80 (74-80)	.11	.01[c]	.15
Physical function							
Ten-time sit-to-stand (seconds), median (IQR)							
UC	20.3 (16.8-24.1)	18.9 (16.4-22.6)	19.3 (15.9-24.5)	18.3 (14.9-24.0)	*.03* [c]	N/A	N/A
TC	20.9 (15.9-23.8)	20.3 (18.5-24.4)	20.7 (17.4-24.5)	20.5 (16.5-22.7)	.18	*.02* [c]	N/A
AT	21.3 (18.6-27.9)	19.7 (14.9-27.3)	19.7 (16.2-28.0)	21.7 (16.7-28.3)	.97	.14	.38
Timed up and go (seconds), mean (SD)							
UC	7.2 (2.9)	6.9 (2.6)	7.2 (2.5)	6.8 (2.6)	.71	N/A	N/A
TC	7.5 (2.9)	7.2 (2.2)	7.3 (2.4)	7.3 (2.3)	.17	.83	N/A
AT	7.8 (3.1)	7.9 (3.6)	7.5 (3.0)	7.9 (3.2)	.21	.47	.93
6 -min walk test (meters), median (IQR)							
UC	345 (299-413)	324 (272-400)	352 (300-384)	350 (335-422)	.51	N/A	N/A
TC	298 (220-450)	336 (244-430)	338 (294-430)	358 (312-467)	.24	.38	N/A
AT	375 (240-400)	342 (244-380)	295 (228-410)	291 (244-445)	.30	.72	.24
Modified shuttle walk test (meters), median (IQR)							
UC	460 (360-570)	520 (390-640)	530 (440-660)	450 (415-590)	.48	N/A	N/A
TC	440 (380-540)	480 (350-570)	555 (440-600)	530 (440-590)	.31	.84	N/A
AT	490 (390-560)	630 (520-640)	555 (480-630)	555 (450-620)	.26	1.0	.86
Quality of life							
SF-36[f] physical health summary score, median (IQR)							
UC	50 (36-72)	51.5 (31.5-77)	51 (31-72)	63 (39-71)	.48	N/A	N/A
TC	60 (38-72)	49 (68-70)	52 (41-68)	52 (43-69)	.76	.97	N/A
AT	53 (39-73)	54.5 (40-77)	61 (36-79)	60 (47-70)	.19	.63	.51
SF-36 mental health summary score, median (IQR)							
UC	68 (53-84)	65.5 (54-83)	67 (42-82)	77 (52-84)	.76	N/A	N/A
TC	72.5 (52-82)	73 (54-80)	69 (54-78)	77 (55-84)	.73	.96	N/A
AT	72 (53-83)	71 (53-83)	73 (62-82)	76 (68-82)	.35	1.0	1.0

[a]UC: usual care.

[b]TC: telephone counseling.

[c]Statistically significant.

[d]N/A: not applicable.

[e]AT: activity tracker.

[f]SF-36: Short-Form 36-Item Health Survey.

The results from the LOCF and per-protocol analyses showed similar results with a few key exceptions. For the LOCF analysis, the UC group showed a significant increase in nonstepping time (*P*=.05) between baseline and 12 months, whereas the AT group showed a significant decrease in self-reported minutes of physical activity per week between baseline and 12 months (*P*=.04). This was also observed in the per-protocol analysis (self-reported minutes per week, *P*=.02;

self-reported MET per minute per week, *P*=.01). The previously observed significant decrease in LM and time taken to perform the TTSTS between baseline and 12 months for the UC group was found to be no longer significant following the LOCF analysis (*P*=.09 and *P*=.07, respectively). Following the per-protocol analysis, a significant decrease in body weight was observed in the UC group (*P*=.05), and DBP in the AT group was found to differ significantly from the TC group (*P*=.05) in addition to the UC group, which was observed in the intention-to-treat analysis.

Discussion

Principal Findings

This study aimed to investigate how wearable ATs may assist in providing ongoing support to maintain physical activity levels and health outcomes in older adults compared with TC and UC over a 12-month intervention. We found that both ATs and TC were similarly effective at successfully maintaining daily step count over the 12-month intervention. The UC group maintained daily step count throughout the 9 months; however, a significant reduction in daily steps was observed at 12 months. Although previous research has suggested that the use of ATs either as part of a broader intervention or as a stand-alone intervention can significantly improve daily step counts [11], this is the first study to investigate the use of ATs to assist in the maintenance of physical activity over a 12-month period following a structured lifestyle intervention. Previous research has indicated that wearable ATs can help maintain physical activity participation (in cancer survivors) in the 3 months following a lifestyle intervention [26]. Using step count as a proxy for physical activity, our data demonstrate that, in older adults, this benefit can last at least a year and is as effective as the next best alternative, TC.

Interestingly, the results of the LOCF analysis showed a significant reduction in self-reported physical activity during the 12-month intervention in the AT group, reporting a mean decrease of 100 min of activity per week. The AT group was the only intervention group to consistently self-report performing at least 150 min of activity throughout the 12-month intervention and was performing between 70 and 87 min more of physical activity than both the UC and TC groups at 12 months. As this is approximately 50% of the recommended levels of physical activity, the AT group exceeded the minimally important clinical difference when compared with the TC and UC groups [27]. This suggests that even though a significant decrease in self-reported physical activity was observed in the AT group, it is unlikely that it had any significant implications for participants because of their overall higher levels of activity. It is also important to consider that results from the LOCF analysis are unlikely to reflect the true results of the study [28], particularly because of the high number of participants (15/113, 13.2%) who withdrew before completing the 3-month assessment. As noted in the *Methods* section, the LOCF analysis was completed as it was specified in the published protocol for this study [15] before the withdrawal of these participants.

Although overall body weight did not change for any of the included intervention groups, some changes in body composition were observed. An increase in BF% was observed in all intervention groups but was significant only in the TC and AT groups. Although elevated BF increases the risk of cardiovascular disease [29] and is associated with increased mortality [30], preservation of physical activity levels reduces mortality risk, even in individuals with elevated BF [31]. Despite the significant increase in BF%, this equated to a mean increase of only 0.3 kg in fat mass for both the AT and TC groups and therefore unlikely to place them at greater risk of chronic health conditions. In addition, the UC group showed a significant decrease in LM, with a reduction of almost 10% in LM between baseline and 12 months. The few previous studies that examined the effects of AT use on body composition and have included LM either reported decreases in LM across all groups potentially due to the incorporation of dietary intake restrictions or reported no change [32,33]. A key difference between this study and previous studies, which may potentially explain why LM was maintained in the AT and TC groups, is that the participants of this study were older and less healthy and therefore at an increased risk of developing sarcopenia. We speculate that the additional physical activity performed by the AT and TC groups helped to preserve their LM and subsequently assisted in delaying obligatory age-related sarcopenia, as has been shown previously in older adults who exercise [34]. In the long term, this preservation of LM could also subsequently reduce functional decline and mortality risk [35,36].

Although unexpected, a significant reduction in DBP and an improvement in physical function as measured by TTSTS was observed in the UC group, with significant differences also observed between the UC and AT groups for DBP and the UC and TC groups for TTSTS. The reason for the decrease in DBP observed in the UC group and the observed difference between the UC and AT groups are potentially related to the 6.8 kg weight loss in the UC group. Although this finding was not statistically significant, it did exceed the minimum clinically important difference for weight loss [37]. As the association between weight loss and reduction in blood pressure is well documented in individuals both with and without hypertension [38], this may have influenced the observed reduction in DBP in the UC group. In relation to the observed reduction in TTSTS time at 12 months in the UC group and differences between the UC and TC groups for the time taken to perform a TTSTS, the reasons are less clear. It has been suggested that weight loss alone can result in significant improvements in physical performance in frail, obese older adults [39]; however, because of the significant loss in LM observed, this may not fully explain the results observed. For a 5-time sit-to-stand test, the minimum clinically important difference is 1.7 seconds [40]. As the MD between the UC and TC groups was 2.3 seconds (95% CI −0.1 to 4.8 seconds) for the TTSTS, it is unclear if this would have a clinically meaningful effect. Furthermore, when analyzed using the LOCF analysis, a significant reduction in TTSTS time in the UC group was no longer observed. However, the difference between the UC and TC groups remained.

As ATs and TC were similarly effective in this study, the advantages and disadvantages of both methods should be considered to help determine if one is more suitable than the other and has more potential to encourage increases or

maintenance in physical activity. ATs have several potential benefits compared with TC. An AT allows individuals to self-monitor physical activity, which can improve physical activity participation [41]. Furthermore, sharing objectively measured data assists health care professionals in providing tailored feedback, which improves self-management [42]. In comparison, TC can be time intensive and resource intensive [9] and typically relies on subjective self-reported data, which can be unreliable in older adults [43]. It is important to note that this study reported that telephone calls lasted an average of 11 min per call, which is significantly less than previously reported studies [8,44]. The reason for this was most likely that the telephone calls provided served as a *check-in* with BCTs used as required to assist the maintenance of activity levels rather than as an intervention aimed at increasing physical activity participation. Currently, as there are limited options available that allow the sharing of AT data with health care professionals, it is acknowledged that providing feedback based on an individual's AT data can also be time intensive and resource intensive because of the data mining and interpretation required. Through improvements in the interoperability of proprietary and third-party mobile apps, the level of data mining would be reduced [45]. In addition, the development of a platform to specifically facilitate patient AT data transfer to health professionals would further reduce the time and resource burden of providing ongoing, tailored support.

Strengths and Limitations

A key strength of this study was the long-term intervention period, as most previous research included interventions of no longer than 6 months. The study design also allowed for direct comparison between feedback provided by an AT with TC, an established method of providing ongoing support. Evaluating the effect of newer technologies such as ATs is important in understanding how they can be incorporated into standard clinical practice. In addition, participants were recruited from a clinical exercise program, offering a good representation of the chronic conditions present in community-dwelling older adults.

A limitation of this study potentially affecting the ability to interpret results was that the a priori sample size was not met. A total of 150 participants were recruited at the start of the S2S program; however, a number of participants did not complete the S2S program and were not randomized. Owing to the cessation of the S2S program, additional participants could not be recruited to account for the withdrawals before the intervention. Further dropouts occurred during the 12-month intervention. The primary reason for withdrawal was ill health, which is not unexpected, given that more than 70% of the study population had 2 or more chronic conditions. Overall, 5 participants withdrew from the AT group because of feeling uncomfortable using the device. Although ATs are well accepted in older adult populations [46], this highlights that ATs may not be suitable for all older adults. Another consideration relates to the frequency in which feedback was provided in each of the intervention groups. Although the AT group received weekly feedback, the TC group received fortnightly feedback for the first 3 months and monthly feedback thereafter. It is possible that more regular feedback for the TC group may have improved participant outcomes. Finally, it is important to note that the placement of the ActivPAL (TM) on the front of the thigh means that some exercises, including the upper body and seated strength exercises, would not be captured.

Conclusions

Both TC- and AT-based interventions are effective at maintaining physical activity levels in older adults following a structured lifestyle intervention. As connected health technologies improve, ATs may provide an alternative to TC to assist older adults to remain active. The costs associated with delivering each intervention and the effects of each intervention on health utility should also be considered and investigated further.

Practical Implications

The following practical implications arose from the study findings:

- A consumer-based wearable AT is an effective alternative to traditional TC support to assist older adults in maintaining daily step counts.
- Clinicians should consider the individual needs of patients to determine whether TC or an AT is better suited to provide ongoing feedback and support.

Acknowledgments

This research was supported by the internal funding provided by the University of Tasmania. In-kind support was provided by Primary Health Tasmania during the recruitment and data collection phase of this research. AW was supported by a Medical Research Future Fund Translating Research into Practice Fellowship.

Authors' Contributions

KB completed data collection, data analysis, and interpretation and preparation of the manuscript. KA assisted with data analysis and interpretation and manuscript revisions. GW assisted with the conception of study design and manuscript revisions. JO assisted with data interpretation and manuscript revisions. AW assisted with the conception of study design, data analysis and interpretation, and manuscript revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of weekly text messages sent to AT participants and Jawbone UP app user interface showing progress towards daily step goal.
[PNG File , 203 KB - mhealth_v9i1e18686_app1.png]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 1614 KB - mhealth_v9i1e18686_app2.pdf]

References

1. Taylor D. Physical activity is medicine for older adults. Postgrad Med J 2014 Jan;90(1059):26-32 [FREE Full text] [doi: 10.1136/postgradmedj-2012-131366] [Medline: 24255119]
2. Giné-Garriga M, Roqué-Fíguls M, Coll-Planas L, Sitjà-Rabert M, Salvà A. Physical exercise interventions for improving performance-based measures of physical function in community-dwelling, frail older adults: a systematic review and meta-analysis. Arch Phys Med Rehabil 2014 Apr;95(4):753-69.e3. [doi: 10.1016/j.apmr.2013.11.007] [Medline: 24291597]
3. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, LIFE study investigators. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. J Am Med Assoc 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]
4. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
5. Lindahl B, Nilssön TK, Borch-Johnsen K, Røder ME, Söderberg S, Widman L, et al. A randomized lifestyle intervention with 5-year follow-up in subjects with impaired glucose tolerance: pronounced short-term impact but long-term adherence problems. Scand J Public Health 2009 Jun;37(4):434-442. [doi: 10.1177/1403494808101373] [Medline: 19181821]
6. van der Bij AK, Laurant MG, Wensing M. Effectiveness of physical activity interventions for older adults: a review. Am J Prev Med 2002 Feb;22(2):120-133. [doi: 10.1016/s0749-3797(01)00413-5] [Medline: 11818183]
7. Müller-Riemenschneider F, Reinhold T, Nocon M, Willich SN. Long-term effectiveness of interventions promoting physical activity: a systematic review. Prev Med 2008 Oct;47(4):354-368. [doi: 10.1016/j.ypmed.2008.07.006] [Medline: 18675845]
8. Goode AD, Reeves MM, Eakin EG. Telephone-delivered interventions for physical activity and dietary behavior change: an updated systematic review. Am J Prev Med 2012 Jan;42(1):81-88. [doi: 10.1016/j.amepre.2011.08.025] [Medline: 22176852]
9. Castro CM, King AC, Brassington GS. Telephone versus mail interventions for maintenance of physical activity in older adults. Health Psychol 2001 Nov;20(6):438-444. [Medline: 11714186]
10. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 9;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
11. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
12. Pellegrini CA, Verba SD, Otto AD, Helsel DL, Davis KK, Jakicic JM. The comparison of a technology-based system and an in-person behavioral weight loss intervention. Obesity (Silver Spring) 2012 Feb;20(2):356-363 [FREE Full text] [doi: 10.1038/oby.2011.13] [Medline: 21311506]
13. van Hoye K, Boen F, Lefevre J. The impact of different degrees of feedback on physical activity levels: a 4-week intervention study. Int J Environ Res Public Health 2015 Jun 9;12(6):6561-6581 [FREE Full text] [doi: 10.3390/ijerph120606561] [Medline: 26067990]
14. 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]
15. Brickwood K, Smith ST, Watson G, Williams AD. The effect of ongoing feedback on physical activity levels following an exercise intervention in older adults: a randomised controlled trial protocol. BMC Sports Sci Med Rehabil 2017;9:1 [FREE Full text] [doi: 10.1186/s13102-016-0066-5] [Medline: 28078090]
16. Miller WR, Rollnick S. Motivational Interviewing: Preparing People to Change Addictive Behavior. New York, USA: The Guilford Press; 1991.
17. Bandura A. Social cognitive theory. In: The International Encyclopedia of Communication. Thousand Oaks, CA: Sage Publications; 2008.
18. The Active Australia Survey: a Guide and Manual for Implementation, Analysis and Reporting. Australian Institute of Health and Welfare. 2003. URL: https://www.aihw.gov.au/reports/physical-activity/active-australia-survey [accessed 2020-12-05]

19. Heesch KC, Hill RL, van Uffelen JG, Brown WJ. Are Active Australia physical activity questions valid for older adults? J Sci Med Sport 2011 May;14(3):233-237. [doi: 10.1016/j.jsams.2010.11.004] [Medline: 21276752]
20. Ritchie JD, Miller CK, Smiciklas-Wright H. Tanita foot-to-foot bioelectrical impedance analysis system validated in older adults. J Am Diet Assoc 2005 Oct;105(10):1617-1619. [doi: 10.1016/j.jada.2005.07.011] [Medline: 16183365]
21. Ritchie C, Trost SG, Brown W, Armit C. Reliability and validity of physical fitness field tests for adults aged 55 to 70 years. J Sci Med Sport 2005 Mar;8(1):61-70. [doi: 10.1016/s1440-2440(05)80025-8] [Medline: 15887902]
22. Gibson AL, Wagner DR, Hetward VH. Advanced Fitness Assessment and Exercise Prescription Sixth Edition. New York, USA: Human Kinetics; 2010.
23. Enright PL. The six-minute walk test. Respir Care 2003 Aug;48(8):783-785 [FREE Full text] [Medline: 12890299]
24. Singh SJ, Morgan MD, Scott S, Walters D, Hardman AE. Development of a shuttle walking test of disability in patients with chronic airways obstruction. Thorax 1992 Dec;47(12):1019-1024 [FREE Full text] [doi: 10.1136/thx.47.12.1019] [Medline: 1494764]
25. Wang R, Wu C, Zhao Y, Yan X, Ma X, Wu M, et al. Health related quality of life measured by SF-36: a population-based study in Shanghai, China. BMC Public Health 2008 Aug 19;8:292 [FREE Full text] [doi: 10.1186/1471-2458-8-292] [Medline: 18710578]
26. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. Maintenance of physical activity and sedentary behavior change, and physical activity and sedentary behavior change after an abridged intervention: Secondary outcomes from the ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2856-2860 [FREE Full text] [doi: 10.1002/cncr.32142] [Medline: 31012968]
27. Fielding RA, Guralnik JM, King AC, Pahor M, McDermott MM, Tudor-Locke C, LIFE study group. Dose of physical activity, physical functioning and disability risk in mobility-limited older adults: Results from the LIFE study randomized trial. PLoS One 2017;12(8):e0182155 [FREE Full text] [doi: 10.1371/journal.pone.0182155] [Medline: 28820909]
28. Lachin JM. Fallacies of last observation carried forward analyses. Clin Trials 2016 Apr;13(2):161-168 [FREE Full text] [doi: 10.1177/1740774515602688] [Medline: 26400875]
29. Larsson SC, Bäck M, Rees JM, Mason AM, Burgess S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: a Mendelian randomization study. Eur Heart J 2020 Jan 7;41(2):221-226 [FREE Full text] [doi: 10.1093/eurheartj/ehz388] [Medline: 31195408]
30. Padwal R, Leslie WD, Lix LM, Majumdar SR. Relationship among body fat percentage, body mass index, and all-cause mortality: a cohort study. Ann Intern Med 2016 Apr 19;164(8):532-541. [doi: 10.7326/M15-1181] [Medline: 26954388]
31. Myers J, McAuley P, Lavie CJ, Despres J, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: their independent and interwoven importance to health status. Prog Cardiovasc Dis 2015;57(4):306-314. [doi: 10.1016/j.pcad.2014.09.011] [Medline: 25269064]
32. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
33. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the idea randomized clinical trial. J Am Med Assoc 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
34. Garatachea N, Pareja-Galeano H, Sanchis-Gomar F, Santos-Lozano A, Fiuza-Luces C, Morán M, et al. Exercise attenuates the major hallmarks of aging. Rejuvenation Res 2015 Feb;18(1):57-89 [FREE Full text] [doi: 10.1089/rej.2014.1623] [Medline: 25431878]
35. Newman AB, Yanez D, Harris T, Duxbury A, Enright PL, Fried LP, Cardiovascular Study Research Group. Weight change in old age and its association with mortality. J Am Geriatr Soc 2001 Oct;49(10):1309-1318. [doi: 10.1046/j.1532-5415.2001.49258.x] [Medline: 11890489]
36. Chang S, Lin P. Systematic literature review and meta-analysis of the association of sarcopenia with mortality. Worldviews Evid Based Nurs 2016 Apr;13(2):153-162. [doi: 10.1111/wvn.12147] [Medline: 26844538]
37. Swift DL, Johannsen NM, Lavie CJ, Earnest CP, Church TS. The role of exercise and physical activity in weight loss and maintenance. Prog Cardiovasc Dis 2014;56(4):441-447 [FREE Full text] [doi: 10.1016/j.pcad.2013.09.012] [Medline: 24438736]
38. Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 2003 Nov;42(5):878-884. [doi: 10.1161/01.HYP.0000094221.86888.AE] [Medline: 12975389]
39. Villareal DT, Chode S, Parimi N, Sinacore DR, Hilton T, Armamento-Villareal R, et al. Weight loss, exercise, or both and physical function in obese older adults. N Engl J Med 2011 Mar 31;364(13):1218-1229 [FREE Full text] [doi: 10.1056/NEJMoa1008234] [Medline: 21449785]
40. Jones SE, Kon SS, Canavan JL, Patel MS, Clark AL, Nolan CM, et al. The five-repetition sit-to-stand test as a functional outcome measure in COPD. Thorax 2013 Nov;68(11):1015-1020. [doi: 10.1136/thoraxjnl-2013-203576] [Medline: 23783372]

41. Huygens MW, Swinkels IC, de Jong JD, Heijmans MJ, Friele RD, van Schayck OC, et al. Self-monitoring of health data by patients with a chronic disease: does disease controllability matter? BMC Fam Pract 2017 Mar 20;18(1):40 [FREE Full text] [doi: 10.1186/s12875-017-0615-3] [Medline: 28320330]
42. Jager M, Lahr MM, Spoorenberg SL, Buskens E, Wynia K. Improving self-management of health through an eHealth application: an action-based study among older adults living in the community. Int J Integr Care 2018 Oct 23;18(s2):113. [doi: 10.5334/ijic.s2113]
43. Wild KV, Mattek N, Austin D, Kaye JA. 'Are you sure?': lapses in self-reported activities among healthy older adults reporting online. J Appl Gerontol 2016 Jun;35(6):627-641 [FREE Full text] [doi: 10.1177/0733464815570667] [Medline: 25669877]
44. Ewald B, Stacey F, Johnson N, Plotnikoff RC, Holliday E, Brown W, et al. Physical activity coaching by Australian exercise physiologists is cost effective for patients referred from general practice. Aust N Z J Public Health 2018 Feb;42(1):12-15. [doi: 10.1111/1753-6405.12733] [Medline: 29165855]
45. 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]
46. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]

Abbreviations

6MWT: 6-min walk test
AAS: Active Australia Survey
AEP: accredited exercise physiologist
AT: activity tracker intervention
BCT: behavior change technique
BF: body fat
DBP: diastolic blood pressure
LM: lean mass
LOCF: last observation carried forward
MD: mean difference
MET: metabolic equivalent
MSWT: Modified Shuttle Walk Test
RCT: randomized controlled trial
S2S: Strength2Strength
SBP: systolic blood pressure
SF-36: Short-Form 36-Item Health Survey
TC: telephone counseling
TTSTS: 10-time sit-to-stand
TUAG: timed up and go
UC: usual care

Edited by G Eysenbach, L Buis; submitted 12.03.20; peer-reviewed by R Li, C Blair; comments to author 12.06.20; revised version received 07.08.20; accepted 15.10.20; published 05.01.21.

Please cite as:
Brickwood KJ, Ahuja KDK, Watson G, O'Brien JA, Williams AD
Effects of Activity Tracker Use With Health Professional Support or Telephone Counseling on Maintenance of Physical Activity and Health Outcomes in Older Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(1):e18686
URL: https://mhealth.jmir.org/2021/1/e18686
doi:10.2196/18686
PMID:33399541

Original Paper

Wrist-Worn Activity Trackers in Laboratory and Free-Living Settings for Patients With Chronic Pain: Criterion Validity Study

Veronica Sjöberg[1], MSc; Jens Westergren[1], MSc; Andreas Monnier[1,2,3], PhD; Riccardo Lo Martire[3], MSc; Maria Hagströmer[3,4], Prof Dr; Björn Olov Äng[1,3,5], Prof Dr; Linda Vixner[1], PhD

[1]School of Education, Health and Social Studies, Dalarna University, Falun, Sweden

[2]Military Academy Karlberg, Swedish Armed Forces, Solna, Sweden

[3]Department of Neurobiology, Care Sciences and Society, Division of Physiotherapy, Karolinska Institutet, Huddinge, Sweden

[4]Academic Primary Health Care Centre, Region Stockholm, Stockholm, Sweden

[5]Center for Clinical Research Dalarna, Uppsala University, Region Dalarna, Falun, Sweden

Corresponding Author:
Veronica Sjöberg, MSc
School of Education, Health and Social Studies
Dalarna University
Högskolegatan 2
Falun, SE-791 88
Sweden
Phone: 46 23 77 87 57
Email: vsj@du.se

Abstract

Background: Physical activity is evidently a crucial part of the rehabilitation process for patients with chronic pain. Modern wrist-worn activity tracking devices seemingly have a great potential to provide objective feedback and assist in the adoption of healthy physical activity behavior by supplying data of energy expenditure expressed as metabolic equivalent of task units (MET). However, no studies of any wrist-worn activity tracking devices' have examined criterion validity in estimating energy expenditure, heart rate, or step count in patients with chronic pain.

Objective: The aim was to determine the criterion validity of wrist-worn activity tracking devices for estimations of energy expenditure, heart rate, and step count in a controlled laboratory setting and free-living settings for patients with chronic pain.

Methods: In this combined laboratory and field validation study, energy expenditure, heart rate, and step count were simultaneously estimated by a wrist-worn activity tracker (Fitbit Versa), indirect calorimetry (Jaeger Oxycon Pro), and a research-grade hip-worn accelerometer (ActiGraph GT3X) during treadmill walking at 3 speeds (3.0 km/h, 4.5 km/h, and 6.0 km/h) in the laboratory setting. Energy expenditure and step count were also estimated by the wrist-worn activity tracker in free-living settings for 72 hours. The criterion validity of each measure was determined using intraclass and Spearman correlation, Bland-Altman plots, and mean absolute percentage error. An analysis of variance was used to determine whether there were any significant systematic differences between estimations.

Results: A total of 42 patients (age: 25-66 years; male: 10/42, 24%; female: 32/42, 76%), living with chronic pain (duration, in years: mean 9, SD 6.72) were included. At baseline, their mean pain intensity was 3.5 (SD 1.1) out of 6 (Multidimensional Pain Inventory, Swedish version). Results showed that the wrist-worn activity tracking device (Fitbit Versa) systematically overestimated energy expenditure when compared to the criterion standard (Jaeger Oxycon Pro) and the relative criterion standard (ActiGraph GT3X). Poor agreement and poor correlation were shown between Fitbit Versa and both Jaeger Oxycon Pro and ActiGraph GT3X for estimated energy expenditure at all treadmill speeds. Estimations of heart rate demonstrated poor to fair agreement during laboratory-based treadmill walks. For step count, the wrist-worn devices showed fair agreement and fair correlation at most treadmill speeds. In free-living settings; however, the agreement for step count between the wrist-worn device and waist-worn accelerometer was good, and the correlation was excellent.

Conclusions: The wrist-worn device systematically overestimated energy expenditure and showed poor agreement and correlation compared to the criterion standard (Jaeger Oxycon Pro) and the relative criterion standard (ActiGraph GT3X), which needs to be considered when used clinically. Step count measured with a wrist-worn device, however, seemed to be a valid estimation, suggesting that future guidelines could include such variables in this group with chronic pain.

(JMIR Mhealth Uhealth 2021;9(1):e24806) doi:10.2196/24806

KEYWORDS

chronic pain; energy expenditure; heart rate; physical activity; step count; validity; wearable devices; wearable; pain; rehabilitation

Introduction

Chronic pain is defined as "pain that persists past normal healing time and hence lacks the acute warning function of physiological nociception" [1,2] and is a leading major public health problem internationally due to its effects on physical, social, and emotional functions [3]. Physical activity is a central part of chronic pain rehabilitation due to the evident health benefits, which include improved cardiovascular health, prolonged lifespan [4,5], positive effects on pain intensity, health-related quality of life, and both physical and psychological functions [6]. The American Heart Association has provided guidelines regarding sufficient weekly amounts of physical activity to reap health benefits for a healthy population, as well as for populations with chronic conditions [4,7]. For patients with chronic pain, recommendations are to spend ≥150 minutes/week engaged in moderate-to-vigorous physical activity (MVPA). Moderate physical activity is defined as equal to or more than 3 and less than 6 metabolic equivalent task units (MET) [8]. One MET is defined as a resting metabolic rate obtained when quietly seated [8]. Despite clear guidelines, it seems that inadequate physical activity levels are common among patients with chronic pain, which can lead to an increased risk of physical and mental illness [5]. In rehabilitation settings, objective estimations of physical activity are rarely used. Instead, subjective measures are common practice due to their high degree of acceptance, cost effectiveness, and relatively low administrative burden [9]. However, despite its perceived benefits, subjective estimations of physical activity domains have estimation biases, such as recall bias and reactivity bias [9]. Several studies [10-14] have indicated the potential of wrist-worn activity tracking devices as tools that can facilitate behavior change and increase the degree to which patients follow individually modulated physical activity levels designed to improve health. Wearable devices for physical activity tracking have received increased interest from both the research community and consumers aiming to quantify domains of physical activity (eg, frequency and duration) in order to optimize health behaviors [10,15,16]; however, before the clinical use of these devices can be introduced, the validity of each device needs to be established [17]. In the past decade, there has been an increasing number of studies [18-22] assessing the validity of wrist-worn tracking devices that measure energy expenditure by comparison to a criterion standard such as indirect calorimetry or accelerometry. The majority of these validation studies were conducted among healthy adult participants [17,23], with studies reporting somewhat conflicting findings—both overestimation [20,23] and underestimation [19,24,25] with Fitbit devices were reported. In a recent systematic review [23] investigating the accuracy of Fitbit devices, it was reported that 49% (43 of 88 comparisons) overestimated energy expenditure, particularly during physical activity. In an earlier systematic review of the field, Evenson et al [17] reported a high validity of different brands of wearable activity tracking devices regarding step count when compared to various criterion standards made in laboratory settings [26,27]. Regarding the validity of heart rate estimations from by wrist-worn activity tracking devices, one study [28] have shown that the agreement between true rate and the estimated rate made by a wrist-worn device is higher during rest than during MVPA in healthy subjects. To our knowledge, there has been no prior research examining wrist-worn activity tracking device criterion validity in estimating energy expenditure (using MET), step count, or heart rate among patients with chronic pain. This lack of research constitutes a substantial knowledge gap given how important it is for patients with chronic pain to achieve adequate amounts of weekly physical activity. Therefore, the aim of this study was to evaluate the criterion validity of each of these measures estimated by a wrist-worn activity tracking device for patients with chronic musculoskeletal pain in both laboratory and free-living settings.

Methods

Study Design

We conducted a laboratory and field validation study. Data were collected between March 2019 to June 2020 (Health and Sports laboratory, Dalarna University). The sample size calculation was based on intraclass correlation (ICC), the primary statistic in the study. In order to achieve 80% power to detect an ICC of 0.80 (excellent agreement) with a 95% distribution (lower limit 0.6), calculation based on published recommendations [29] showed a requirement of 26 to 49 participants. This study was approved by Swedish Ethical Review authority (registration number 2018-307).

Recruitment and Study Sample

The inclusion criteria were adult age (between 18 and 67 years), with chronic (>3 months) musculoskeletal (neck or low back) pain or widespread pain, currently undergoing assessment or treatment (for chronic pain) in a primary or specialized health care clinic, and having the ability to understand information in Swedish. The exclusion criteria were having given birth within the previous 3 months, pregnant in the second or third trimester, requiring a walking aid indoors, currently undergoing heart assessment or investigation, with pain caused by malignancy or systematic disease, or having a known allergy to plaster or adhesive tape. Participants were recruited from 8 primary and specialized health care clinics in Region Dalarna. Patients who matched the study criteria (age, duration of pain, language) were asked by clinicians for consent to be contacted by a study representative, who conducted additional screening for eligibility. At the test site, for safety reasons, all participants declared whether they had been diagnosed with or experienced a heart condition, chest pain, dizziness, high or low blood pressure, any respiratory disorder, or diabetes before any tests were performed. Participants' height and weight were manually measured using a stadiometer (Holtain Limited) and a weighing

scale (Sartorius AG). A self-rated questionnaire captured date of birth; biological sex; education level; work status; years lived with pain; and pharmaceutical, caffeine, and nicotine consumption in the previous 24 hours. Participants also completed the Swedish National Board of Health and Welfare's questionnaire on physical activity level (minutes per week spent in exercise and in physical activity) [30,31]. In addition, participants completed the Multidimensional Pain Inventory (in Swedish) to describe psychosocial and behavioral consequences of pain [32].

Equipment

A wrist-worn activity tracker (Fitbit Versa, Fitbit Inc), chosen for its high degree of user-friendliness, because it can be used with web or smartphone apps, and it is suitable for water activities. The Fitbit Versa estimates movement (eg, active minutes) using a triaxial accelerometer and MET/minutes based on a combination of basal metabolic rate (adjusted for sex, age, height, and weight), accelerometry-based activity counts, and heart rate measured through optical sensors [33,34].

The criterion standard (gold standard) for energy expenditure in our laboratory setting was indirect calorimetry from pulmonary gas exchange. Oxygen uptake (VO_2) and carbon dioxide production (VCO_2) was measured using a mixing-chamber system (Jaeger Oxycon Pro) that measures respiratory gas exchange through a mouthpiece and tube [35]. Jaeger Oxycon Pro provides an assessment of resting energy expenditure and activity-related energy expenditure based on type and amount of substrate oxidized and the amount of energy produced by biological oxidation—MET values are based on the equation: 1 MET = 3.5 mL/min/kg VO_2 [8]. Before the start of the testing protocol, ambient conditions were recorded, and automatic volume and gas calibration was performed using a high-precision gas mixture (Air Liquide AB). The Jaeger Oxycon Pro has been validated by comparison to the Douglas bag-method and has been found a reliable criterion standard for indirect calorimetry [36]. Real-time VO_2 and heart rate data were recorded throughout the entire laboratory protocol.

The relative criterion standard was a research-grade hip-worn accelerometer: ActiGraph GT3X-BL (ActiGraph LLC) and appurtenant software Actilife (version 6.13.3; ActiGraph LLC). The ActiGraph GT3X is a research-based triaxial accelerometer commonly used as a criterion standard both in free-living and in laboratory settings, within various populations as it is a valid and reliable tool to quantify physical activity [11,37,38].

Procedures

According to current guidelines [39,40], in investigations aiming to evaluate the criterion validity of a wrist-worn activity tracker, data collection should be conducted in laboratory and free-living settings. In the laboratory setting, energy expenditure data were concurrently collected from Jaeger Oxycon Pro and Fitbit Versa during rest (sitting quietly seated for 10 minutes) and during treadmill walking (18 minutes). Heart rate data were also collected with a chest band (Polar HR10). Step count was estimated by ActiGraph GT3X and Fitbit Versa. The last 2 minutes of each activity (rest, treadmill speed) was included in data analysis providing data during a steady state environment [41]. During rest, participants were seated (wearing the facemask with tube) in an inclined chair with supported arms, under a blanket to avoid feeling cold. The room temperature was set at 20 °C, and the laboratory was kept quiet during the resting period. The treadmill walk protocol consisted of 6 minutes at each speed of 3.0 km/h, 4.5 km/h, and 6.0 km/h. At the end of each 6 minutes, participants rated perceived exertion according to (Borg Rating of Perceived Exertion, rating from 6-20) [42], and after the third final speed, pain intensity was also assessed using a visual analog scale (0 mm to 100 mm) [43]. In the free-living setting, step count was concurrently estimated by Fitbit Versa and ActiGraph GT3X for the subsequent 72 hours after the laboratory testing [39]. Participants were instructed to wear the devices simultaneously for at least 10 hours each day, to remove the devices for sleeping, showering, and bathing, and to record their wear-time in a logbook. Data collection started once participants left the laboratory. A schematic overview for the study procedure is shown in Table 1.

Table 1. A schematic overview of the study procedure, measurements, and outcomes in both settings.

Activity	Duration	Instruments and devices	Outcomes
Laboratory setting			
Baseline measurements	N/A[a]	Sartorius weighting scale	Weight
		Holtain Stadiometer	Height
Questionnaires	N/A	Multidimensional Pain Inventory	Personal characteristics
			Pain characteristics
		National Board of Health and Welfare questions for physical activity level	Physical activity level
Seated rest measurements	10 minutes	Jaeger Oxycon Pro	Energy expenditure, heart rate
		ActiGraph GT3X	Energy expenditure
		Fitbit Versa	Energy expenditure, heart rate
Treadmill walk measurements	18 minutes (6 minutes at 3.0, 4.5, 6.0 km/h each)	Jaeger Oxycon Pro	Energy expenditure, heart rate
		ActiGraph GT3X	Energy expenditure, step count
		Fitbit Versa	Energy expenditure, heart rate, step count
		Borg´s RPE scale (6-20)	Perceived exertion
		Visual Analogue Scale (0-100)	Pain intensity post–treadmill walk
Free-living setting			
Free-living activities	72 hours	ActiGraph GT3X	Energy expenditure, step count
		Fitbit Versa	Energy expenditure, step count
		Logbook	Wear-time

[a]N/A: not applicable.

Experimental Measurement

The Fitbit Versa was initialized, and participants' age, height, length, and biological sex were registered. The device was synchronized to its app (Fitbit Dashboard) and fitted on participants' nondominant wrist according to the manufacturer's recommendations. To retrieve data (energy expenditure, step count, heart rate) we deployed a web-based application programming interface [44] with assistance from an experienced computer programmer. Through such script, Fitbit allows users to download defined data by minute resolution. After the devices were returned, they were resynchronized before data was downloaded.

Criterion Standard

Participants' biological sex, height, and weight were entered into the software. Data (energy expenditure, heart rate) retrieved from Jaeger Oxycon Pro and Polar HR10 were manually aggregated to minute resolution (from 15 s to 60 s) to correspond with Fitbit Versa and ActiGraph GT3X data output.

Relative Criterion Measurement

ActiGraph GT3X was initialized at the 30 Hz sample rate and participants' date of birth, height, length, and biological sex were entered. The device was fitted on participants' waists, to the right of the spine, using an elasticated belt. Data (counts per axis, step count) were downloaded in epochs of 60 seconds, which is commonly used in corresponding research [45]. After download, we applied a cut-off (combining the Work-Energy Theorem and the Freedson equation) in Actilife software (version 6.13.4; ActiGraph LLC) that combines to calculate energy expenditure [46]. Actilife calculates MET values based on brand-specific activity counts and chosen cut points. We applied the Freedson cut-point to score MET per minute [47].

Data Management and Statistics

Frequency analysis of data was performed to identify potential errors. Manual checking of random samples (20% of the data)was carried out and deemed satisfactory with <3% error rate. Descriptive statistics were used to describe participant characteristics. The Shapiro-Wilk test was used to determine whether data were normally distributed. The criterion validity was determined through assessment of agreement as well as assessment of correlation between estimations and measurements of primary outcomes energy expenditure, heart rate, and step count in laboratory and free-living [39,40]. Agreement was assessed with ICC coefficient analysis (2-way random, average measures, 95% CI, absolute agreement) [48,49]. An ICC below 0.4 was considered poor, an ICC between 0.4 and 0.59 fair, an ICC between 0.6 and 0.74 good, and an ICC above 0.75 was considered as excellent [50]. Analysis of variance (ANOVA) was used to determine any significant systematic differences between estimations. To visualize the absolute, unscaled agreement [48,51], Bland-Altman plots with 95% CI (ie, limits of agreement, LOA) were calculated. Values beyond ±3 SD were identified as outliers and were excluded from analysis after sensitivity analysis. To determine correlation between estimations of energy

expenditure, step count, and heart rate, Spearman (ρ) bivariate correlation analysis was used, and ρ<0.2 was considered poor, 0.2≤ρ<0.6 was considered fair, 0.6≤ρ<0.8 was considered moderate, 0.8≤ρ<0.9 was considered very strong, 0.9≤ρ<1 was considered perfect [52,53]. In addition, mean absolute percentage error (MAPE) were calculated as a measure of accuracy for both measured energy expenditure, steps, and heart rate as the mean difference between estimations of the wrist-worn activity tracker and estimations of the criterion measurement (Jaeger Oxycon Pro or ActiGraph GT3X) multiplied by 100, divided by the mean of the criterion measurement (Jaeger Oxycon Pro or ActiGraph GT3X) [27]. An MAPE value <1% was acceptable in the laboratory context [28,54] and a MAPE <10% of the criterion value was considered an acceptable rate of error in the free-living setting [9]. Missing data analysis was performed as recommended by Fox-Wasylyshyn [55] to evaluate any significant association between missing data and participant characteristics at baseline. Our predetermined significance level for *P* values was .05

Results

Participants

A total of 42 patients (female: 32/42, 76%; male: 10/42, 24%) participated in the study, but only 41 participants completed the protocol due to the malfunction of 1 device. The participants' mean age was 43.8 years (SD 11.8). Participants' mean BMI was 29.4 (SD 5.8), 66% of participants (27/41) were working/studying at the time of the study, and 49% (20/41) stated that they were physically active 150 minutes/week or more (Table 2). Most participants (36/41, 88%) completed all 3 treadmill speeds, while the remaining participants (5/41, 12%) discontinued the treadmill test at the highest speed due to high physical exertion or increased pain. Missing analysis revealed 1 significant result—all participants who discontinued the treadmill walk at the highest speed reported being physically active <150 minutes/week at baseline, while 44% (16/36) among those who completed all 3 treadmill speeds rated <150 minutes/week (*P*=.05). The mean ratings of perceived exertion at the end of each treadmill walk were 9 (SD 2), 12 (SD 2), and 14 (SD 2) for 3.0 km/h, 4.5 km/h, 6.0 km/h. Pain intensity ranged from 1 mm to 96 mm, mean 43 mm (SD 29 mm) on the visual analog scale after completion of the treadmill walk. Within the 24 hours prior to testing, 15 of the 41 participants (37%) used analgesics, and 2 (<5%) used beta blockers. Because 3 participants did not return their logbooks, data from 38 participants were included in the free-living analyses. The mean wear-time of the devices during the free-living period was 31 hours and 23 minutes (SD 6 hours and 21 minutes).

Table 2. Personal and pain characteristics of participants.

Characteristic	Value (n=41)
Demographic characteristic	
Sex, n (%)	
Female	31 (76)
Male	10 (24)
Age (years), mean (SD)	43.8 (11.8)
BMI, mean (SD)	29.4 (5.8)
Education level, n (%)	
Elementary	1 (2)
Secondary	28 (68)
University	12 (29)
Other unspecified	1 (2)
Working/studying, n (%)	
Yes	27 (66)
No	14 (34)
Treatment, n (%)	
Primary health care	33 (80)
Specialized	8 (20)
Pain characteristics	
Multidimensional pain inventory, Swedish version (0-6), part 1, mean (SD)	
Pain intensity	3.6 (1.1)
Pain interference	3.7 (0.8)
Life control	3.4 (1.0)
Affective distress	2.9 (0.9)
Social support	3.6 (1.3)
Number of pain locations (0-36)	14.0 (9.5)
Years lived with pain, n (%)	
0-5[a]	20 (49)
6-10	5 (12)
<10	15 (37)
Pharmaceutical consumption last 24 hours, n (%)	
Analgesics	15 (37)
Beta blockers, n (%)	2 (<5)
Physical activity level[b]	
Exercise[c] (minutes/week), n (%)	
0-30	15 (37)
31-90	11 (27)
91-120	11 (27)
>120	4 (10)
Physical activity[d] (minutes/week), n (%)	
0-60	8 (20)
61-150	13 (32)

Characteristic	Value (n=41)
151-300	7 (17)
>300	13 (32)

[a]All participants had experienced pain >3 months.

[b]National Board of Health and Welfare's questions for physical activity level.

[c]Structured physical activity requiring physical effort and aims to improve health and fitness.

[d]Any bodily movement produced by skeletal muscles that requires energy expenditure.

Criterion Validity

The mean energy expenditure, heart rate, and step count of the criterion standard (Jaeger Oxycon Pro), the relative criterion measure (ActiGraph GT3X), and the experimental measure (Fitbit Versa) are presented in Table 3. The ICC (95% CI), mean difference with upper and lower LOA, Spearman correlation, and MAPE for all statistical calculations are presented in Table 4 and Table 5. The Bland-Altman plots for energy expenditure, step count, and heart rate are shown in Figures 1-5.

Table 3. Energy expenditure, heart rate, and step count during treadmill walking and in free-living setting.

Measure	Seated rest	Treadmill walk				Free-living setting
		3.0 km/h	4.5 km/h	6.0 km/h	Overall	
Energy expenditure						
MET/minute, mean (SD)						
Jaeger Oxycon Pro	0.73 (0.17)[a]	2.76 (0.36)	3.50 (0.37)	5.10 (0.42)[b]	3.80 (0.33)[b]	N/A[c]
ActiGraph GT3X	1.00 (0.00)	1.31 (0.40)	3.91 (1.40)	5.48 (1.07)[b]	3.56 (0.84)[b]	2.53 (0.52)[d]
Fitbit Versa	1.0 (0.02)	5.73 (0.56)[b]	6.41 (0.58)[a]	7.56 (1.17)[b]	6.56 (0.64)[e]	3.73 (0.86)[d]
***P* value**						
Jaeger Oxycon Pro–Fitbit Versa	<.001	<.001	<.001	<.001	<.001	N/A
ActiGraph GT3X–Fitbit Versa	N/A	<.001	<.001	<.001	<.001	<.001
Heart rate						
bpm, mean (SD)						
Jaeger Oxycon Pro	71.25 (9.82)[a]	99.15 (15.54)[a]	109.61 (14.89)[a]	132.14 (19.48)[b]	113.53 (16.14)[b]	N/A
Fitbit Versa	72.11(11.55)	101.83 (10.92)[a]	108.67 (8.18)	121.95 (9.63)[b]	110.85 (7.05)[b]	N/A
***P* value**						
Jaeger Oxycon Pro–Fitbit Versa	.81	.37	.77	.002	0.34	N/A
Step count						
Steps/minute						
ActiGraph GT3X	N/A	92.70 (8.78)[a]	110.83 (7.08)	124.93 (6.52)[e]	108.61 (6.11)[e]	18.64 (8.51)[d]
Fitbit Versa	N/A	91.89 (8.75)	106.72 (7.05)	114.20 (11.09)[e]	103.95 (5.62)[e]	11.34 (5.87)[f]
***P* value**						
ActiGraph GT3X–Fitbit Versa	N/A	.53	<.001	<.001	<.001	<.001

[a]n=40.

[b]n=39.

[c]N/A: not applicable.

[d]n=38.

[e]n=36.

[f]n=37.

Table 4. Comparison between experimental measurement (Fitbit Versa) and the criterion standard (Jaeger Oxycon Pro) in the laboratory setting.

Test and measure	Jaeger Oxycon Pro vs Fitbit Versa (n=41)			
	Energy expenditure–energy expenditure	Heart rate–heart rate	Energy expenditure–heart rate	Energy expenditure–step count
Seated rest[a]				
ICC[b] (95% CI)	0.003 (–0.16 to 0.20)	0.99 (0.98 to 0.99)	N/A[c]	N/A
Mean difference (LOA[d])	0.27 (–0.07 to 0.61)	0.09 (–4.35 to 4.52)	N/A	N/A
ρ (*P* value)	–0.03 (.86)	0.96 (<.001)	0.27 (.09)	
MAPE[e]	28.46	2.24	N/A	N/A
Treadmill walk				
3.0 km/h[f]				
ICC (95% CI)	0.01 (–0.04 to 0.04)	0.09 (–0.72 to –0.52)	N/A	N/A
Mean difference (LOA)	2.97 (1.61 to 4.34)	–2.68 (–39.01 to 33.65)	N/A	N/A
ρ (*P* value)	–0.14 (.39)	0.24 (.14)	0.10 (.53)	–0.04 (.82)
MAPE	51.52	11.54	N/A	N/A
4.5 km/h[g]				
ICC (95% CI)	–0.03 (–0.09 to –0.07)	0.20 (–0.55 to 0.58)	N/A	N/A
Mean difference (LOA)	2.91 (1.39 to 4.43)	0.75 (–30.73 to 32.22)	N/A	N/A
ρ (*P* value)	–0.31 (.05)	0.16 (.33)	–0.11 (.51)	–0.19 (.24)
MAPE	44.80	10.19	N/A	N/A
6.0 km/h[h]				
ICC (95% CI)	–0.05 (–0.19 to –0.16)	0.40 (–0.09 to 0.68)	N/A	N/A
Mean difference (LOA)	2.46 (–0.11 to 5.03)	10.19 (–25.56 to 45.95)	N/A	N/A
ρ (*P* value)	–0.11 (.51)	0.44 (.01)	0.17 (.34)	0.01 (.97)
MAPE	31.59	12.10	N/A	N/A
Overall speeds[i]				
ICC (95 % CI)	–0.03 (–0.08 to 0.08)	0.19 (–0.58 to 0.59)	N/A	N/A
Mean difference (LOA)	2.76 (1.21 to 4.31)	–2.68 (–35.30 to 29.95)	N/A	N/A
ρ (*P* value)	–0.22 (.20)	0.23 (.17)	–0.05 (.75)	0.07 (.71)
MAPE	35.39	10.56	N/A	N/A

[a]n=40.

[b]ICC: intraclass correlation.

[c]N/A: not applicable.

[d]LOA: limits of agreement.

[e]MAPE: mean absolute percent error.

[f]n=39 for energy expenditure–energy expenditure; n=40 for heart rate–heart rate and energy expenditure–heart rate.

[g]n=40 for energy expenditure–energy expenditure and heart rate–heart rate.

[h]n=36 for energy expenditure–energy expenditure; heart rate–heart rate, and energy expenditure–heart rate; n=39 for energy expenditure–step count.

[i]n=35 for energy expenditure–energy expenditure and energy expenditure–step count; n=36 for energy expenditure–heart rate and heart rate–heart rate.

Table 5. Comparison between the experimental measurement (Fitbit Versa) and the relative criterion measurement (ActiGraph GT3X) in both settings (laboratory and free-living).

Test and measure	ActiGraph GT3X vs Fitbit Versa (n=41)			
	Energy expenditure–energy expenditure	Step count–step count	Energy expenditure (ActiGraph GT3X)–step count (Fitbit Versa)	Step count (ActiGraph GT3X)–energy expenditure (Fitbit Versa)
Seated rest				
ICC[a] (95% CI)	N/A[b]	N/A	N/A	N/A
Mean difference (LOA[c])	0.00 (–0.33 to 0.34)	N/A	N/A	N/A
ρ (*P* value)	N/A	N/A	N/A	N/A
MAPE[d]	0.36	N/A	N/A	N/A
Treadmill walk				
3.0 km/h[e]				
ICC (95% CI)	–0.01 (–0.03 to 0.04)	0.71 (0.44 to 0.84)	N/A	N/A
Mean difference (LOA)	4.43 (2.92 to 5.94)	0.84 (–15.73 to 17.40)	N/A	N/A
ρ (*P* value)	–0.26 (.11)	0.66 (<.001)	–0.08 (.60)	0.42 (.01)
MAPE	76.86	5.39	N/A	N/A
4.5 km/h[f]				
ICC (95% CI)	0.02 (–0.12 to 0.21)	0.69 (0.29 to 0.85)	N/A	N/A
Mean difference (LOA)	2.55 (–0.36 to 5.45)	4.11 (–8.13 to 16.35)	N/A	N/A
ρ (*P* value)	0.11 (.51)	0.66 (<.001)	0.07 (.66)	0.60 (<.001)
MAPE	39.44	4.98	N/A	N/A
6.0 km/h[g]				
ICC (95% CI)	–0.14 (–0.50 to 0.24)	0.05 (–0.51 to 0.35)	N/A	N/A
Mean difference (LOA)	2.08 (–1.30 to 5.45)	10.79 (–15.02 to 36.61)	N/A	N/A
ρ (*P* value)	–0.07 (.67)	0.09 (.60)	0.23 (.18)	0.37 (.03)
MAPE	25.01	11.15	N/A	N/A
Overall speeds[h]				
ICC (95% CI)	–0.04 (–0.13 to 0.12)	0.60 (0.03 to 0.82)	N/A	N/A
Mean difference (LOA)	3.02 (0.78 to 5.26)	–4.98 (–16.68 to 6.71)	N/A	N/A
ρ (*P* value)	–0.10 (.56)	0.51 (<.002)	0.28 (.11)	0.31 (.07)
MAPE	39.41	5.41	N/A	N/A
Free-living				
Overall days[i]				
ICC (95% CI)	0.46 (–0.16 to 0.80)	0.70 (–0.21 to 0.91)	N/A	N/A
Mean difference (LOA)	1.20 (0.17 to 2.24)	–7.12 (–16.25 to 2.00)	N/A	N/A
ρ (*P* value)	0.79 (<.001)	0.87 (<.001)	0.41 (.01)	0.55 (<.001)
MAPE	31.11	82.45	N/A	N/A

[a]ICC: intraclass correlation.

[b]N/A: not applicable.

[c]LOA: limits of agreement.

[d]MAPE: mean absolute percent error.

[e]n=39 for energy expenditure–energy expenditure; n=40 for step count–step count, n=38 for energy expenditure–step count.

[f]n=40 for energy expenditure–energy expenditure and step count–energy expenditure.

[g]n=36 for energy expenditure–energy expenditure; n=34 for step count–step count; n=35 for energy expenditure–step count and step count–energy expenditure.

[h]n=35 for energy expenditure–energy expenditure; n=34 for step count–step count; n=35 for energy expenditure–step count; n=34 for step count–energy expenditure.

[i]n=38 for energy expenditure–energy expenditure and step count–energy expenditure; n=37 for step count–step count and energy expenditure–step count.

Figure 1. Bland-Altman plot visualizing agreement of energy expenditure (MET) estimated by Fitbit Versa and criterion measurement Jaeger Oxycon Pro during overall treadmill walk. The middle green line shows the mean difference (bias) between devices. The dashed lines indicate upper (+1.96 SD) and lower (–1.96 SD) limits of agreement and the black line represents the regression line illustrating association between estimations.

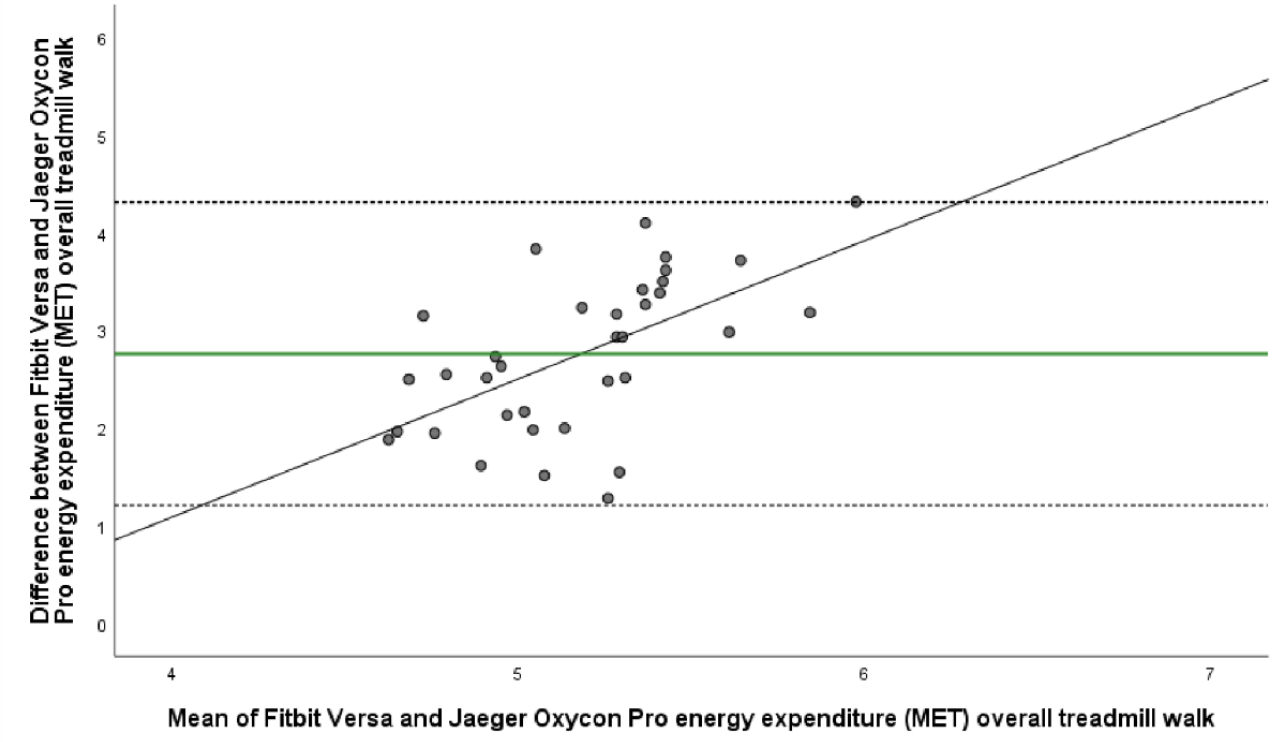


Figure 2. Bland-Altman plot visualizing agreement of heartrate estimated by Fitbit Versa and criterion measurement Jaeger Oxycon Pro during overall treadmill walk. The middle green line shows the mean difference (bias) between devices. The dashed lines indicate upper (+1.96 SD) and lower (–1.96 SD) limits of agreement and the black line represents the regression line illustrating association between estimations.

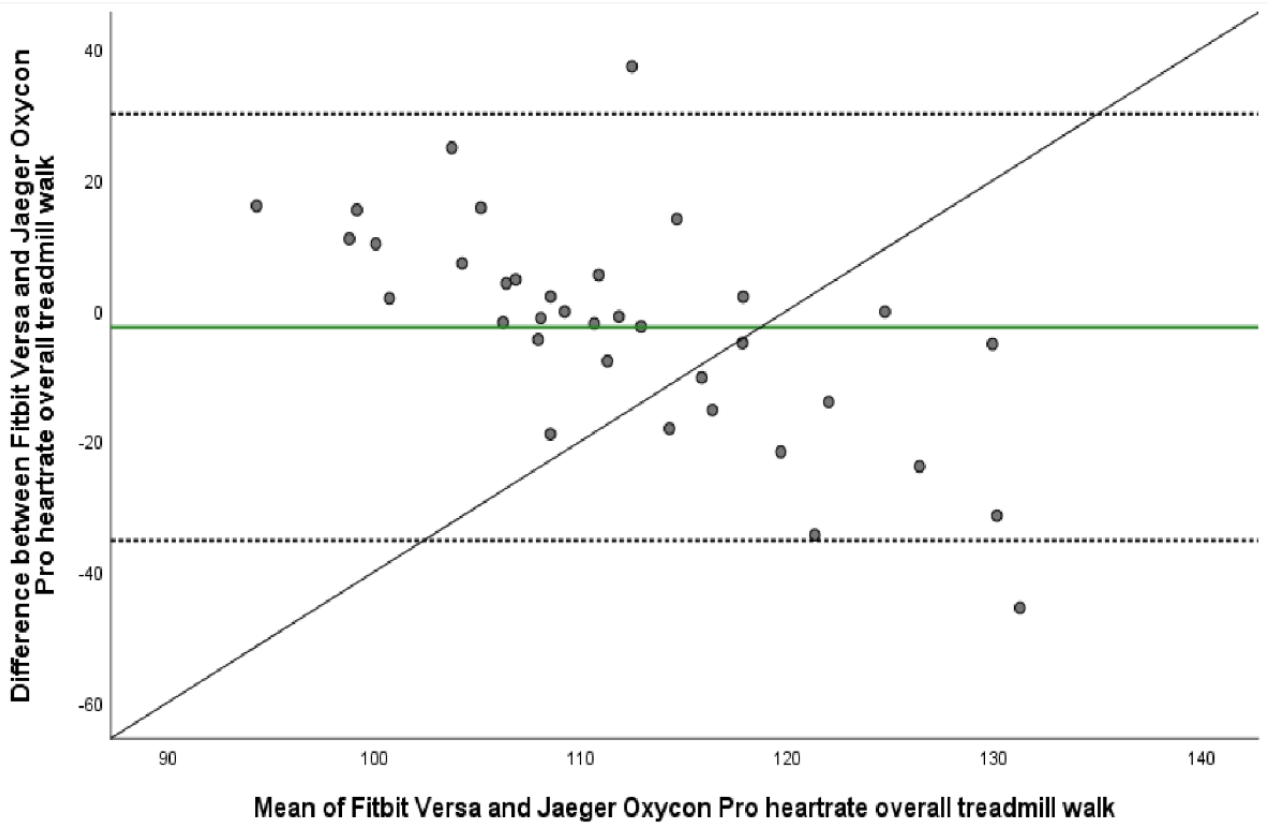


Figure 3. Bland-Altman plot visualizing agreement of energy expenditure (MET) estimated by Fitbit Versa and relative criterion measurement ActiGraph GT3X during overall treadmill walk. The middle green line shows the mean difference (bias) between devices. The dashed lines indicate upper (+1.96 SD) and lower (–1.96 SD) limits of agreement and the black line represents the regression line illustrating association between estimations.

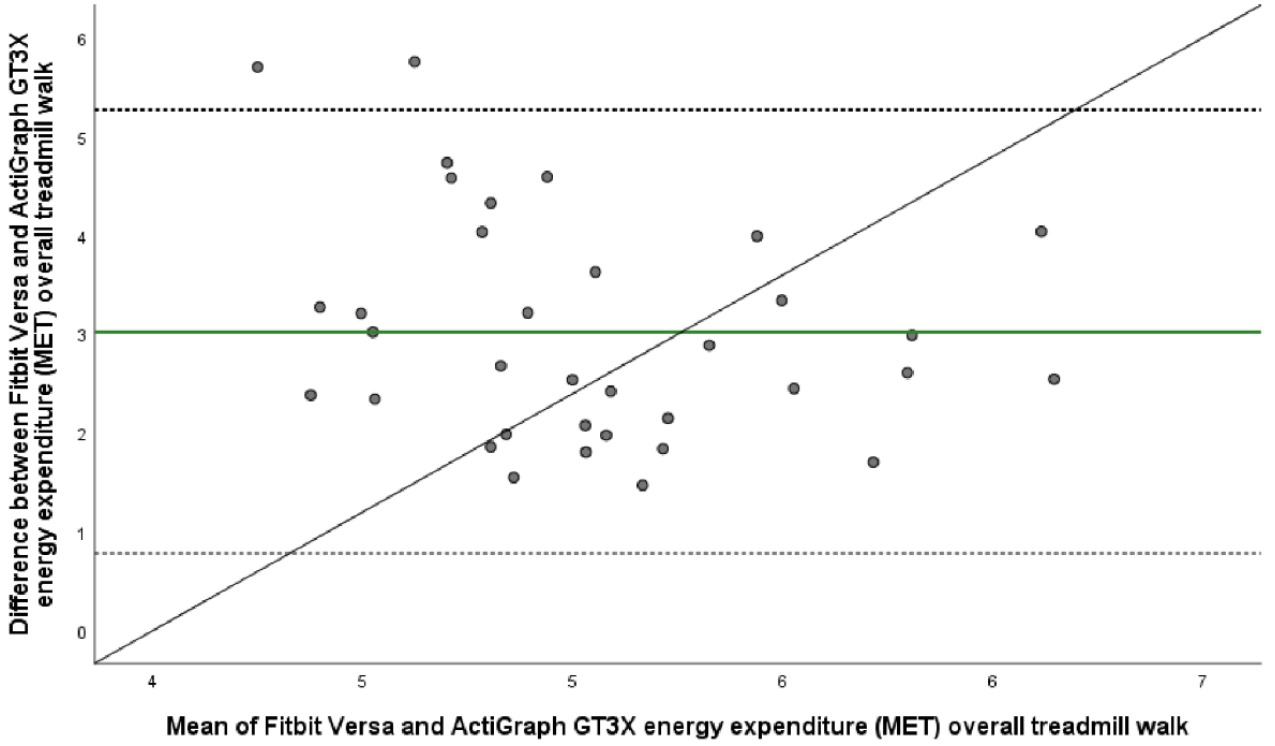


Figure 4. Bland-Altman plot visualizing agreement of step count estimated by Fitbit Versa and relative criterion measurement ActiGraph GT3X during overall treadmill walk. The middle green line shows the mean difference (bias) between devices. The dashed lines indicate upper (+1.96 SD) and lower (–1.96 SD) limits of agreement and the black line represents the regression line illustrating association between estimations.

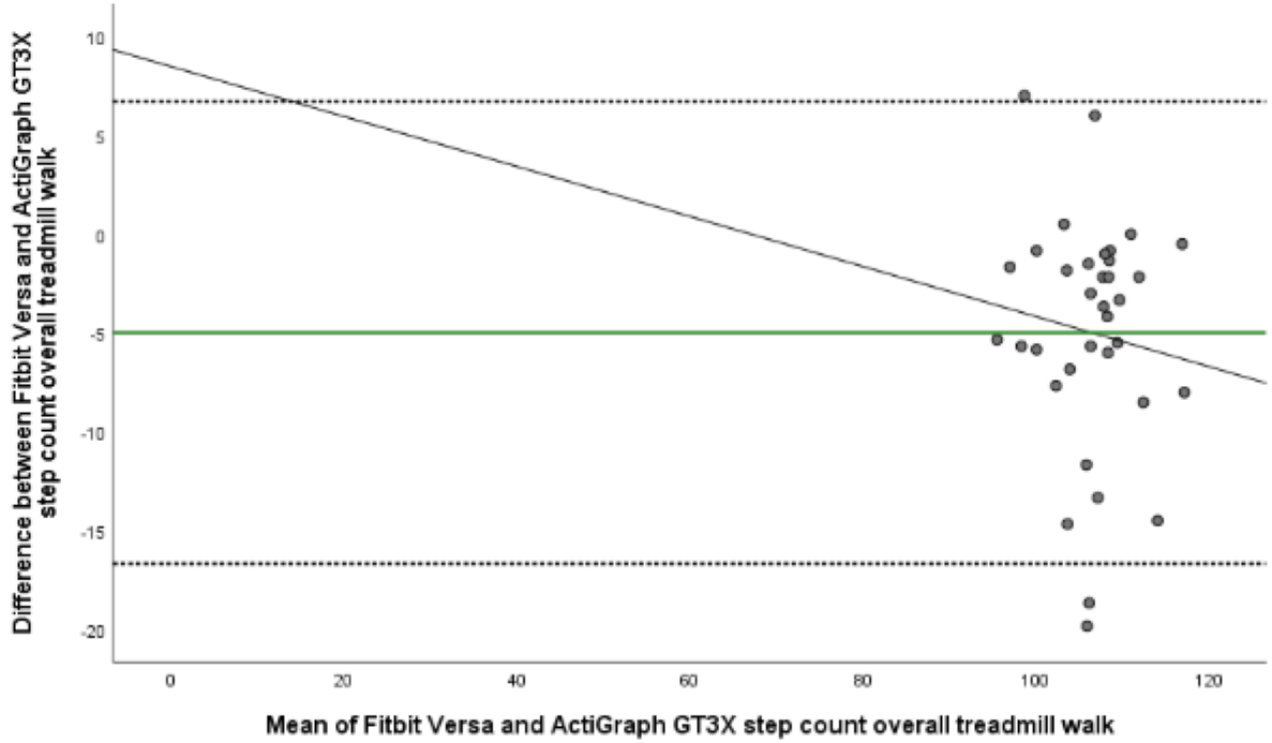
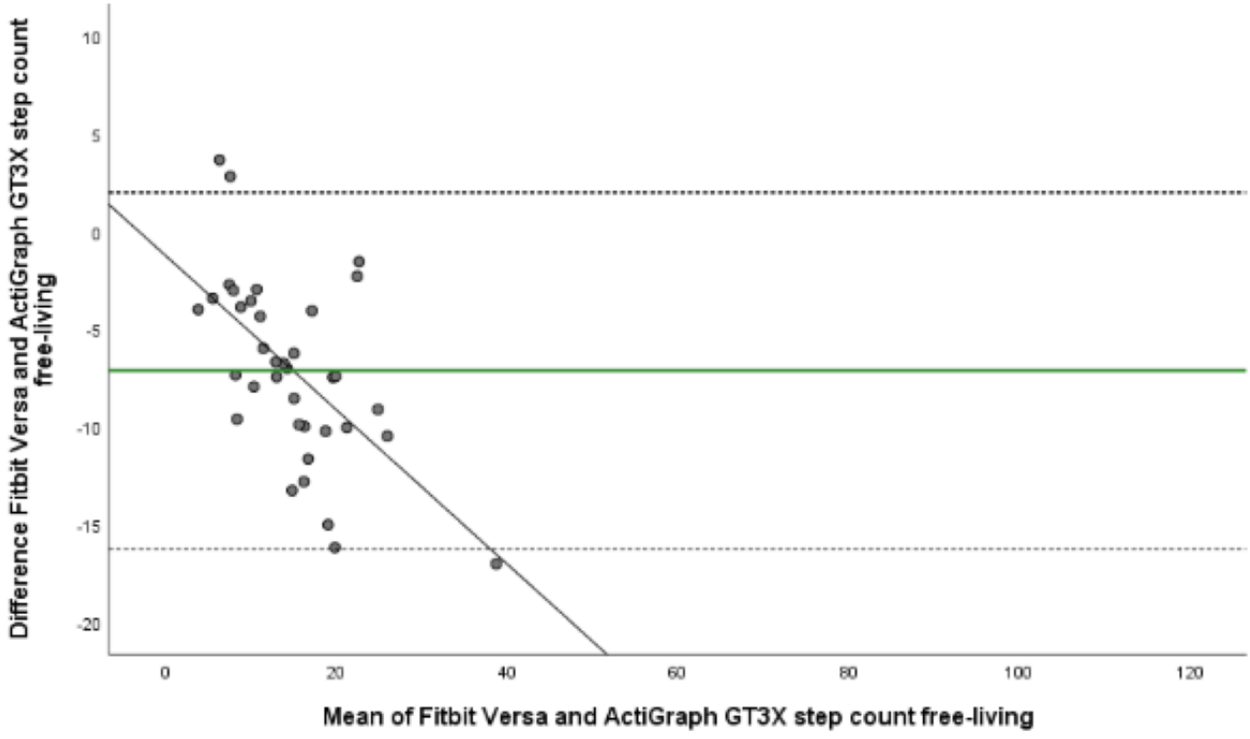


Figure 5. Bland-Altman plot visualizing agreement of step count estimated by Fitbit Versa and relative criterion measurement ActiGraph GT3X during free-living. The middle green line shows the mean difference (bias) between devices. The dashed lines indicate upper (+1.96 SD) and lower (–1.96 SD) limits of agreement and the black line represents the regression line illustrating association between estimations.



Fitbit Versa versus Jaeger Oxycon Pro

In the laboratory setting we found that Fitbit Versa showed poor agreement of estimated energy expenditure with corresponding estimations by Jaeger Oxycon Pro in the overall treadmill walk (ICC –0.03, 95% CI –0.08 to 0.08). There were also significant systematic differences between estimations in all treadmill speeds as well as in the overall treadmill walk ($P \leq .001$). In

addition, the Bland-Altman plot showed a broad range for energy expenditure estimation, also indicated the overestimation, with a mean difference of 2.76 MET, LOA 1.21 to 4.31 for overall speeds (Table 4, Figure 1). A narrow mean difference was found during rest, 0.27 MET, LOA –0.07 to 0.61. In addition, the correlation of energy expenditure estimated by Fitbit Versa and Jaeger Oxycon Pro was weak at all measured timepoints. Overall treadmill speed MAPE for energy expenditure was 35.39 and ranged from 31.59 at 6 km/h to 51.52 at 3.0 km/h.

There was poor agreement between Fitbit Versa's estimation of heart rate compared to Jaeger Oxycon Pro at overall treadmill (ICC 0.19, 95% CI –0.58 to 0.59). At the specific treadmill speeds ICC ranged from poor (ICC 0.09, 95% CI –0.72 to 0.52) at 3.0 km/h to fair (ICC 0.40, 95% CI –0.09 to 0.68) at the final treadmill speed (6 km/h). However, agreement of estimations was excellent (ICC 0.99, 95% CI 0.98 to 0.99) and correlation very strong (ρ=.96, *P*≤.001) during seated rest. ANOVA results showed no systematic differences between estimations of heart rate during rest (*P*=.81), at 3 km/h (*P*=.37), at 4.5 km/h (*P*=.77), or during the overall treadmill walk (*P*=.34). This was also confirmed by the Bland-Altman plot; the mean difference of heart rate estimation during the overall treadmill walk were –2.68 bpm, LOA –35.30 to 29.95 bpm. It ranged from –2.68 bpm, LOA –39.01 to 33.65 at 3.0 km/h to a broader range, 10.19 bpm, LOA –25.45 to 46.57 at 6 km/h (Table 4, Figure 2). Corresponding MAPE ranged from 2.24 at seated rest to 12.10 at 6 km/h, with the overall treadmill walk at 10.51.

We found only weak correlations between energy expenditure by Jaeger Oxycon Pro and heart rate by Fitbit Versa, and between energy expenditure by Jaeger Oxycon Pro and step count by Fitbit Versa, during both seated rest and during all treadmill speeds (Table 4).

In accordance with findings of poor agreement between Fitbit Versa and the criterion measurement's (Jaeger Oxycon Pro) estimations of energy expenditure, we also found poor agreement between corresponding estimations by Fitbit Versa and the relative criterion measurement ActiGraph GT3X, at all treadmill speeds (Table 4). For the overall treadmill walk, the agreement was poor (ICC –0.04, 95% CI –0.13 to 0.12) as it also was at specific treadmill speeds (Table 5).

Fitbit Versa versus ActiGraph GT3X

Due to zero variation in data, ICC calculations of energy expenditure estimated by ActiGraph GT3X and Fitbit Versa during seated rest were not possible to perform. The Bland-Altman plot provided a mean difference of 0.00 MET, LOA –0.33 to 0.34 to for seated rest indicating a high agreement in estimations of heart rate between the devices (Table 4). Also, there were minimal individual differences between measurements during rest (MAPE 0.36) but greater differences (MAPE 76.86) at 3.0 km/h, however they decreased as treadmill speed increased (MAPE 39.44 at 4.5 km/h, MAPE 25.01 at 6 km/h) (Table 5).

Findings suggest a fair agreement (ICC 0.54, 95% CI 0.02 to 0.78) and a strong significant correlation (ρ=0.51, *P*≤.001) of step count estimations by Fitbit Versa and ActiGraph GT3X at the overall treadmill level (Table 5, Figure 3). At specific treadmill speeds, the agreement was good at both 3.0 km/h (ICC 0.71, 95% CI 0.44 to 0.84) and at 4.5 km/h (ICC 0.69, 95% CI 0.29 to 0.85), but decreased at 6 km/h (ICC 0.05, 95% CI –0.51 to 0.35) (Table 5).

There was fair and significant correlation in step count between devices in 2 out of 3 treadmill speeds (3.0 km/h, 4.5 km/h) and the overall treadmill walk (ρ=0.51, *P*≤.001). The ANOVA results were significant for the overall treadmill walk and at the 2 higher treadmill speeds (*P*≤.001) while the Bland-Altman plots showed a mean difference at the overall speed by –4.98 steps, LOA –16.68 to 6.71 (Table 5, Figure 4). MAPE ranged from 5.39 at 3.0 km/h to 11.15 at 6 km/h, with 5.41 for the overall treadmill walk (Table 5). The correlation between ActiGraph GT3X estimations of energy expenditure and Fitbit estimations of step count were weak for the treadmill walk in the laboratory setting. However, the correlation between ActiGraph GT3X estimations of step count and Fitbit estimations of energy expenditure were significant and fair for the slowest (ρ=0.42, *P*=.01) and fastest (ρ=0.37, *P*=.37) treadmill speed. Moderate and significant correlation (ρ=0.60, *P*≤.001) was found at 4.5 km/h (Table 5).

In the free-living setting, we found fair agreement between Fitbit Versa and ActiGraph GT3X's estimations of energy expenditure (ICC 0.46, 95% CI –0.16 to 0.80), and a significant and strong correlation (ρ=0.79, *P*≤.001). ANOVA results show no systematic differences between estimations (*P*≤.001), which is confirmed by the Bland-Altman plot mean difference by 1.20 MET, LOA 0.17 to 2.24 MET and MAPE 31.11 (Table 5). The agreement between Fitbit Versa and ActiGraph GT3X's estimations of step count were good (ICC 0.70, 95% CI –0.21 to 0.91) and the correlation between estimations was strong (ρ=0.87, *P*≤.001). ANOVA results showed no systematic differences between step count estimations (*P*≤.001). Bland-Altman plot showed a mean difference with –7.12 steps, LOA –16.25 to 2.00 confirming an agreement (Figure 5). MAPE, on the other hand, was 82.45, indicating great individual bias (Table 5).

The correlation between ActiGraph GT3X estimations of energy expenditure and Fitbit Versa estimations of step count were significant and fair (ρ=0.41, *P*=.01). A corresponding association was found (ρ=0.55, *P*≤.001) between ActiGraph GT3X's step count, and Fitbit Versa's estimation of energy expenditure (Table 5).

Discussion

Principal Findings

To our knowledge, this is the first study that has evaluated criterion validity of Fitbit Versa's estimations of energy expenditure, step count, and heart rate for patients with chronic pain. Evaluations of criterion validity wrist-worn outputs of energy expenditure, heart rate, and step count is essential before any clinical application may be implemented [39]. Poor agreement (ICC, mean difference and LOA, MAPE) as well as poor correlation were found between the criterion measurement (Jaeger Oxycon Pro) and the experimental measurement (Fitbit

Versa) regarding energy expenditure for the overall treadmill walk as well as the 3 specific treadmill speeds (Table 4). However, good agreement and fair correlation emerged between estimations of step count by Fitbit Versa and ActiGraph GT3X for the majority of the treadmill walk as well as the overall treadmill walk (Table 5). Good agreement and correlation were shown for the estimation of heart rate during seated rest as well, but this decreased during all treadmill speeds.

Comparison With Previous Studies

Findings suggest that Fitbit Versa systematically overestimated energy expenditure across the full range of the testing protocol when compared to both the criterion (Jaeger Oxycon Pro) and the relative criterion measurement (ActiGraph GT3X). A strict comparison of our study findings with other research within this population was not possible due to the lack of such studies. On the other hand, studies have been conducted aiming to evaluate criterion validity in wrist-worn activity trackers among an elderly population [56] as well as among populations suffering from chronic cardiac conditions [57,58]. These reports also suggest an overestimation of energy expenditure. Herkert et al [57] studied the accuracy of Fitbit Charge 2 among patients with chronic heart conditions and compared estimations of energy expenditure with indirect calorimetry (Oxycon Mobile) during several household activities and a treadmill walk (4.0 km/h, 5.5 km/h, 4.0 km/h + 5% slope). While their findings are not strictly applicable to our sample, both samples included patients with physically impairments and findings suggested a clear overestimation of energy expenditure [57].

The results of other studies [19,24], conducted primarily among healthy participants and examining the validity of other Fitbit models' ability to estimate energy expenditure, contradict our results—certain Fitbit models (Fitbit, Fitbit Ultra/Fitbit Zio, Fitbit Flex) underestimated energy expenditure and step count when data were compared to the criterion measurements. Furthermore, we found good agreement between step count by Fitbit Versa and ActiGraph GT3X for the first 2 treadmill speeds, as well as fair agreement for the overall treadmill test, but agreement decreased at 6 km/h. In the free-living setting, we found good agreement and excellent correlation between step count estimation by Fitbit Versa and ActiGraph GT3X. This corresponds to the findings of a study [59] that compared step count estimation of healthy participants in a free-living setting using a Fitbit device (Fitbit One) and ActiGraph GT3X, and reported excellent agreement between the 2 measurements.

The overestimation of energy expenditure found in our study could be explained by the proprietary algorithms applied by Fitbit, which are not tailored to specific populations. In the specific population related to this study, altered movement patterns is indicated due to changed motor control and kinematics as well as due to a fear of pain causing a protective avoidance in movement and activity [60-62]. Fitbit's estimation of energy expenditure is based on both body composition metrics as well as (if available in the device) estimated heart rate [33,34], which requires a valid heart rate measurement. Our findings indicate a poor to fair criterion validity in estimations of heart rate for all treadmill speeds, which is consistent with previous studies [28,63]. Another important factor that may have influenced our findings is the placement of devices on the body. Our experimental device was placed on participants' wrists according to the manufacturer's instructions, and the relative criterion measurement device were placed around participants' waists, near their right hip (also according to the manufacturer's instructions). Feehan with colleges concludes that placing devices on the wrist generally leads to an overestimation of energy expenditure which may be explained by the waist sensor being placed closer to the center of the body [23].

Strengths and Limitations

Data collection was performed in both a laboratory, limiting many confounding variables, and in a natural free-living context, increasing the ecological validity, which is in line with current recommendations for validity studies examining wearable monitors for physical activity [39,40]. In this study, conventional statistics were applied in order to conduct a comprehensive evaluation of the criterion validity and report findings in a standardized manner [54]. Equivalence testing is also recommended as it provides both a risk evaluation of measurement agreement and zones of equivalence between estimations is established by consensus [54]. This method was not applied in this study, which may be a limitation, as knowledge of statistically significant risk of misclassified physical activity–level would certainly contribute to the interpretation of the results. The sample size was in accordance with the initial sample size calculation and equivalent to corresponding research studies [20,22,57,59,64]. Furthermore, when conducting validation studies, guidelines state that one should perform measurements in a large range of physical activity intensities [39]. In this study, treadmill speed was set to a maximum of 6 km/h which may be interpreted as a low intensity; however, we suspected that a higher speed could have been problematic for some participants. The fact that 5 of the 41 participants were unable to complete the treadmill walk at 6 km/h indicates that a higher treadmill speed could have resulted in a higher number of discontinuations. A possible source of bias in validation studies is pharmacological use of beta blockers because it affects heart rate and possibly biases evaluation of physical activity intensity level. In this study, only 2 participants (<5%) reported taking beta blockers within the last 24 hours which may have affected participants perceived exertion during treadmill use. However, we estimate this having a very little impact on our results.

Our sample included 76% women (31/41), which is in line with other studies describing people with chronic pain in Sweden [65,66] and other countries [67]. Patients in our sample had lived with pain for a shorter time than what is described in other studies of this population [68,69], and they rated their pain severity at baseline as equal to what have been reported in another study [70] describing patients participating in primary care management of low back pain patients. In our study, 37% of participants reported an exercise level of >90 minutes/week and 49% reported being physically active >150 minutes/week. A previous study [71] reported that that 38% of participating men and 37% of participating women with chronic pain were physically active >60 minutes/week. Since almost half of our participants report reaching the recommended weekly amounts of physical activity, it seems that our sample is slightly more

physically active than the population with chronic pain, in general. In all, we believe that the external validity can be extended to the major group of individuals with chronic pain seeking care in primary and specialist care.

Conclusions

This study provides new knowledge on the criterion validity of Fitbit Versa's estimations of energy expenditure, heart rate, and step count in patients with chronic pain. Findings show that Fitbit Versa overestimates energy expenditure when compared to criterion estimations in a controlled laboratory setting as well as in free-living settings, which needs to be considered when used clinically for patients with chronic pain. Step count measured from the wrist, however, seems to provide a valid estimation, suggesting that future guidelines should include this variable in this major patient group. Findings may contribute to the solicited documentation of estimation properties of wrist-worn activity tracking devices within specific patient groups and may therefore guide future application in further clinical research.

Acknowledgments

The authors would like to thank all our study participants who kindly devoted their time to the study. We are also truly grateful to all the involved staff at the primary and specialized health care clinics in Region Dalarna for their invaluable support in the recruitment process. We want to acknowledge Linnea McCarthy, MSc, at Dalarna University for her support with cross-checking data, and to Anton Westman, PhD, for his initial support with the Actilife software.

The study was funded by the Swedish Research Council for Health, Working Life and Welfare (registration number 2017-00491 and 2017-00177), the Swedish Research Council (registration number 2018-02455 and 2015-02512), and research funding from Dalarna University. Sponsors had no part in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication.

Authors' Contributions

LV and BÄ are responsible for the conception and design of the study. VS and JW contributed to study design, recruited participants, and collected all data. VS conducted data management, data reduction, and all statistical analyses with guidance from AM, BÄ, RLM, and LV. VS, JW, LV, and BÄ had full access to all data. BÄ was the principal investigator and was involved in all methodological decisions. RLM assisted with the retrieval of data from the commercial activity tracker and provided valuable insights regarding the interpretation of the results. MH contributed with field expertise throughout all stages of data collection, statistical analysis, and writing of the manuscript. VS drafted the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. IASP terminology. International Association for the Study of Pain (IASP). URL: https://www.iasp-pain.org/terminology?navItemNumber=576 [accessed 2020-08-18]
2. Merskey H, Bogduk N. The present classification. In: Classification of Chronic Pain. Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms. Seattle: IASP Press; 1994:11-11.
3. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Apr 24;4:CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
4. Piercy KL, Troiano RP. Physical activity guidelines for Americans from the US Department of Health and Human Services. Circ Cardiovasc Qual Outcomes 2018 Nov;11(11):e005263. [doi: 10.1161/CIRCOUTCOMES.118.005263] [Medline: 30571339]
5. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
6. Ambrose KR, Golightly YM. Physical exercise as non-pharmacological treatment of chronic pain: why and when. Best Practice & Research Clinical Rheumatology 2015 Feb;29(1):120-130. [doi: 10.1016/j.berh.2015.04.022]
7. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]
8. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [Medline: 10993420]
9. Shepard R, Tudor-Locke C. A history of physical activity measurement in epidemiology. In: Shephard RJ, Tudor-Locke C, editors. The objective monitoring of physical activity: Contributions of Accelerometry to Epidemiology, Exercise, Science and Rehabilitation. The Objective Monitoring of Physical Activity: Contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation. 1st edition. Switzerlandpringer International Publishing; 2016:39-83.

10. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
11. Bai Y, Welk GJ, Nam YH, Lee JA, Lee J, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2015 Jul 6. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
12. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
13. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
14. Le Hello C, Trombert B, Morel A, Chieh A, Brouard B, Boissier C. Performance analysis of walking of 10,000 regular users of a connected activity tracker. J Med Vasc 2018 Jul;43(4):231-237. [doi: 10.1016/j.jdmv.2018.04.001] [Medline: 29981731]
15. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]
16. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mhealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594. [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
17. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
18. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140. [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]
19. Adam NJ, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol 2013 Oct;37(7):456-462. [doi: 10.3109/03091902.2013.831135] [Medline: 24007317]
20. Gusmer R, Bosch T, Watkins A, Ostrem J, Dengel D. Comparison of FitBit® Ultra to ActiGraph™ GT1M for assessment of physical activity in young adults during treadmill walking. TOSMJ 2014 Apr 04;8(1):11-15. [doi: 10.2174/1874387001408010011]
21. Stahl ST, Insana SP. Caloric expenditure assessment among older adults: criterion validity of a novel accelerometry device. J Health Psychol 2014 Nov;19(11):1382-1387 [FREE Full text] [doi: 10.1177/1359105313490771] [Medline: 23818504]
22. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
23. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
24. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2017 May 08. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]
25. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. Med Sci Sports Exerc 2013 Nov;45(11):2105-2112 [FREE Full text] [doi: 10.1249/MSS.0b013e318299d2eb] [Medline: 23669877]
26. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
27. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van DSCP, de GM. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
28. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
29. 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]
30. Bästa frågan till patienter om fysisk aktivitet. GIH. URL: https://www.gih.se/OM-GIH/Press-och-media/Pressmeddelanden/Basta-fragan-om-fysisk-aktivitet-till-patienter/ [accessed 2020-08-20]
31. Forskningsstudie om validering av indikatorfrågor till patienter om fysisk aktivitet. Socialstyrelsen. 2013. URL: http://www.hfsnatverket.se/static/files/1185/forskningsstudie_validering_av_indikatorfragor_till_patienter_om_fysisk_aktivitet.pdf [accessed 2020-08-20]

32. Bergström G, Jensen IB, Bodin L, Linton SJ, Nygren AL, Carlsson SG. Reliability and factor structure of the Multidimensional Pain Inventory--Swedish language version (MPI-S). Pain 1998 Mar;75(1):101-110. [doi: 10.1016/s0304-3959(97)00210-8] [Medline: 9539679]
33. How does my Fitbit device calculate calories burned? Fitbit. URL: https://myhelp.fitbit.com/articles/en_US/Help_article/1141#calories [accessed 2020-08-30]
34. How does my Fitbit device calculate active minutes? Fitbit. URL: https://help.fitbit.com/articles/en_US/Help_article/1379.htm [accessed 2020-08-30]
35. Haugen HA, Chan L, Li F. Indirect calorimetry: a practical guide for clinicians. Nutr Clin Pract 2007 Aug;22(4):377-388. [doi: 10.1177/0115426507022004377] [Medline: 17644692]
36. Rietjens GJ, Kuipers H, Kester AD, Keizer HA. Validation of a computerized metabolic measurement system (Oxycon-Pro) during low and high intensity exercise. Int J Sports Med 2001 May;22(4):291-294. [doi: 10.1055/s-2001-14342] [Medline: 11414673]
37. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]
38. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
39. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S32-S38 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399cf7] [Medline: 22157772]
40. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exercise and Sport Sciences Reviews 2019;47(4):206-214. [doi: 10.1249/jes.0000000000000206]
41. Dowd KP, Harrington DM, Donnelly AE. Criterion and concurrent validity of the activPAL™ professional physical activity monitor in adolescent females. PLoS One 2012;7(10):e47633 [FREE Full text] [doi: 10.1371/journal.pone.0047633] [Medline: 23094069]
42. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
43. Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. Acad Emerg Med 2001 Dec;8(12):1153-1157. [Medline: 11733293]
44. Web API reference. Fitbit Platform. URL: https://dev.fitbit.com/build/reference/web-api/ [accessed 2021-01-04]
45. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 7;19(1). [doi: 10.1186/s12874-019-0668-1]
46. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol 2011 Feb;111(2):187-201 [FREE Full text] [doi: 10.1007/s00421-010-1639-8] [Medline: 20842375]
47. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
48. Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with continuous measurements. J Biopharm Stat 2007;17(4):529-569. [doi: 10.1080/10543400701376480] [Medline: 17613641]
49. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation - a discussion and demonstration of basic features. PLoS One 2019;14(7):e0219854 [FREE Full text] [doi: 10.1371/journal.pone.0219854] [Medline: 31329615]
50. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
51. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
52. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
53. Chan YH. Biostatistics 104: correlational analysis. Singapore Med J 2003 Dec;44(12):614-619 [FREE Full text] [Medline: 14770254]
54. Welk GJ, Bai Y, Lee J, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Medicine & Science in Sports & Exercise 2019;51(8):1767-1780. [doi: 10.1249/mss.0000000000001966]
55. Fox-Wasylyshyn SM, El-Masri MM. Handling missing data in self-report measures. Res Nurs Health 2005 Dec;28(6):488-495. [doi: 10.1002/nur.20100] [Medline: 16287052]
56. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. PLoS ONE 2019 May 21;14(5):e0216891. [doi: 10.1371/journal.pone.0216891]
57. Herkert C, Kraal JJ, van Loon EMA, van Hooff M, Kemps HMC. Usefulness of modern activity trackers for monitoring exercise behavior in chronic cardiac patients: validation study. JMIR Mhealth Uhealth 2019 Dec 19;7(12):e15045 [FREE Full text] [doi: 10.2196/15045] [Medline: 31855191]
58. Vetrovsky T, Siranec M, Marencakova J, Tufano JJ, Capek V, Bunc V, et al. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. PLoS One 2019;14(9):e0222569 [FREE Full text] [doi: 10.1371/journal.pone.0222569] [Medline: 31518367]

59. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]
60. Meulders A. From fear of movement-related pain and avoidance to chronic pain disability: a state-of-the-art review. Current Opinion in Behavioral Sciences 2019 Apr;26:130-136. [doi: 10.1016/j.cobeha.2018.12.007]
61. Burnett AF, Cornelius MW, Dankaerts W, O'sullivan PB. Spinal kinematics and trunk muscle activity in cyclists: a comparison between healthy controls and non-specific chronic low back pain subjects-a pilot investigation. Man Ther 2004 Nov;9(4):211-219. [doi: 10.1016/j.math.2004.06.002] [Medline: 15522646]
62. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85(3):317-332. [doi: 10.1016/s0304-3959(99)00242-0]
63. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
64. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
65. Tseli E, Vixner L, LoMartire R, Grooten WJA, Gerdle B, Äng BO. Prognostic factors for improved physical and emotional functioning one year after interdisciplinary rehabilitation in patients with chronic pain: Results from a national quality registry in Sweden. J Rehabil Med 2020 Feb 27;52(2):jrm00019 [FREE Full text] [doi: 10.2340/16501977-2648] [Medline: 31995224]
66. Gerdle B, Björk J, Cöster L, Henriksson K, Henriksson C, Bengtsson A. Prevalence of widespread pain and associations with work status: a population study. BMC Musculoskelet Disord 2008 Jul 15;9:102 [FREE Full text] [doi: 10.1186/1471-2474-9-102] [Medline: 18627605]
67. Macfarlane GJ, Morris S, Hunt IM, Benjamin S, McBeth J, Papageorgiou AC, et al. Chronic widespread pain in the community: the influence of psychological symptoms and mental disorder on healthcare seeking behavior. J Rheumatol 1999 Feb;26(2):413-419. [Medline: 9972978]
68. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
69. Gerdle B, Boersma K, Åsenlöf P, Stålnacke B, Larsson B, Ringqvist Å. Influences of sex, education, and country of birth on clinical presentations and overall outcomes of interdisciplinary pain rehabilitation in chronic pain patients: a cohort study from the Swedish Quality Registry for Pain Rehabilitation (SQRP). J Clin Med 2020 Jul 25;9(8) [FREE Full text] [doi: 10.3390/jcm9082374] [Medline: 32722367]
70. Johansson E, Lindberg P. Low back pain patients in primary care: Subgroups based on the multidimensional pain inventory. Int J Behav Med 2000 Dec;7(4):340-352. [doi: 10.1207/S15327558IJBM0704_05]
71. Nilsen TIL, Holtermann A, Mork PJ. Physical exercise, body mass index, and risk of chronic pain in the low back and neck/shoulders: longitudinal data from the Nord-Trondelag Health Study. Am J Epidemiol 2011 Aug 01;174(3):267-273. [doi: 10.1093/aje/kwr087] [Medline: 21633119]

Abbreviations

ANOVA: analysis of variance
ICC: intraclass correlation coefficient
LOA: limits of agreement
MAPE: mean absolute percent error
MVPA: moderate-to-vigorous physical activity

Edited by G Eysenbach; submitted 06.10.20; peer-reviewed by J Kraal, B Boudreaux; comments to author 28.10.20; revised version received 06.12.20; accepted 12.12.20; published 12.01.21.

Please cite as:
Sjöberg V, Westergren J, Monnier A, Lo Martire R, Hagströmer M, Äng BO, Vixner L
Wrist-Worn Activity Trackers in Laboratory and Free-Living Settings for Patients With Chronic Pain: Criterion Validity Study
JMIR Mhealth Uhealth 2021;9(1):e24806
URL: http://mhealth.jmir.org/2021/1/e24806/
doi:10.2196/24806
PMID:33433391

Original Paper

Physical Activity Monitoring Using a Fitbit Device in Ischemic Stroke Patients: Prospective Cohort Feasibility Study

Irene Katzan[1,2], MD, MSc; Andrew Schuster[1], BA; Tyler Kinzy[1], MSc

[1]Neurological Institute Center for Outcomes Research and Evaluation, Cleveland Clinic, Cleveland, OH, United States

[2]Cerebrovascular Center, Cleveland Clinic, Cleveland, OH, United States

Corresponding Author:
Irene Katzan, MD, MSc
Neurological Institute Center for Outcomes Research and Evaluation
Cleveland Clinic
9500 Euclid Avenue, S80
Cleveland, OH, 44195
United States
Phone: 1 12164452616
Email: katzani@ccf.org

Abstract

Background: Continuous tracking of ambulatory activity in real-world settings using step activity monitors has many potential uses. However, feasibility, accuracy, and correlation with performance measures in stroke patients have not been well-established.

Objective: The primary study objective was to determine adherence with wearing a consumer-grade step activity monitor, the Fitbit Charge HR, in home-going ischemic stroke patients during the first 90 days after hospital discharge. Secondary objectives were to (1) determine accuracy of step counts of the Fitbit Charge HR compared with a manual tally; (2) calculate correlations between the Fitbit step counts and the mobility performance scores at discharge and 30 days after stroke; (3) determine variability and change in weekly step counts over 90 days; and (4) evaluate patient experience with using the Fitbit Charge HR poststroke.

Methods: A total of 15 participants with recent mild ischemic stroke wore a Fitbit Charge HR for 90 days after discharge and completed 3 mobility performance tests from the National Institutes of Health Toolbox at discharge and Day 30: (1) Standing Balance Test, (2) 2-Minute Walk Endurance Test, and (3) 4-Meter Walk Gait Speed Test. Accuracy of step activity monitors was assessed by calculating differences in steps recorded on the step activity monitor and a manual tally during 2-minute walk tests.

Results: Participants had a mean age of 54 years and a median modified Rankin scale score of 1. Mean daily adherence with step activity monitor use was 83.6%. Mean daily step count in the first week after discharge was 4376. Daily step counts increased slightly during the first 30 days after discharge (average increase of 52.5 steps/day; 95% CI 32.2-71.8) and remained stable during the 30-90 day period after discharge. Mean step count difference between step activity monitor and manual tally was –4.8 steps (–1.8%). Intraclass correlation coefficients for step counts and 2-minute walk, standing balance, and 4-meter gait speed at discharge were 0.41 (95% CI –0.14 to 0.75), –0.12 (95% CI –0.67 to 0.64), and 0.17 (95% CI –0.46 to 0.66), respectively. Values were similarly poor at 30 days.

Conclusions: The use of consumer-grade Fitbit Charge HR in patients with recent mild stroke is feasible with reasonable adherence and accuracy. There was poor correlation between step counts and gait speed, balance, and endurance. Further research is needed to evaluate the association between step counts and other outcomes relevant to patients, including patient-reported outcomes and measures of physical function.

(JMIR Mhealth Uhealth 2021;9(1):e14494) doi:10.2196/14494

KEYWORDS

physical activity; accelerometer; ischemic stroke; step activity monitor

Introduction

Stroke is a leading cause of long-term disability in adults [1]. Ambulation difficulties contribute substantially to long-term disability and health care utilization poststroke [2]. Because of the importance of the ability to ambulate to perform routine activities, ambulatory ability is frequently included as part of the assessment of physical functioning of patients with stroke

[3]. In addition, ambulation is recommended for stroke survivors because it has a wide range of benefits that support recovery and cardiovascular health [4].

The ability to passively capture and track continuous ambulatory activity over time in real-world settings for patients with stroke has many potential uses. For example, such tracking may be able to detect relevant changes in a patient's ability to ambulate more accurately and efficiently than other common strategies, which include self-reported ambulation questionnaires or mobility performance test [5,6]. Self-reported ambulation questionnaires have poor accuracy [7-9]. Performance tests such as gait speed require measurement by a trained assessor, typically in a clinical setting, and they are resource intensive and not always feasible [5]. Another potential use of tracking ambulatory activity in real-time is to provide greater insight into the trajectories of an individual's recovery after stroke and the effectiveness of new therapies. The use of clinical information derived from real-world settings has been advocated as an avenue for comparative effectiveness research in the field of stroke and other conditions [10]. In addition, passive tracking of ambulatory activity may be able to identify slower recovery patterns of stroke patients who would benefit from targeted interventions; this use has been previously postulated in patients recovering from orthopedic surgery [11,12]. Further, monitoring ambulatory activity can also be used to encourage greater physical activity and physical fitness, which has previously been demonstrated to be effective in studies of people with diabetes [13], and in patients who have recently undergone knee and hip arthroplasty [14].

Step counts can be monitored using step activity monitors. The devices developed for research settings have been expensive and difficult to use [15]. Consumer-oriented step activity monitors are being used more broadly, as they have become more robust and comfortable to wear, with more available features. One in six (15%) consumers in the US currently use health care wearables, including smart-watches or fitness bands [16]. Earlier versions of consumer-oriented step activity monitors were pedometers consisting of mechanical sensors or uniaxial accelerometers, which measure acceleration. These had lower step count accuracy than research-grade devices; many that are currently available now contain more accurate triaxial accelerometers [17,18]. Step counts of newer step activity monitors, including several tested Fitbit devices, have excellent correlation with research-grade accelerometers [15,19,20].

Despite the theoretical potential for the use of easy-to-wear step activity monitors to improve the clinical care and outcomes of patients poststroke, the feasibility and utility of using consumer-oriented step activity monitors in stroke patients are poorly known. The consumer-oriented step activity monitors, Fitbit Charge HR and Garmin Vivosmart, were found to accurately measure step count in 37 patients attending either inpatient or outpatient therapy for stroke who wore the step activity monitors 5 to 10 hours per day for 2 days [15]. In addition to accuracy, other central aspects of the feasibility of using a step activity monitor as part of research or clinical care include adherence, patient experience, and the ability of step counts to serve as a correlate of other outcome measures [21]. Understanding the fluctuations in step counts, specifically in patients with stroke who may have different ambulatory patterns compared to other patient groups, will also be helpful to assess the feasibility of identifying changes or trends in ambulatory activity.

Therefore, we performed a prospective cohort pilot study to assess feasibility of monitoring step counts in ambulatory patients with recent mild ischemic stroke using a consumer step activity monitor, the Fitbit Charge HR. The primary study objective was to determine adherence with wearing a Fitbit Charge HR in home-going ischemic stroke patients during the first 90 days after hospital discharge. Secondary objectives were to (1) determine accuracy of step counts of the Fitbit Charge HR compared with a manual tally; (2) calculate correlations between the Fitbit step counts and the mobility performance scores at discharge and 30 days after stroke; (3) determine variability and change in weekly step counts over 90 days; and (4) evaluate patient experience with using the Fitbit Charge HR poststroke.

Methods

Procedures

Overview

All admissions to the stroke service were monitored by a trained research coordinator to identify potentially eligible patients. Prior to hospital discharge, participants were given a Fitbit Charge HR step activity monitor and instructed to wear this continuously throughout the day for 90 days. Participants were asked to complete a diary of their physical activity for the first 7 days postdischarge (See Multimedia Appendix 1: Participant Activity Diary). No recommendations were given on amount of physical activity and there were no daily step targets. Participants underwent mobility performance testing of balance, walking endurance, and gait speed shortly before hospital discharge and then again at 30 days postdischarge. They were reimbursed $30 for their time completing the performance testing at the follow-up visit, provided a $10 parking voucher, and were allowed to keep the Fitbit Charge HR for their own personal use at the end of their study participation. Participants were contacted by phone on Day 14 and Day 90 (study end) to be asked about issues with the step activity monitor and the presence of any adverse events stemming from its use. They were then sent a voluntary participant experience survey at the conclusion of the study and asked to mail back the completed survey.

Informed consent was obtained from all patient participants. The study was approved by the Cleveland Clinic Institutional Review Board.

Participants

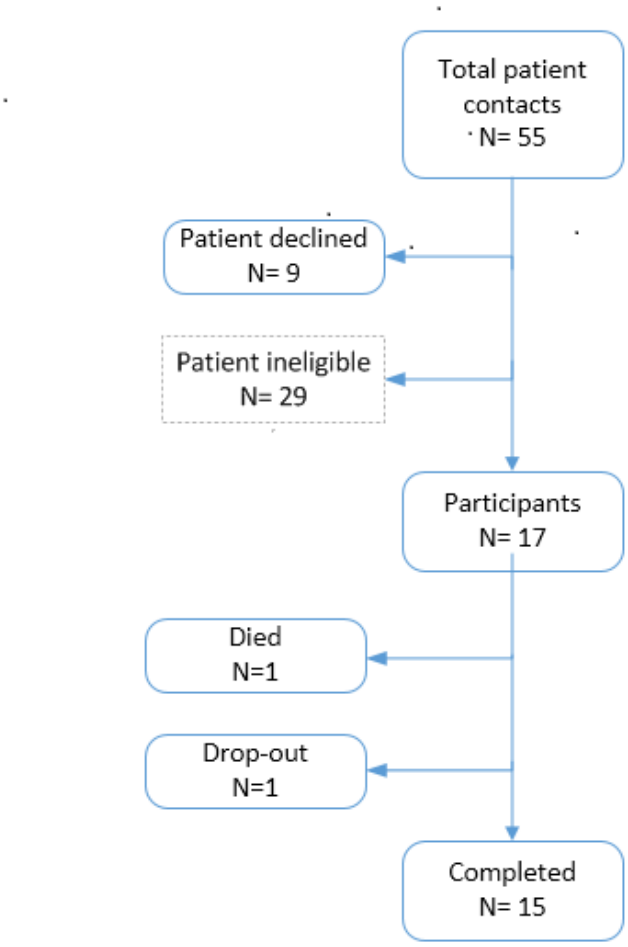
Participants were recruited through daily screening of the inpatient stroke service at Cleveland Clinic. All patients who met eligibility criteria were approached for consent. Initial inclusion criteria were the following: (1) admission to the Stroke Service with an admission diagnosis of ischemic stroke; (2) reside within Cuyahoga County or 5 surrounding counties; (3)

informed consent obtained from the patient, caregiver, or a legal representative; (4) discharged home; (5) mild stroke operationalized as mild disability at discharge as defined by a modified Rankin scale score of 1 to 2; and (6) ambulatory at the time of hospital discharge. The modified Rankin scale is a 1-item clinician-reported scale with scores ranging from 0 (representing no symptoms) to 6 (death) [22]. Patients with modified Rankin scale scores of 1 to 2 have residual symptoms but are able to ambulate without assistance from another person. Exclusion criteria were the following: (1) aged <18 years; (2) prisoners; (3) ischemic stroke due to vasculitis, moya-moya, complication from surgical procedure, or trauma; (4) non-English speaking patient with no available proxy; (5) residing in hospice or receiving palliative care prior to enrollment; (6) active recreational drug use; and (7) a medical condition that would impair the patient's ability to participate in this study in the opinion of the investigator.

Most patients discharged home after stroke from the inpatient stroke service had few neurological deficits as assessed by the clinician-reported National Institutes of Health Stroke Scale [23], a commonly used scale of severity of deficits in patients with stroke. To include participants with slightly greater deficits, the inclusion criteria were modified in the final 2 months of the recruitment period to include stroke patients discharged from inpatient rehabilitation units.

A total of 55 individuals were approached to participate in the study. Of these, 11 declined participation and 29 no longer met the inclusion criteria after initial interest in participating (Figure 1). Reasons these patients failed to meet inclusion criteria were (1) not having an internet-facing electronic device that could link to the Fitbit and upload step count data through the internet (n=10); (2) nonstroke final diagnosis (n=8); (3) poor medical compliance (n=3); (4) cognitive deficits (n=4); (5) unavailable prior to hospital discharge (after first contact) (n=2); and (6) already had Fitbit account on their personal device precluding synchronization of their device with a research account required for study participation (n=2). A total of 17 patients were enrolled in the study; one participant died due to factors unrelated to this study approximately 2 weeks postdischarge, and another participant dropped out of the study shortly after completing the assessment at hospital discharge because she felt the Fitbit Charge HR was uncomfortable on her wrist. These 2 patients were not included in the analyses.

Figure 1. Study enrollment and retention.



Step Activity Monitor Data

The Fitbit Charge HR was provided to patients at the time of enrollment. It is a wrist-worn device with a triaxial accelerometer marketed as a consumer device that displays step count and wirelessly synchronizes with smartphones. Fitbit devices have been shown to provide step counts similar to those provided by research grade accelerometers [19,20], including in patients with stroke [15]. MyInertia platform (Motion Connected LLC, Green Bay, WI) was used for administrative access to the Fitbit total daily step count tallies of study participants. The MyInertia platform did not provide a more precise timing of steps within a 24-hour period, so we were unable to determine when patients ambulated or wore the Fitbit within a 24 hour period. A MyInertia account linked to each patient's Fitbit device was activated at the beginning of the study and deactivated at the end of the study period.

Mobility Performance Tests

We administered 3 mobility performance tests that are part of the National Institutes of Health (NIH)Toolbox at discharge and at a return visit on Day 30: (1) Standing Balance Test, (2) 2-Minute Walk Endurance Test, and (3) 4-Meter Walk Gait Speed Test. The NIH Toolbox is comprised of numerous standardized royalty-free measures in neurological and behavioral health. It is designed for use in people ages 3 to 85 to allow for comparison of data across multiple studies [24]. The tests are administered through a downloaded application onto an iPad and each provides automated scoring. Balance has been shown to be independently associated with step counts [25], and the 2-Minute Walk Endurance Test and 4-Meter Walk Gait Speed Test are standard performance tests in stroke. The National Institutes of Neurological Disorders and Stroke (NINDS) Stroke Clinical Data Elements (CDE) Project considers walking speed a "highly recommended" outcome measure in patients with stroke [26].

The NIH Toolbox balance test involves the participant maintaining 5 poses for 50 seconds each. Postural sway is recorded for each pose using an accelerometer worn by the participant which synchronizes with the NIH Toolbox app. Normative scores are provided that adjust for age, sex, ethnicity, and education. The average score of the normative population is 50 (SD 10), with higher scores indicating better balance.

The 4-Meter Gait Speed Test measures the time required to walk 4 meters, which is then transformed into gait speed in meters per second. The 2-Minute Walk Endurance Test records the distance that the participant is able to walk back and forth on a 50-foot course in 2 minutes, which is converted to normative scale scores. As with balance scores, higher scores indicate better performance. To determine accuracy of the step activity monitor, we manually counted the number of steps during the 2-Minute Walk Endurance Test using a hand counter at both discharge and Day 30 assessments.

Clinician-Reported Measures

The National Institutes of Health Stroke Scale [23] is the standard scale for measuring neurological impairment. It consists of 15 items with scores ranging from 0 to 42, with higher scales indicating greater impairment. The modified Rankin scale is a 1-item measure of global disability, with scores ranging from 0 to 6 where 0 represents no symptoms [22]. These measures were part of the study inclusion criteria and were used to describe the study participants.

Participant Experience Survey (Secondary Objective 4)

Questions were adapted from those used in previously published research [27-30] with the addition of new questions specific to our study (See Multimedia Appendix 2: Participant Experience Survey).

Statistical analysis

Primary Objective: Adherence With Wearing a Step Activity Monitor

The MyInertia system used to access step count data for this study only provided daily step counts. Participant adherence was defined as having a daily step count of 100 or more. This threshold was chosen to minimize the chance of incorrectly classifying participants as nonadherent. Available data from the literature indicates that stroke survivors are sedentary [4,6]. Our study participants were recently discharged home after a hospital admission for stroke, and it was possible they would not be very physically active, especially in the first few days after discharge.

Adherence was summarized descriptively and modeled using a logistic Generalized Estimating Equation. A first-order auto-regressive (AR(1)) correlation structure was used to take into account the dependent nature of consecutive days. The intercept of the Generalized Estimating Equation model provided the population-level probability of adherence. This approach to calculating adherence allowed more accurate estimates of error around adherence rate [31].

Secondary Objectives

Secondary Objective 1: Determine Fitbit Accuracy

To determine accuracy of the Fitbit, we performed a manual count of the number of steps taken during the 2-Minute Walk Endurance Test using a hand tally counter at both discharge and Day 30 assessments. We performed an analysis of differences in the number of steps recorded on the Fitbit during 2-Minute Walk Endurance Tests and manual tally by calculating mean and median differences and mean absolute difference. Differences between device and manually counted steps were graphically displayed using Bland-Altman plots for repeated measures [32]. Accuracy of step counts was considered "acceptable" if variance between the manual tally and Fitbit step count was within 10% in either direction [33-35].

Secondary Objective 2: Calculate Correlation Between Step Counts and Mobility Performance Scores

Spearman rank correlations were used to examine the relationship between step counts and mobility performance test scores for the weeks following discharge and the Day 30 visit. The mean daily steps of the 7-day period following assessments was used to summarize step counts. Confidence intervals were estimated using the bootstrap with 10,000 repetitions. As per standard NIH Toolbox scoring procedures, these mobility performance tests were adjusted for age, sex, and education to

provide normative data for patients falling within similar age, sex, and educational attainment categories [36].

Secondary Objective 3: Determine Variability in Step Counts

Variability in steps-per-day was measured over the 90-day study period and for each 7-day period. Previous studies have found that a 7-day monitoring period is required to obtain a stable and representative average of levels of walking activity in healthy people [37] and those with neurological impairment [38]. Variability was assessed using a mixed effects model with a subject-level random intercept. Between-subjects variability was estimated using the subject-level random effect variance. Within-subject variability was indirectly assessed by calculating the intraclass correlation defined here as the proportion of total variance explained by between-subjects variability; the higher the intraclass correlation coefficient value, the greater the between-subjects variability.

Secondary Objective 4: Evaluate Participant Experience

Descriptive statistics were used to summarize participant survey responses.

Sensitivity Analyses for Definition of Daily Adherence

The suitability of the definition of daily adherence used in our analysis was evaluated using self-reported daily diaries during the first 7 days of monitoring. Daily diary entries were considered adherent if a patient reported wearing the device for more than 12 hours per day. Diary entries indicating that the Fitbit was not worn for 12 or more hours were categorized as nonadherent. Agreement between the diaries and adherence was summarized as a percentage. In a sensitivity analysis, the percentage of participants identified as "adherent" was calculated for different daily step count thresholds.

In addition, a sensitivity analysis was conducted to evaluate the effect of varying adherence thresholds on correlations between mobility performance scores and step counts. This was done by computing a series of average daily step counts over the 7-day period immediately following the performance measure assessment using only the days where the total step count was at or above specific thresholds (no step count restriction, ≥100, ≥500, ≥1000, ≥2000, ≥3000, ≥4000, and ≥5000 steps). Correlations between the mobility performance scores and each series of daily step counts was then calculated.

Sample Size Calculations

Sample size calculations were based on simulations to determine power required for the primary objective. We assumed that the percentage of adherent days could vary from 50% to 96% [37,39-41] with an average of 76%. With a sample size of 15 participants and 90 days of data from each participant, the estimated power was greater than 73% to detect a 1% or greater difference above the null hypothesis adherence rate of 75% [37,41].

Power for secondary objective 2, the correlation between mobility performance scores and step counts, was estimated using 1000 simulations of bivariate normal distributed data with correlations of 0.65. Power for a sample size of 15 was estimated to be greater than 60% to detect a correlation greater than zero.

All computations were done in R, version 3.4.1 [42]. All tests were two-sided and *P* values less than .05 were considered statistically significant.

Results

Characteristics of Participants

Participants had a mean age of 54.4 years (SD 12.1, range 34-81 years) and 60% (9/15) were male. Most had at least some college education (11/15, 73.3%). Participants were 60% (9/15) non-Hispanic White. Discharge National Institutes of Health Stroke Scale score was 0 or 1 in 73% (11/15) of participants (Table 1). Of the 15 participants, 3 (20%) indicated they used a cane at the Day 30 visit, although no ambulatory assistance devices were used during performance measure testing. The number of mean daily steps during the first week of monitoring was 4368 (SD 3968), with individual participants means ranging from 1140 to 14610. The majority of participants (12/15, 80%) could be classified as sedentary (ie, mean daily step count under 5000 [43]).

Scores for the Standing Balance Test, 2-Minute Walk Endurance Test, and 4-Meter Gait Speed Test measured at discharge and Day 30 are presented in Table 2. Fully-corrected T-scores were adjusted for age, education, race, and sex, and have a mean of 50 (SD 10) in the normative population.

Table 1. Participant characteristics (N=15).

Variable	Value
Age, years	
Mean (SD)	54.4 (12.1)
Range	34-81
NIHSS[a], n (%)	
0	6 (40)
1	5 (33.3)
3	3 (20)
4	1 (6.7)
Male, n (%)	9 (60)
White Non-Hispanic, n (%)	9 (60)
Marital status, n (%)	
Married	9 (60)
Divorced	1 (6.7)
Widowed	1 (6.7)
Single	4 (26.7)
Discharge Modified Rankin Scale, n (%)	
0	2 (13.3)
1	8 (53.3)
2	5 (33.3)
Received PT[b] and/or OT[c] post-discharge, n (%)	9 (60)
Use of ambulatory assistance device[d], n (%)	3 (20)
Education, n (%)	
High School	4 (26.7)
Some college	6 (40)
Associate's degree	3 (20)
Bachelor’s degree	1 (6.7)
Master's degree	1 (6.7)
Right-handed, n (%)	15 (100)

[a]NIHSS: National Institutes of Health Stroke Scale.

[b]PT: physical therapy.

[c]OT: occupational therapy.

[d]Used a cane at Day 30 visit.

Table 2. Mobility performance tests at Discharge and Day 30. Scores adjusted for age, sex, race/ethnicity, and education. Average T-Score is 50; higher scores indicate better performance. 2-Minute Walk Test is a measure of endurance; 4-Meter Distance Test is a measure of gait speed.

Measure	Discharge		Day 30 follow-up		Difference		
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	*P* value[a]
2-minute walk T-Score	27.94 (9.78)	14.52 to 45.16	33.53 (8.09)	19.39 to 48.75	5.60 (7.38)	–3.52 to 19.71	.01
Balance T-Score	40.75 (13.67)	24.47 to 64.26	40.55 (12.98)	22.96 to 63.68	0.64 (10.85)	–19.6 to 17.4	.83
4-meter distance T-Score	43.2 (13.46)	24.95 to 61.94	50.15 (15.23)	27.68 to 72.67	6.95 (11.86)	–16.34 to 28.34	.04
2-minute distance (m)	105.43 (40.9)	31.9 to 164.59	130.92 (33.08)	79.25 to 180.98	25.49 (29.31)	–20.42 to 83.82	.004
4-meter speed (m/sec)	0.94 (0.35)	0.2 to 1.43	1.16 (0.36)	0.67 to 1.79	0.22 (0.28)	–0.24 to 0.75	.01

[a]*P* value from paired-samples *t* test with unequal variance.

On average, participants performed worse than the general population after adjustment for education, sex, and age on all 3 tests, especially on the 2-Minute Walk test. Participant scores for both mobility tests were significantly improved at the Day 30 follow-up; balance test scores were similar at both time points.

Primary Objective: Adherence With Wearing the Step Activity Monitor

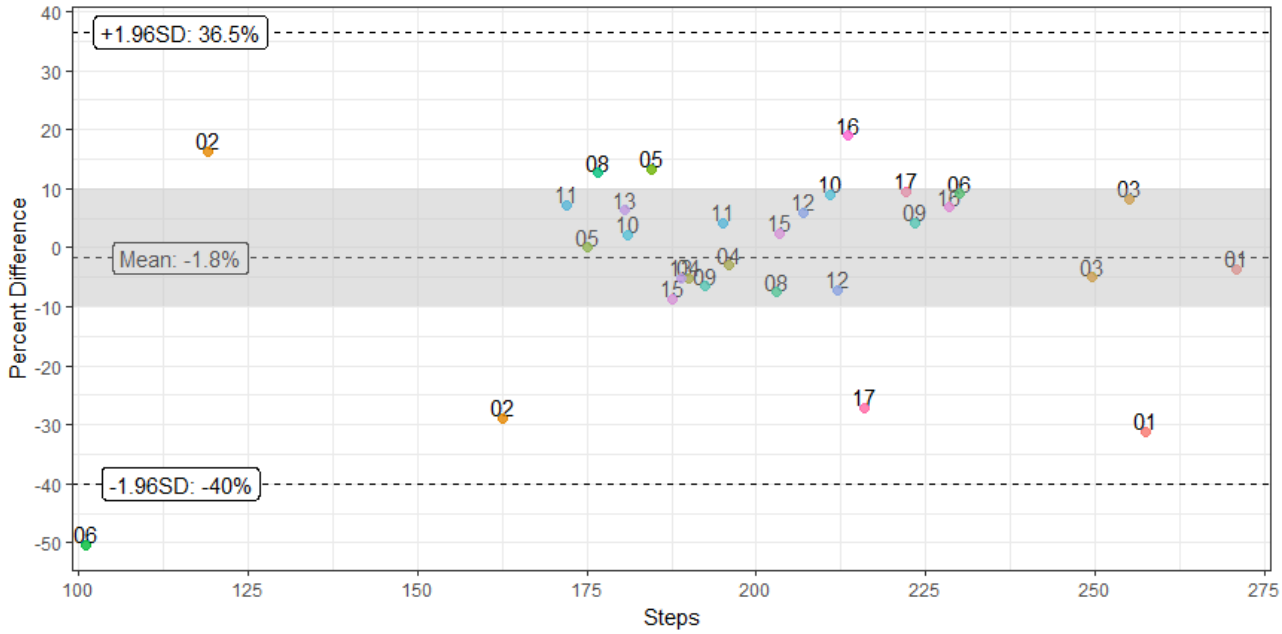
Using the adherence definition of ≥100 steps, participants were adherent 83.6% of days (approximately n=75) during the 90 day study period. An intercept-only logistic Generalized Estimating Equation using an AR(1) correlation, which adjusts for the correlation of adjacent daily observations, estimated the probability of adherence as 0.83 (95% CI 0.73-0.90). Three participants with low-adherence reported difficulties with the device: the wristband broke, step activity monitor was lost when traveling, the charger was lost. We replaced 2 of these devices. Removing these days from the calculations, probability of adherence was 0.86 (95% CI 0.78-0.92).

Secondary Aims

Secondary Objective 1 : Determine Fitbit Accuracy

While individual observation step counts recorded by the Fitbit displayed variability in comparison to manual tallies, the overall mean difference was only –4.8 steps (–1.8%). Mean absolute difference was higher, at approximately 21.7 steps (10.9%). A Bland-Altman plot showing percent difference between Fitbit and manual counts versus manual count appears in Figure 2. Observations were within an acceptable margin of error, defined as a 10% difference from manual counts, 73.3% of the time.

Figure 2. Bland Altman plot of accuracy of step activity monitor step counts compared to manual counts. Percent difference between Fitbit and manual step counts (y-axis) versus the manual step counts (x-axis). Each participant had 2 trials. The shaded region represents the "acceptable" range of error (10% in either direction), while dotted lines depict the mean percent difference and their 95% confidence intervals. Participants 2, 6, and 8 used a cane for ambulation, but no ambulatory assistance devices were used during gait testing.



Secondary Objective 2: Calculate Correlations Between Step Counts and Mobility Performance Scores

Correlations between step counts and mobility performance scores at both time points were very poor. Values for 95% confidence intervals all straddled zero. Correlations with endurance (2-Minute Walk Endurance Test v2.0), balance (Standing Balance Test v2.0), and gait (4-Meter Walk Gait Speed Test v2.0) at discharge were 0.44 (95% CI –0.14 to 0.75), –0.12 (95% CI –0.67 to 0.64), and 0.17 (95% CI –0.46 to 0.66), respectively. Correlations at Day 30 were 0.22 (95% CI –0.45 to 0.71), –0.30 (95% CI –0.83 to 0.41), and 0.21 (95% CI –0.48 to 0.76) respectively.

Secondary Objective 3: Determine Variability in Step Counts

Intraclass correlation showed moderate within-subject correlation (intraclass correlation coefficient=0.47) in step counts. Table 3 depicts intraclass correlation and between-subject variability (expressed as standard deviation) for 7 day periods over the course of the study. Values suggest moderate to large between-subject and within-subject variability. As the study progressed, between-subject variability decreased while within-subject variability increased, as shown by lower intraclass correlations.

Table 3. Variability in weekly step counts.

Week	Within-subject variability, intraclass correlation coefficient (95% CI)[a]	Between-subject variability, standard deviation (95% CI)[b]
1	0.73 (0.49-0.84)	3338.02 (1997.56-5218.09)
2	0.48 (0.18-0.66)	2367.04 (1539.35-3639.75)
3	0.39 (0.10-0.60)	2089.70 (1294.69-3372.90)
4	0.68 (0.40-0.81)	2967.29 (2011.54-4377.15)
5	0.40 (0.13-0.61)	2016.81 (1281.03-3175.19)
6	0.54 (0.26-0.70)	2504.30 (1665.16-3766.31)
7	0.34 (0.09-0.53)	1859.55 (1143.88-3022.99)
8	0.53 (0.21-0.71)	2222.68 (1435.35-3441.88)
9	0.41 (0.10-0.61)	2040.17 (1242.61-3349.64)
10	0.48 (0.18-0.66)	2252.09 (1445.39-3509.02)
11	0.45 (0.14-0.65)	2196.83 (1354.26-3563.62)
12	0.36 (0.07-0.57)	1978.23 (1185.71-3300.49)
13[c]	0.39 (0.04-0.62)	1639.84 (972.28-2765.745

[a]95% confidence interval estimated using semiparametric bootstrap with 10,000 resamples.

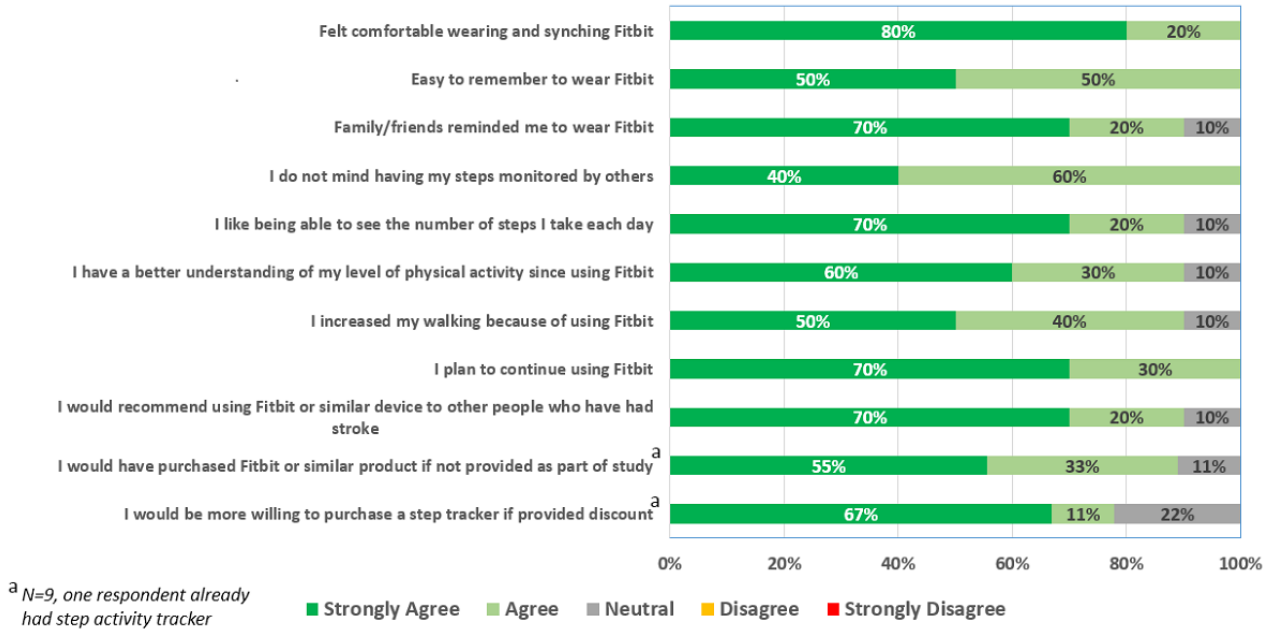
[b]95% confidence interval calculated using the Wald-type test.

[c]6-day period from day 85 to 90. Over time, between-subject variability decreased while within-subject variability increased, as shown by lower intra-class correlations.

Secondary Objective 4: Evaluate Participant Experience

Of the 15 participants, 10 (67%) completed the participant experience survey. All felt comfortable wearing the Fitbit device, thought it was easy to remember to wear, and planned to continue to use it (Figure 3). None of the respondents minded having their step count activity monitored by others. The majority (9/10, 90%) felt they had a better understanding of their physical activity levels and would recommend a step activity monitor to other people that have had a stroke.

Figure 3. Participant experience survey responses (n=10).



Sensitivity Analyses for Definition of Daily Adherence

The suitability of the definition of daily adherence used in our analysis was evaluated using self-reported daily diaries during the first 7 days of monitoring**.** Of the 15 participants, 8 (53.3%) filled out daily diaries of physical activity for the first 7-day period that they wore their Fitbit. These participants reported complete adherence (>12 hours/day Fitbit use) for that 1-week time period. During that 1-week period, step counts for these 8 participants were over 100 steps per day; thus, there was 100% agreement between the diary and step counts using definition of adherence of 100 or more steps. The minimum daily tally during the 7-day period was 346 steps. The percentage agreement for adherence decreased to 96.2% when using 1000 steps as a threshold, as 2 of the 8 participants (25%) had a single daily step count entry with less than 1000 steps.

An additional sensitivity analysis was performed to evaluate if the use of different thresholds affected the correlations between mobility performance scores and daily step counts. Altering the thresholds did not improve correlations consistently between the different mobility performance measure scores and average 7-day step count (See Multimedia Appendix 3: Spearman rank correlations between performance evaluations and different daily step count thresholds).

Discussion

Principal Findings

Our study demonstrated that wearing a consumer-grade step activity monitor after stroke is feasible in patients with recent mild acute ischemic stroke. Participants were adherent 83.6% of days over the 90-day study period, which included several instances of device malfunction or misplacement. We defined adherence a priori as ≥100 steps per day, although varying the step-counts thresholds had little effect either on the percentage of days participants were labeled adherent or with the correlation between step counts and participant performance on mobility tests. Studies of adherence with step activity monitors are sparse, and the differences in definition of adherence, populations studied, and duration of monitoring make comparisons across studies difficult. A recent clinical trial of step activity monitor use in patients with rheumatoid arthritis [44], which monitored step counts over a 21-week period, demonstrated an adherence rate of 88.8% per study day, similar to our study. That study defined adherence as the proportion of study days that steps were recorded.

The accuracy of step activity monitors is a critical consideration when choosing a device, especially in patients with neurological impairment. This study suggests that the Fitbit Charge HR provides reasonably accurate step counts in patients with mild stroke who have similar gait speeds to those in our study. While observed step counts of individuals displayed high variability in accuracy, overall mean error was only –4.8 steps (–2.4%). The mean gait speed of participants in our study was 0.94 m/sec and the accuracy of the Fitbit Charge HR in patients with slower gait speeds than observed in our study is unclear [45].

Approximately 58% (15/26) of eligible patients completed the study. Almost 20% of patients approached to participate (10/55) were ineligible because they did not have an internet-facing electronic device that could link to the Fitbit, an important factor to consider when planning future studies of step counts in stroke. This rate is likely to decline with time but could be a significant factor limiting recruitment over the next few years. Acceptability of wearing the Fitbit was quite high. The majority of participants who responded to the survey indicated they would continue to use the devices to monitor their physical activity after the study ended. The respondents also indicated they would not mind having this information be available to providers.

There are several advantages to using consumer-grade step activity monitors to monitor ambulation in patients with stroke and other conditions. Designed for the consumer, they may be

more intuitive for patients to use, provide user-friendly displays, and often collect other data such as sleep time and heart rate. These factors may have contributed to the high adherence rates and acceptability of the device in our study. Healthcare wearables, such as step activity monitors, have been predicted to grow from 1 million purchases in 2015 to 97.6 million by 2021 [46]. More stroke patients will already own and wear their own step activity monitor in the coming years, another significant advantage. In addition, several electronic health record companies are developing conduits to pull in patient-generated heath data from commonly used consumer devices making them easier to incorporate into providers' clinical care workflows.

Although this study did not directly assess the clinical utility of step activity monitors for patients with mild stroke, the results provide some useful insights for the design of future studies that address this question. First, most of the participants in our study were sedentary, walking less than 5000 steps per day. Our findings are consistent with others who found that many stroke survivors have low levels of physical activity [4,6] and highlights the merit of research regarding the effectiveness of step targets in increasing the ambulation of persons with stroke. Second, there was wide variability in weekly step counts both within and across study participants, which has implications for determining sample sizes of future studies that aim to identify trends. A significant trade-off between sensitivity and specificity to detect change may be required when using step activity monitors in stroke survivors.

A notable and somewhat unexpected finding in our study was the dismal correlation between step counts and performance measures for gait speed, endurance, and balance. The endurance and gait speed tests are standard performance tests in stroke [26] and, along with the balance test, have been shown to be independently associated with step counts [5,6,25,47]. Testing was performed using tools in the NIH Toolbox, which are validated and recommended for use in clinical trials to allow for standard assessment. The low correlations highlight important distinctions between step counts and mobility performance measures that, in light of these findings, may be especially pertinent in patients recovering from mild stroke. Performance measures assess what a patient is capable of doing in ideal conditions, whereas step count measures activity in real-world conditions. Patients' habits, social environment, and attitudes towards engaging in physical activities affect the number of steps taken apart from their physical abilities. Low mood, for example, can contribute to lower physical activity levels [6]. It is possible that there is a stronger association between step counts and patient-reported outcomes, which are also impacted by patient environment and attitudes. It will be important to assess the relationship between step counts in stroke patients and patient-reported and other outcome measures to provide clinical relevance and interpretation of step count values.

Strengths and Limitations

An important strength of our study included the 90-day monitoring period. A 7-day protocol for step activity monitors is commonly used in research protocols [48-51]. Assessment of adherence over 90 days provides information on the feasibility of using step activity monitors during the initial 3 month period after stroke (the timeframe in which the most rapid recovery occurs [52]) and for studies of behavioral intervention that may last months rather than weeks. Another strength is the collection of performance measures across 2 time points, which allowed assessment of whether correlations between step count and mobility performance tests varied according to the time since stroke.

This study also has several limitations. Patient participants had overall mild degree of disability and findings from our study may not be generalizable to stroke survivors with more severe deficits. Our initial target population consisted of patients discharged home after stroke admission, since ambulation in the acute rehabilitation setting can be constricted [53]. We modified the protocol to include patients discharged from acute rehabilitation to assess step counts in patients with greater deficits. The feasibility study included 15 participants and findings would be more compelling with a larger sample size. In addition, only two-thirds (10/15) of the participants completed the participant experience survey. Another important limitation is that we were unable to evaluate within-day variability, intensity, and patterns of step counts. A recent study of patients with Parkinson's disease [54] found that intensity of physical activity, but not step counts, declined over time, suggesting that intensity or type of activity could be more sensitive to detect change in physical function. The lack of detailed step count activity within a 24-hour period also constrained our assessment of adherence. However, benefits of using daily step counts include their measurement simplicity and higher likelihood of being available within a patient's electronic health record. Identifying when a participant is not wearing the step activity monitor is difficult even when detailed step activity data is available within a 24-hour period [55]. An important avenue of further research is to determine the optimal definition of adherence when only daily step count totals are available. One approach may be individualization of step count threshold for each person based on their baseline step counts and activity patterns.

There are also drawbacks with using consumer-grade devices in general. The algorithms used to measure steps and other metrics are typically proprietary and may not be available to investigators [56]. Consumer-grade devices such as the Fitbit Charge HR are often worn on the wrist, and these are not suitable for monitoring step counts in patients using rollator ambulation aids [15].

Conclusions

The use of patient-generated health data in clinical care and research is part of an evolving paradigm shift in health care. Wearable technologies such as step activity monitors offer an excellent mechanism for gathering data from patients effectively, continuously, and in real time. This evaluation represents one of the first applications of patient-generated health data from consumer wearable devices in patients with stroke. The results of this study suggest that the use of the consumer-grade Fitbit Charge HR in patients with recent mild stroke is feasible with reasonable adherence, accuracy, and high level of acceptability. There was poor correlation between step counts and gait speed,

balance, and endurance as assessed using the mobility performance tests in the NIH Toolbox. Further research is needed to evaluate the association between step counts and other outcomes of relevance to patients, including patient-reported outcomes and measures of physical function. This feasibility study will serve as important groundwork for further studies of the use of step activity monitors to optimize the care and outcomes of patients with stroke.

Acknowledgments

This study was funded in part by the Cleveland Clinic Research Program Committee. IK led the study concept and design, interpreted the data, and drafted the manuscript. AS led the data compilation and collection, interpreted the data, and critically revised the manuscript for intellectual content. TK performed the data analysis, interpreted the data, and critically revised the manuscript for intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant activity diary.
[DOCX File , 23 KB - mhealth_v9i1e14494_app1.docx]

Multimedia Appendix 2
Patient experience survey.
[DOCX File , 21 KB - mhealth_v9i1e14494_app2.docx]

Multimedia Appendix 3
Spearman rank correlations between performance evaluations and different daily step count thresholds.
[DOCX File , 19 KB - mhealth_v9i1e14494_app3.docx]

References

1. Virani S, Alonso A, Benjamin E, Bittencourt M, Callaway C, Carson A, et al. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020;141:e139-e596.
2. Rundek T, Mast H, Hartmann A, Boden-Albala B, Lennihan L, Lin I, et al. Predictors of resource use after acute hospitalization: the Northern Manhattan Stroke Study. Neurology 2000 Oct 24;55(8):1180-1187. [doi: 10.1212/wnl.55.8.1180] [Medline: 11071497]
3. Lord SE, McPherson K, McNaughton HK, Rochester L, Weatherall M. Community ambulation after stroke: how important and obtainable is it and what measures appear predictive? Arch Phys Med Rehabil 2004 Feb;85(2):234-239. [doi: 10.1016/j.apmr.2003.05.002] [Medline: 14966707]
4. Billinger SA, Arena R, Bernhardt J, Eng JJ, Franklin BA, Johnson CM, et al. Physical Activity and Exercise Recommendations for Stroke Survivors. Stroke 2014 Aug;45(8):2532-2553. [doi: 10.1161/str.0000000000000022]
5. Fulk GD, He Y, Boyne P, Dunning K. Predicting Home and Community Walking Activity Poststroke. Stroke 2017 Feb;48(2):406-411. [doi: 10.1161/strokeaha.116.015309]
6. Field MJ, Gebruers N, Shanmuga Sundaram T, Nicholson S, Mead G. Physical Activity after Stroke: A Systematic Review and Meta-Analysis. Int Sch Res Notices 2013 Nov 07;2013:1-13 Article ID 464176. [doi: 10.1155/2013/464176]
7. Vaughn NH, Dunklebarger MF, Mason MW. Individual Patient-reported Activity Levels Before and After Joint Arthroplasty Are Neither Accurate nor Reproducible. Clinical Orthopaedics and Related Research 2019;477(3):536-544. [doi: 10.1097/corr.0000000000000591]
8. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty?: A systematic review. Clin Orthop Relat Res 2010 Jul 28;468(7):1891-1904 [FREE Full text] [doi: 10.1007/s11999-009-1135-9] [Medline: 19862586]
9. Verlaan L, Bolink S, Van Laarhoven S, Lipperts M, Heyligers I, Grimm B, et al. Accelerometer-based Physical Activity Monitoring in Patients with Knee Osteoarthritis: Objective and Ambulatory Assessment of Actual Physical Activity During Daily Life Circumstances. Open Biomed Eng J 2015 Jul 31;9(1):157-163 [FREE Full text] [doi: 10.2174/1874120701509010157] [Medline: 26312077]
10. Horn SD, Gassaway J. Practice Based Evidence. Medical Care 2010;48:S17-S22. [doi: 10.1097/mlr.0b013e3181d57473]
11. Toogood PA, Abdel MP, Spear JA, Cook SM, Cook DJ, Taunton MJ. The monitoring of activity at home after total hip arthroplasty. The Bone & Joint Journal 2016 Nov;98-B(11):1450-1454. [doi: 10.1302/0301-620x.98b11.bjj-2016-0194.r1]
12. Schmal H, Holsgaard-Larsen A, Izadpanah K, Brønd JC, Madsen CF, Lauritsen J. Validation of Activity Tracking Procedures in Elderly Patients after Operative Treatment of Proximal Femur Fractures. Rehabil Res Pract 2018 Jun 19;2018:3521271-3521279 [FREE Full text] [doi: 10.1155/2018/3521271] [Medline: 30018824]

13. Baskerville R, Ricci-Cabello I, Roberts N, Farmer A. Impact of accelerometer and pedometer use on physical activity and glycaemic control in people with Type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2017 May;34(5):612-620. [doi: 10.1111/dme.13331] [Medline: 28173623]
14. Van der Walt N, Salmon LJ, Gooden B, Lyons MC, O'Sullivan M, Martina K, et al. Feedback From Activity Trackers Improves Daily Step Count After Knee and Hip Arthroplasty: A Randomized Controlled Trial. J Arthroplasty 2018 Nov;33(11):3422-3428. [doi: 10.1016/j.arth.2018.06.024] [Medline: 30017217]
15. Rozanski GM, Aqui A, Sivakumaran S, Mansfield A. Consumer Wearable Devices for Activity Monitoring Among Individuals After a Stroke: A Prospective Comparison. JMIR Cardio 2018 Jan 04;2(1):e1 [FREE Full text] [doi: 10.2196/cardio.8199] [Medline: 31758760]
16. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. PLoS Med 2016 Feb 2;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
17. Gebruers N, Vanroy C, Truijen S, Engelborghs S, De Deyn PP. Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures. Arch Phys Med Rehabil 2010 Feb;91(2):288-297. [doi: 10.1016/j.apmr.2009.10.025] [Medline: 20159136]
18. Rand D, Eng JJ, Tang P, Jeng J, Hung C. How Active Are People With Stroke? Stroke 2009 Jan;40(1):163-168. [doi: 10.1161/strokeaha.108.523621]
19. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12(1):42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
20. Fulk G, Combs S, Danks K, Nirider C, Raja B, Reisman DS. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. Phys Ther 2014 Feb;94(2):222-229. [doi: 10.2522/ptj.20120525] [Medline: 24052577]
21. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
22. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1988 May;19(5):604-607. [doi: 10.1161/01.str.19.5.604] [Medline: 3363593]
23. Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. Stroke 1989 Jul;20(7):864-870. [doi: 10.1161/01.str.20.7.864] [Medline: 2749846]
24. Hodes RJ, Insel TR, Landis SC. The NIH Toolbox: Setting a standard for biomedical research. Neurology 2013 Mar 11;80(Issue 11, Supplement 3):S1-S1. [doi: 10.1212/wnl.0b013e3182872e90]
25. Michael KM, Allen JK, Macko RF. Reduced ambulatory activity after stroke: the role of balance, gait, and cardiovascular fitness. Arch Phys Med Rehabil 2005 Aug;86(8):1552-1556. [doi: 10.1016/j.apmr.2004.12.026] [Medline: 16084807]
26. Stroke v2. National Institute of Neurological Disorders and Stroke: NINDS Common Data Elements. URL: https://www.commondataelements.ninds.nih.gov/Stroke [accessed 2020-04-20]
27. Wakefield BJ, Holman JE, Ray A, Scherubel M. Patient perceptions of a remote monitoring intervention for chronic disease management. J Gerontol Nurs 2011 Apr 16;37(4):16-20. [doi: 10.3928/00989134-20110302-05] [Medline: 21417194]
28. Britton BP. First home telehealth clinical guidelines developed by the American Telemedicine Association. Home Healthcare Nurse 2003 Oct;21(10):703-706. [doi: 10.1097/00004045-200310000-00014] [Medline: 14557713]
29. Finkelstein SM, MacMahon K, Lindgren BR, Robiner WN, Lindquist R, VanWormer A, et al. Development of a remote monitoring satisfaction survey and its use in a clinical trial with lung transplant recipients. J Telemed Telecare 2012 Jan 08;18(1):42-46 [FREE Full text] [doi: 10.1258/jtt.2011.110413] [Medline: 22067286]
30. Casey M, Hayes PS, Glynn F, ÓLaighin G, Heaney D, Murphy AW, et al. Patients' experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care. Br J Gen Pract 2014 Jul 28;64(625):e500-e508. [doi: 10.3399/bjgp14x680989]
31. Huh D, Flaherty BP, Simoni JM. Optimizing the analysis of adherence interventions using logistic generalized estimating equations. AIDS Behav 2012 Feb 8;16(2):422-431 [FREE Full text] [doi: 10.1007/s10461-011-9955-5] [Medline: 21553253]
32. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007 Jul 05;17(4):571-582. [doi: 10.1080/10543400701329422] [Medline: 17613642]
33. Crouter SE, Schneider PL, Karabulut M, Bassett DR. Validity of 10 Electronic Pedometers for Measuring Steps, Distance, and Energy Cost. Medicine & Science in Sports & Exercise 2003;35(8):1455-1460. [doi: 10.1249/01.mss.0000078932.61440.a2]
34. Schneider PL, Crouter SE, Lukajic O, Bassett DR. Accuracy and Reliability of 10 Pedometers for Measuring Steps over a 400-m Walk. Medicine & Science in Sports & Exercise 2003;35(10):1779-1784. [doi: 10.1249/01.mss.0000089342.96098.c4]
35. Schneider PL, Crouter SE, Bassett DR. Pedometer Measures of Free-Living Physical Activity: Comparison of 13 Models. Medicine & Science in Sports & Exercise 2004;36(2):331-335. [doi: 10.1249/01.mss.0000113486.60548.e9]
36. Beaumont JL, Havlik R, Cook KF, Hays RD, Wallner-Allen K, Korper SP, et al. Norming plans for the NIH Toolbox. Neurology 2013 Mar 11;80(Issue 11, Supplement 3):S87-S92. [doi: 10.1212/wnl.0b013e3182872e70]

37. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc 2002 Aug;34(8):1376-1381. [doi: 10.1097/00005768-200208000-00021] [Medline: 12165695]
38. Hale LA, Pal J, Becker I. Measuring free-living physical activity in adults with and without neurologic dysfunction with a triaxial accelerometer. Arch Phys Med Rehabil 2008 Sep;89(9):1765-1771. [doi: 10.1016/j.apmr.2008.02.027] [Medline: 18760161]
39. Van Coevering P, Harnack L, Schmitz K, Fulton JE, Galuska DA, Gao S. Feasibility of using accelerometers to measure physical activity in young adolescents. Med Sci Sports Exerc 2005 May;37(5):867-871. [doi: 10.1249/01.mss.0000162694.66799.fe] [Medline: 15870643]
40. Barak S, Wu SS, Dai Y, Duncan PW, Behrman AL, Locomotor Experience Applied Post-Stroke (LEAPS) Investigative Team. Adherence to accelerometry measurement of community ambulation poststroke. Phys Ther 2014 Jan;94(1):101-110 [FREE Full text] [doi: 10.2522/ptj.20120473] [Medline: 24029299]
41. Cooper A, Page A, Fox K, Misson J. Physical activity patterns in normal, overweight and obese individuals using minute-by-minute accelerometry. Eur J Clin Nutr 2000 Dec 27;54(12):887-894. [doi: 10.1038/sj.ejcn.1601116] [Medline: 11114687]
42. R Core Team. A language and environment for statistical computing. In: R Foundation for Statistical Computing. Vienna, Austria: www.R-project.org; 2017.
43. Tudor-Locke C, Craig C, Thyfault J, Spence JC. Appl Physiol Nutr Metab 2013;38(2):100-114. [doi: 10.1139/apnm-2012-0235]
44. Katz P, Margaretten M, Gregorich S, Trupin L. Physical Activity to Reduce Fatigue in Rheumatoid Arthritis: A Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018 Jan 06;70(1):1-10. [doi: 10.1002/acr.23230] [Medline: 28378441]
45. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014 Jun;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006] [Medline: 24656871]
46. Wearable Devices for Healthcare Markets - Executive Summary. Tractica LLC. 2016. URL: www.tractica.com/wp-content/uploads/2016/04/HCW-16-Executive-Summary.pdf [accessed 2020-04-20]
47. Vahlberg B, Bring A, Hellström K, Zetterberg L. Level of physical activity in men and women with chronic stroke. Physiother Theory Pract 2019 Oct 16;35(10):947-955. [doi: 10.1080/09593985.2018.1460646] [Medline: 29659314]
48. Tinlin L, Fini N, Bernhardt J, Lewis LK, Olds T, English C. Best practice guidelines for the measurement of physical activity levels in stroke survivors. International Journal of Rehabilitation Research 2018;41(1):14-19. [doi: 10.1097/mrr.0000000000000253]
49. Tudor-Locke C, Burkett L, Reis JP, Ainsworth BE, Macera CA, Wilson DK. How many days of pedometer monitoring predict weekly physical activity in adults? Prev Med 2005 Mar;40(3):293-298. [doi: 10.1016/j.ypmed.2004.06.003] [Medline: 15533542]
50. Rowe D, Kemble CD, Robinson TS, Mahar MT. Daily walking in older adults: day-to-day variability and criterion-referenced validity of total daily step counts. J Phys Act Health 2007;4(4):434-446. [doi: 10.1123/jpah.4.4.435]
51. Nicolai S, Benzinger P, Skelton DA, Aminian K, Becker C, Lindemann U. Day-to-day variability of physical activity of older adults living in the community. J Aging Phys Act 2010 Jan;18(1):75-86. [doi: 10.1123/japa.18.1.75] [Medline: 20181995]
52. Lee KB, Lim SH, Kim KH, Kim KJ, Kim YR, Chang WN, et al. Six-month functional recovery of stroke patients. International Journal of Rehabilitation Research 2015;38(2):173-180. [doi: 10.1097/mrr.0000000000000108]
53. Dorsch AK, Thomas S, Xu X, Kaiser W, Dobkin BH, SIRRACT investigators. SIRRACT: An International Randomized Clinical Trial of Activity Feedback During Inpatient Stroke Rehabilitation Enabled by Wireless Sensing. Neurorehabil Neural Repair 2015 Jun 26;29(5):407-415 [FREE Full text] [doi: 10.1177/1545968314550369] [Medline: 25261154]
54. Cavanaugh JT, Ellis TD, Earhart GM, Ford MP, Foreman KB, Dibble LE. Capturing ambulatory activity decline in Parkinson's disease. J Neurol Phys Ther 2012 Jun;36(2):51-57 [FREE Full text] [doi: 10.1097/NPT.0b013e318254ba7a] [Medline: 22592060]
55. Mâsse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S544-S554. [doi: 10.1249/01.mss.0000185674.09066.8a] [Medline: 16294117]
56. Lee S. Can Fitbits be trusted in science? buzzfeednews. 2015. URL: www.buzzfeednews.com/article/stephaniemlee/can-fitbits-be-trusted-in-science [accessed 2020-04-20]

Abbreviations

NIH: National Institutes of Health

Edited by G Eysenbach; submitted 25.04.19; peer-reviewed by M Bass, A Joeris, K Tamura, J Bussmann, F Lanfranchi, J Farzi; comments to author 14.12.19; revised version received 01.05.20; accepted 13.05.20; published 19.01.21.

Please cite as:
Katzan I, Schuster A, Kinzy T
Physical Activity Monitoring Using a Fitbit Device in Ischemic Stroke Patients: Prospective Cohort Feasibility Study
JMIR Mhealth Uhealth 2021;9(1):e14494
URL: http://mhealth.jmir.org/2021/1/e14494/
doi:10.2196/14494
PMID:33464213

Original Paper

Habit Formation in Wearable Activity Tracker Use Among Older Adults: Qualitative Study

Wei Peng[1], PhD; Lin Li[1], MA; Anastasia Kononova[1], PhD; Shelia Cotten[2], PhD; Kendra Kamp[3], PhD; Marie Bowen[4], MSED

[1]College of Communication Arts and Sciences, Michigan State University, East Lansing, MI, United States

[2]Clemson University, Clemson, SC, United States

[3]University of Washington, Seattle, WA, United States

[4]State of Michigan, Lansing, MI, United States

Corresponding Author:
Wei Peng, PhD
College of Communication Arts and Sciences
Michigan State University
404 Wilson Room 409
East Lansing, MI, 48824
United States
Phone: 1 5174328235
Email: pengwei@msu.edu

Abstract

Background: Wearable activity trackers are popular devices used to motivate behavior change. Wearable activity trackers are especially beneficial for encouraging light physical activity such as walking, which is an ideal behavior for older adults or individuals who cannot be physically active at moderate and vigorous levels. A common problem is that people do not continue to use these wearable devices, with initial behavioral change gains eroding as people disengage. Limited research is available regarding the continued use of wearable activity trackers. The habit formation literature may provide insights into the long-term use of wearables and other health informatics devices.

Objective: This study aims to uncover the mechanism underlying the long-term continued use of wearable devices among older adults through the theoretical lens of habit formation.

Methods: In-depth interviews were conducted with 20 participants who were aged 65 years or older and had used wearable activity trackers for more than 6 months to understand their experiences and the strategies they employed to support continued use.

Results: Thematic analysis of data revealed 8 themes related to habit formation, including aspects in initiation and goal setting, use of contextual cues, action planning, and coping planning. Long-term users tended to have meaningful initiation of wearable activity trackers. They usually started with a small behavioral change goal and gradually increased it. They used consistent time and locational cues to make the use of wearable activity trackers routine. Long-term users also used creative contextual cues and reminders to facilitate action planning, engaged in coping planning to deal with anticipated problems, and had a positive mindset and inventive strategies for managing unfulfillment and lapses.

Conclusions: The results of this qualitative study of long-term users of wearable activity trackers suggest specific ways to enhance long-term habit formation among older adults. These best practices by long-term users can inform the future design of technology-based behavior interventions.

(JMIR Mhealth Uhealth 2021;9(1):e22488) doi:10.2196/22488

KEYWORDS

habits; action planning; coping planning; activity trackers; fitness trackers; continued use; mobile phone; older adults; health behavior; mHealth

Introduction

Long-term Use of Health-Related Devices

A multitude of devices and gadgets, from the early pedometer to the wearable activity tracker, have been developed over the past few decades to help individuals change from a sedentary existence to a more active and healthier lifestyle. Wearable activity trackers are sensor-enabled devices used for monitoring physical activity, sleep, and other fitness and health-related states to facilitate behavioral change and improve health [1]. Examples of wearable activity trackers include commercial products such as Fitbit, Garmin, or Apple Watch. Studies have shown that using these devices can be effective in motivating and facilitating physical activity increases [2-6]. Wearable activity trackers have the potential to facilitate and sustain behavior change to help individuals be more physically active [5-9]. If individuals have continued engagement with wearable activity trackers, it is more likely that these devices can positively affect behavior change.

However, the problem is that many people start using the devices but abandon them quickly. For instance, wearable activity trackers have not lived up to expectations, with most people discontinuing their use within 6 months of starting [10-12]. Most of the empirical research focuses on identifying reasons why people did *not* use or continue to use personal informatics devices, such as forgetting to put it on, inconsistency between users' anticipations and the real functionalities of the device, mismatch between devices and the user's self-concept, difficulties in managing the device over time, reaching saturation in knowledge from the monitoring, or achieving behavior change goals and thus losing the sense of benefit from wearable activity trackers [13-19].

Limited research is available regarding continued use of these devices [20-23]. Without the continued use of wearable activity trackers, their benefits cannot be fully achieved. Therefore, understanding why individuals continue to use wearable activity trackers and what strategies they employ to do so is critical for wearable activity trackers to reach their full potential for health promotion.

Habit Formation and Maintenance

The habit formation literature provides insights into the long-term use of wearable activity trackers and other health informatics devices [24-27]. Habit is defined as "behavioral patterns enacted automatically in response to a situation in which the behavior has been performed repeatedly and consistently in the past [28]," following what is commonly accepted in the field of psychology. Once a habit is formed, individuals engage in action *automatically* without requiring much cognitive effort [29]. People do not need to *remember* to engage in the action; they do so out of *second nature* [30]. A meta-analysis revealed a moderate to strong effect size (random effect r=0.46) for the correlation between the behaviors of healthy eating and physical activity and the self-reported habit index [31]. A 12-week randomized control trial found that habit formation was significantly related to taking more steps per day and adherence to the intervention [27]. The critical issue is to understand the process of habit formation and maintenance. Therefore, the aim of this study is to investigate *how* long-term users of wearable activity trackers develop the habit of using wearable activity trackers to manage their physical activity.

Habits are acquired through the repetition of behavior in a consistent context [32,33]. The habit formation process involves 3 progressing phases: initiation, learning, and stability. In the initiation phase, decisions are made to perform a behavior and the context in which it will be performed is selected. Instilling sufficient motivation, setting small and realistic goals, and choosing a simple action that will help individuals move toward the goal are essential in this phase [34]. Before actions are repeated in a consistent context, intention to act, action planning, and coping planning are needed to prepare the initiation of habit formation. Action planning involves detailed specification of the features of situations and intended responses to the situation. One example of action planning is the *if-then* rule; if situation X happens, then one will do Y [35]. Action planning helps to establish a link between contextual cues and behaviors. Reminders can help people remember their plans [36]. Another type of planning, coping planning, deals with anticipating difficulties that might hinder the action and creating concrete plans to overcome potential obstacles [37].

The second phase is the learning phase in which the behavior is repeated to strengthen the context-behavior association. The critical factor for habit formation is the consistent context cue: repetition must occur in association with the same event triggers, similar time and location, or similar people or objects. For instance, people usually form a habit of closing their garage door when they drive out of the garage. Through numerous repetitions of pushing the button to close the garage door at the same event of driving out of the garage (ie, a contextual cue), people do not need to think of pushing the button deliberately. Sometimes, people cannot remember whether they have closed the garage door or not and just return to find out that the garage door is indeed closed—they did it out of habit automatically.

When a new action is performed, a mental association between the situation (contextual cues) and the action is created. Through repetition, the association of contextual cues and behaviors strengthens; dependence on conscious attention or motivational process to perform an action is reduced, and only the contextual cues are sufficient for the action. Behavior achievement promotes continued repetition [28].

The third phase of habit formation is the stability phase in which the behavior persists over time with minimal effort and deliberation [34]. At this stage, the habit has formed, and its strength has peaked and plateaued. Once the habit is formed, even after conscious motivation or interest fades away, the action or behavior may still be performed when contextual cues are present. The habit concept yields potential for health-related behavior change, which is typically a long-term process characterized by the initiation of new behavior and maintenance (ie, repetition) over time [28].

In this study, we focus on the first 2 phases, as initiation and learning phases bring most insights to individuals whose habit has not formed. We propose to study how long-term *power users* developed the habit of using a wearable activity tracker

regularly to manage physical activity via the following research questions (RQ):

- RQ1: How did long-term users initiate the use of wearable activity trackers?
- RQ2: How did long-term users set goals in the context of tracker use?
- RQ3: What consistent contextual cues did long-term users use to form the habit of wearing wearable activity trackers?
- RQ4: How did long-term users engage in action planning and coping planning to support the continued use of wearable activity trackers?

Methods

Overview

We adopted a qualitative approach in this study because limited evidence exists to examine the habit formation process and specific strategies for continued use, particularly for older adults. Our qualitative approach with long-term users has the benefit of gaining an in-depth understanding of the habit of continued technology use and behavior maintenance. This study was approved by the institutional review board of Michigan State University (IRB# x16-1583eD).

Participants

This study is situated in the context of wearable activity tracker use among older adults. In-depth interviews were conducted with 20 older adults (aged 65 years and above) who had used wearable activity trackers for more than 6 months. Older adults were selected for this study for several reasons. First, previous research has found that wearable activity trackers are beneficial for enhancing physical activity among older adults [3,38,39]. Wearable activity trackers typically track step counts, and walking is a significant contributor to physical activity in older adults [40]. Increasing physical activity in older adults can help stave off chronic disease development and progression. The very act of wearing a wearable activity tracker that monitors and encourages movement may remind older adults to be more proactive in increasing their physical activity. This study focused on long-term *power users* or individuals who have used behavioral modification technology over 6 months. Previous research on technology-based behavioral interventions primarily investigated efficacy and barriers while missing the opportunity to identify successful strategies employed by *power users*. These best practices by long-term users can inform the future design of technology-based behavior interventions.

Long-term wearable activity tracker users were recruited in January 2017 by distributing a survey to a Qualtrics panel of older adults (aged 65 years and above) who had an experience with wearable activity trackers. A total of 314 respondents completed the survey, of which 163 indicated that they had used wearable activity trackers for more than 6 months. Among the 163 long-term *power users*, 71 stated that they were interested in our follow-up interview study exploring their experiences with wearable activity trackers. Of the 71 participants, 20 were randomly chosen for the in-depth phone interview, which took place from February to April 2017. Among the 20 participants, the average age was 67.95 years (SD 2.01) and 55% (11/20) were female. The average length of the wearable activity tracker used in the sample was 31.9 months (SD 25.64), with 15 reporting that they used a Fitbit. Other devices reported by the participants included Garmin, Miband, and a cell phone. Detailed demographic information of the interviewees is listed in Table 1.

Table 1. Participant demographic information (n=20).

Participant ID	Gender	Age (years)	Race or ethnicity	State
1	Female	67	White	Minnesota
2	Female	66	White	Oregon
3	Male	67	Asian	Hawaii
4	Female	68	White	Michigan
5	Male	65	White	Pennsylvania
6	Male	66	White	Alabama
7	Female	68	White	Arkansas
8	Male	70	White	North Carolina
9	Male	68	White	New Jersey
10	Female	67	White	Ohio
11	Female	71	White	Colorado
12	Female	70	White	Georgia
13	Male	66	Hispanic or Latino	Texas
14	Female	69	White	California
15	Female	72	White	Tennessee
16	Female	66	White	Texas
17	Male	69	White	Indiana
18	Male	65	White	Maryland
19	Female	70	White	Wisconsin
20	Male	68	White	Nevada

Procedure

The 20 randomly selected participants were contacted via telephone by a researcher to confirm the availability and length of the wearable activity tracker use. We sent an informed consent document a week in advance and then read an abbreviated version over the phone and obtained verbal consent. At the end of the first interview, the participants were asked to use their smartphones to take photographs that provided insights into the use of wearable activity trackers in their lives. As the participants were long-term users of wearable activity trackers and their habit of use had already developed, these photos helped the participants recall and become more conscious of certain automatic behaviors for the interview discussion. The participants then texted or emailed the photos with annotation to the research team. The photo-taking and sharing exercise lasted for a week, and then a follow-up interview was conducted. After completing the first interview, participants received a US $30 Target, Walmart, or Amazon gift card via mail or email. After completing the photo-taking exercise and the second interview, the participants received a US $50 gift card to the store of their choice. This study reports the elements in the interviews that pertain to habit formation for the continued use of wearable activity trackers.

Interview Materials

Semistructured interviews were conducted. The interview guide was developed based on the habit formation literature [32-34,41]. For the first interview, researchers began with the consent process. Once individuals agreed to participate, researchers started with ice-breaker questions. Then the participants were asked questions related to general use, habit formation, and data usage of the wearable activity tracker. Example questions are included in Table 2. At the end of the interview, researchers explained that they would participate in a photo-taking task for about a week to capture what, where, when, and why they used their trackers and strategies for long-term use of the trackers. The second interview included questions about the photos taken by each participant, the context of tracker use, benefits of wearable activity trackers, and successful strategies of long-term use. Table 2 summarizes the interview guide topics and provides examples of questions.

Table 2. Interview guide and example questions.

Interview	Example questions
First interview	
Interview topics	
Icebreaker	Can you describe how you use your tracker on a typical day?
General use	• What features do you use the most and least and why? • Did you set a goal for yourself when using the wearable activity tracker? • How do you feel when you don't reach your goals?
Habit formation	• Tell me the story of when you started using your tracker. What made you decide to start using one? • Why have you continued to use your tracker for so long? • What got you in the habit of using your tracker? • Are there times of the day you use or notice your activity tracker more than other times of the day? • What makes it difficult to use the activity tracker on those days? How do you deal with the difficulties? • If a friend wanted to use an activity tracker every day for a year, what are the three main pieces of advice you would tell your friend?
Tracker data usage	• How often do you check your tracker data? • Do you share your tracker data? • How do you use the data provided by the tracker?
Second interview	
Interview topics	
Photo diary opening question	What physical activity were you doing when you took this photo?
Context	Where were you when you took this photo?
Benefits of wearable activity trackers	• How did you feel physically while doing this activity? • How do you see the wearable activity tracker contributed to your physical health? • Were you with anyone when you used the activity trackers? • How do you see the wearable activity track contributed to your social life?
Successful strategies of long-term use	Please explain to us the successful strategies you employed to continue using the wearable activity tracker, which are captured in these photos.

Data Analysis

Both rounds of interviews were recorded and transcribed verbatim with the use of a web-based transcription service, Rev. The transcribed interview data were systematically approached by applying the principles of thematic analysis [42-45]. Thematic analysis is a flexible, efficient, and accessible method that is widely used in qualitative psychology to generate rich data that provide social and psychological interpretations of phenomena and illuminate unpredicted insights relevant to the topic of the study [42]. While acknowledging the subjective nature of collected data, we used a postpositivist empirical approach to data analysis, which implies that the reality can be known through the accounts of humans—in our case, experiences and voices of long-term users of wearable activity trackers [42,43]. This approach is described as essentialist and experiential [42,43]. Furthermore, we used a mix of inductive and deductive techniques in data coding and theme generation. Data coding was predominantly an inductive, data-driven process to identify a wide array of patterns in participants' descriptions of long-term tracker use. Aggregating the codes within each theme was approached both inductively and deductively; we applied the literature of habit formation (deduction) and accounted for unanticipated findings (induction) to define themes [42,43].

Thematic analysis of the transcripts was performed in 5 stages, as suggested by previous literature [42,43]. These stages included (1) becoming familiar with the transcribed interview data, (2) coding, (3) outlining initial themes, (4) reviewing and revising themes, and (5) defining final themes [42,43]. First, each author read the transcripts to become familiar with the data (stage 1). Then, the transcripts were iteratively analyzed by 4 researchers using NVivo 10 and NVivo 11 software (QSR International) for qualitative data analysis [46,47]. All 4 researchers analyzed the same 3 randomly selected transcripts to generate nodes (stage 2). The researchers were then divided into 2 groups, and each group analyzed half of the transcripts. After completing the initial coding, the researchers reviewed the nodes and discussed disagreements and discrepancies. If 2 researchers coding the same transcript could not come to an agreement, a third researcher was invited to resolve the disagreement. The nodes were iteratively compared and revised during the coding process. Throughout the discussions of nodes, the initial theme formation started (stage 3). Themes were then generated, reviewed, and finalized based on the final nodes identified, and most relevant quotes were selected in support of the themes (stages 4 and 5). As the scope of the interviews was broader than in this study, only nodes and themes relevant to the RQs in this study are reported.

Trustworthiness and consistency of the findings were established through a number of procedures. The instrument for the interview was developed and refined by all members of the research team and reflected the goals of this study. The interviewers were properly trained to recruit participants and conduct telephonic interviews. As the data collection continued, the researchers met weekly to discuss the study progress and the highlights of the interviews. A codebook was established to analyze the data and derive nodes. Multiple coders coded the transcripts and met regularly to discuss progress and resolve inconsistencies and disagreements. Finally, quote relevance was carefully assessed to support the final themes [45].

Results

Overview

A total of 8 themes were identified to answer the 4 RQs. Table 3 summarizes the themes with exemplary quotes from the participants.

Table 3. Summary of identified themes.

RQ[a]	Themes	Example quotes
RQ1: How did long-term users initiate the use of wearable activity trackers?	Meaningful initial start	"She [The physician] laid it out to me there, end of February: 'You're going to have a choice. We're going to give you injections, or it's borderline [diabetes]. Exercise and eating right can reduce it." (Participant 20)
RQ2: How did long-term users set goals?	Goal setting: start with a small goal and gradually increase	"The tracker came preset with the 10,000 step goal. I think that's what the American Heart Association recommends. I didn't really push my steps for that goal because if I was walking, I would still have some pain in my chest……After I had that stress test after a year is when I really started picking up the pace and doing a lot more and working on the 10,000 steps every single day. There have been days where it's tracked 15,000 so I would be significantly over the goal some days. It did take me a while to work up to it." (Participant 4)
RQ3: What consistent contextual cues did long-term users use to form the habit of wearing wearable activity trackers?	Consistent cues: use time and locational cues to make it a routine	"If you develop a routine and a habit by having a place you put it every night, an agreement with yourself that it's the first thing you do when your feet hit the ground in the morning is put it on, then it will become a habit. Anything you do for 15 days in a row becomes a habit. If you can do it for 15 days, you probably will continue to do it." (Participant 15)
RQ4: How did long-term users engage in action planning and coping planning to support the continued use of wearable activity trackers?	• Action planning: creative contextual cues • Action planning: use reminders • Coping planning: anticipate problems and have a plan to deal with it • Coping planning: mindset for managing unfulfillment and lapses • Coping planning: try to have fun and try something new	• "While we were talking, I've been walking some because I knew I still lacked a few steps. And so, that's right. Anytime I talk on the phone, you know, since we all have the cell phones, I walk while I talk." (Participant 9) • "When you set up the account you can turn on the alerts and the alerts can go to your phone and can go to also your email. That is... I did both and it's a duplicate because the email goes to the phone, too, but...whatever gets my attention at the moment, that's fine. And then I go and charge. And it doesn't take long to charge." (Participant 10) • "I carry one with me. I have it in my bag that I carry to work, but I do have two chargers. So that's probably a good idea, too, to have more than one. I know one person I know has it in their car. Charges it in their car." (Participant 7) • "Put it on and don't be discouraged the first few days or even weeks or months you don't reach what you want, because it's like everything else. It's something new. It's something you've never done before." (Participant 1) • "Maybe in the spring so you can hear the birds and just do something like that to motivate you and look at the other side of town. When you walk over and say, hey I haven't been down in this area for quite a while. You've got to come up with something that motivates you to walk or to see something different, or whatever." (Participant 19)

[a]RQ: research questions.

Meaningful Initial Start

Some long-term users have been physically active for a long time, and wearable activity trackers are just another tool for them to monitor their physical activity automatically. However, many of our long-term users became physically active after they started to use wearable activity trackers. Most of the participants indicated that they took it very seriously when they started using the wearable activity tracker to increase their physical activity levels. Although some purchased the wearable activity tracker on their own and some received it as a gift, the majority of participants experienced significant events in their life that made them realize that they must do something to change their lifestyles. The primary event was declining health. Some participants experienced a gradual decline, followed by a significant health event such as surgery. Some participants were given either severe warning or gentle encouragement from their health providers to exercise to address the health issue:

> *Anyhow, but so what I'm saying is over those years my health had been declining without me really recognizing it because it wasn't just an abrupt thing. With the heart surgery, that being an abrupt event in my life, I realized that I definitely needed to do more*

> *to try to get the exercise and the sleep, everything that I needed as well as diet.* [Participant 4]
>
> *She [The physician] laid it out to me there, end of February: 'You're going to have a choice. We're going to give you injections, or it's borderline [diabetes]. Exercise and eating right can reduce it.* [Participant 20]

The aforementioned quotations were from individuals facing severe health challenges. Although not all participants had such dramatic life-threatening events that triggered them to begin using wearable activity trackers to help them exercise more, they mentioned other health-related reasons such as weight loss or weight management, better diet, increased physical activity and metabolism, and general preventative practice for healthy retirement life. It seems that for individuals to start forming a habit of long-term use, they needed to see a meaningful purpose behind the action to make up their mind and be determined to make a change.

Goal Setting: Start With a Small Goal and Gradually Increase

Goal setting was one of the most mentioned technological features of wearable activity trackers that assisted users in becoming long-term users. Before a habit is formed, goal setting is important. Different wearable devices on the market have a goal-setting feature and the goal is typically set at 10,000 steps for everyone by default. However, this goal target of 10,000 steps may not work for everyone, especially older adults or individuals who are relatively sedentary to start. Our participants did not always follow the default goal suggested by the device and instead set realistic goals based on their conditions and gradually modified their goals. Each round of goal setting and goal accomplishment increased participants' confidence and self-efficacy in using wearable activity trackers to monitor physical activity. It likely helped them form a habit and become long-term users. Participants acknowledged the importance of starting with a reasonable step count. Health problems often limited participants; however, gradually increasing their goal enabled the participants to meet their goals:

> *She [My daughter] has set it 10,000, which I found was just a little bit too much for me, so I changed it down to 4,000... Well, get it down to where it's more reasonable to start out with, and then move up... I would say set your goals realistically. Don't set them because you think this is what you should do.* [Participant 1]
>
> *The tracker came preset with the 10,000 step goal. I think that's what the American Heart Association recommends. I didn't really push my steps for that goal because if I was walking, I would still have some pain in my chest...After I had that stress test after a year is when I really started picking up the pace and doing a lot more and working on the 10,000 steps every single day. There have been days where it's tracked 15,000 so I would be significantly over the goal some days. It did take me a while to work up to it.* [Participant 4]

Consistent Cues: Use Time and Locational Cues to Make It a Routine

It sounds effortless: "just put it on." However, many individuals stop using wearable activity trackers because they cannot remember to put it on [19]. What are the tricks among the long-term wearable activity tracker users? As the habit formation literature predicts, volitional goal setting is the first step before the autonomous habit is formed, and consistent cues that trigger behaviors repeatedly move individuals toward habituation. Commonly discussed temporal cues are morning and night routines. Long-term users typically put on the wearable activity trackers the first thing in the morning. A time cue of going to bed reminds some users to plug in the wearable activity tracker to charge so that it is available for use the next day:

> *If you develop a routine and a habit by having a place you put it every night, an agreement with yourself that it's the first thing you do when your feet hit the ground in the morning is put it on, then it will become a habit. Anything you do for 15 days in a row becomes a habit. If you can do it for 15 days, you probably will continue to do it.* [Participant 15]

Another essential contextual cue is the location. Participants leave the wearable activity trackers at the same location at night or for charging, usually a place where they can easily spot, for example, nightstand, besides the clothes. Time and locational cues enabled participants to develop the habit of wearing a wearable activity tracker. Consistent cues regarding time and location provided the older adults a framework on which to develop a habit. As many of the participants had previously worn watches, the process of developing a routine for wearing a wearable activity tracker was identified as similar to wearing a watch. Many of the participants recognized that once they were able to develop the habit of wearing the wearable activity tracker, it became an automatic process:

> *It's just right there on the, where I lay my clothes out, so it's like okay, this is part of it. It's like putting your earrings in. This is just what you do.* [Participant 9]
>
> *You don't take it off except to charge it or take a bath or shower; you put it in a spot where you're gonna put it right back on if you do have to take it off. Just do it all the time and it becomes a habit.* [Participant 7]
>
> *I would say that the most important thing is the issue of getting into a regimen. They can get on an automatic thing like I do. Basically in the morning you just put it on. You don't even think about. It's like putting on my watch. It's an automatic thing. That's number one.* [Participant 18]

Action Planning: Creative Contextual Cues

Action planning helps establish the link between contextual cues and behaviors. Action planning involves detailed specification of the features of situations and intended responses to the situation, creating an if-then rule. Because long-term users have planned ahead of time what to follow once the contextual cues are triggered, they could easily engage in the planned actions. The long-term users in the study had a number of

creative contextual cues, such as commercial breaks, phone conversations, and ice and snow outside. For instance, if there was a commercial break or if participants were having a phone conversation, they would get up and walk. Wearable activity trackers most likely did not help participants plan action or create *if-then* rules; however, the devices provided the concrete behavior consequence (eg, step count), which is a way to validate their action planning:

> *At night or when I start watching TV in the evening, I get up during the commercials. All the commercials I walk.* [Participant 12]

> *While we were talking, I've been walking some because I knew I still lacked a few steps. And so, that's right. Anytime I talk on the phone, you know, since we all have the cell phones, I walk while I talk.* [Participant 9]

> *In the winter, because of the ice and snow, I don't get as much activity, steps and stuff like that, but I will walk around the house to try and get to my, the number of steps. If somebody would drive by and look in the window, they'd see me back and forth, back and forth, like I'm pacing.* [Participant 17]

Action Planning: Use Reminders

Long-term users also acknowledged that it was easy to forget to put on the wearable activity trackers when they first started, especially after they took it off for charging. One of the strategies they adopted was using reminders, for example, on their phones. Some participants used the alert feature available from the wearable activity tracker website to alert their phone or email regarding the need to charge the tracker. Participants tried to avoid charging the tracker last minute as often they would forget to put on the tracker. They actively planned to charge the tracker while engaging in sedentary behaviors and use reminders:

> *When you set up the account you can turn on the alerts and the alerts can go to your phone and can go to also your email. That is... I did both and it's a duplicate because the email goes to the phone, too, but...whatever gets my attention at the moment, that's fine. And then I go and charge. And it doesn't take long to charge.* [Participant 10]

Coping Planning: Anticipate Problems and Have a Plan to Deal With It

Coping planning, anticipating difficulties, and developing plans to overcome challenges was also a common practice among long-term users. They anticipated potential obstacles that might prevent them from wearing the wearable activity tracker and planned for how to deal with them. Once wearing wearable activity trackers becomes a daily routine, or in the participants' words, "becoming automatic," the long-term wearable activity tracker users are challenged to maintain a certain level or even increase physical activity levels. One participant identified coping planning to address the potential of a dead battery:

> *I carry one with me. I have it in my bag that I carry to work, but I do have two chargers. So that's probably a good idea, too, to have more than one. I know one person I know has it in their car. Charges it in their car.* [Participant 7]

Coping Planning: Mindset for Managing Unfulfillment and Lapses

Adopting a new behavior does not always mean that the behavior will be continuously maintained. Throughout behavior change, maintenance, and dealing with occasional lapses are essential [48]. Sometimes, people can be discouraged when lapses happen, and they stop using wearable activity trackers altogether [15]. The long-term tracker users shared that they were prepared for occasional unfulfillment and lapses with a positive mindset. The expectation of ups and downs helped prevent users from quitting at the beginning of forming the habit:

> *Put it on and don't be discouraged the first few days or even weeks or months you don't reach what you want, because it's like everything else. It's something new. It's something you've never done before.* [Participant 1]

Besides the positive mindset to view early road bumps as inevitable and prepare for it, long-term users also tried to find a positive angle to view lapses. For instance, they focused on the larger picture, such as a weekly goal achievement, rather than on a 1-day relapse:

> *I think just consistency is the most important thing. Wear it every day and there might be some days when you can't achieve your goal but if you look at it on a weekly basis, and say 'Okay, my goal for the week is 70,000 steps, every day it's 10,000. If something happens, if you're sick one day or you have business there or whatever and you only get in the 5,000, if you get in 12,000 two other days or 12,500, then you're still at your goal for the week.' That's one thing I kept in mind is that you look at the goal on a weekly basis as well as on a daily basis.* [Participant 4]

Coping Planning: Try to Have Fun and Try Something New

Engaging in the same behavior over and over again can be dull, even after one has already established as a habit. Most of our long-term users wear wearable activity trackers to track walking steps as a way to monitor their activity level. Taking the same route for walking every day certainly is a good way of keeping the number of step counts; however, it also can make wearing the wearable activity tracker unnecessary as today's step count would not be much different from yesterday's. One reason for the abandonment of a wearable activity tracker is that it is no longer needed once users have gauged their routine activity level [15]. The participants tried to explore new routes or new environments, making physical activity more interesting and justifying the need to wear the wearable activity tracker to gauge step counts:

> *Maybe in the spring so you can hear the birds and just do something like that to motivate you and look*

at the other side of town. When you walk over and say, hey I haven't been down in this area for quite a while. You've got to come up with something that motivates you to walk or to see something different, or whatever. [Participant 19]

Discussion

Summary of Findings

This qualitative study obtained valuable insights from 20 long-term wearable activity tracker users who were at least 65 years old to inform future development of interventions to increase wearable activity tracker use and physical activity among older adults. The value of this study is in its focus on factors that contribute to the long-term use of wearable activity trackers among older adults, explored through the theoretical framework of habit formation. Numerous studies have focused on the initial stages of health-related technology use [19,49,50], often neglecting the issue of long-term, sustainable engagement. Existing studies examining the continued use of wearable activity trackers and other personal informatics devices primarily focus on reasons for abandonment rather than long-term use strategies [21-23]. Studying sustainable wearable activity tracker use in the older adult population is seen as especially beneficial, as this particular group may greatly benefit from physical activity [40,51]. Older adults are susceptible to developing chronic diseases such as diabetes, cardiovascular diseases, and obesity, which could be protected against by regular physical activity [52]. Factors contributing to long-term use, in particular, successful strategies to support habit formation and long-term use, have implications not only for wearable activity tracker use but also for other information technology facilitated behavior change.

Our qualitative results confirmed 4 types of strategies, consistent with the habit formation literature, which were used by our participants to become long-term users of wearable activity trackers. Specifically, they set up realistic and adaptive goals in the early phase of habit formation and creatively relied on contextual cues to trigger goal-related behaviors automatically. They engaged in action planning, a form of mental stimulation for how to act when contextual cues are triggered, translating intentions, and goals to action. Finally, they deliberately did coping planning, another type of mental stimulation to anticipate obstacles and alternative behaviors to overcome those obstacles.

Goal Setting

Among the features that are available in wearable activity trackers, goal setting was one of the most mentioned helpful features. Goal setting, as an important step in the early phase of habit formation, is common in most wearable activity trackers. The default goal of 10,000 steps is not always obtainable for older adults, especially those who are struggling with health issues. One of the suggestions to help form sustained wearable activity tracker use habits that came from the interviewees was to start with realistic goals. If one wants to make a change, the desired outcome is achievable [34]. This approach not only encourages older adults to move but also boosts their self-efficacy as they achieve their daily activity goals [28,34]. Previous research on health apps has also shown that the ability to set customizable goals helps boost individuality, which contributes to the sustained motivation of improving physical fitness or weight loss [53]. For older adults who prioritize positive emotions and emotionally meaningful goals in life, adaptive goals that are attainable at different stages of behavior change may be especially important for their sense of meaning and well-being [54]. Designers of personal health informatics devices may consider realistic and adaptive goal setting based on users' initial use data rather than providing a default goal for everyone. For instance, when users start using the wearable activity tracker, it is crucial to analyze their age, current activity level, and health status to craft a realistic goal and then adaptively modify the goal based on the daily step count data from the user.

Consistent Contextual Cues

Automaticity emerged as the core *element* of the habit formation process. Most of the tricks to maintain use of wearable activity trackers revolved around making wearable activity tracker use routines less effortful or noticeable. Repetition and consistency play an important role [26]. To get into the habit, one must put the wearable activity tracker on at the same time every day. Participants tended to keep their wearable activity trackers in the same place, put them on as part of a grooming routine, or charge them at the same time. These consistent contextual cues helped them establish a routine of wearing wearable activity trackers that were incorporated into their everyday lives. Specifically, these time and location cues are the bases for action planning as they signal *when* and *where* actions should take place. Most recently, the HeartSteps study found that contextually tailored activity suggestions can be helpful for users to increase their step counts [55]. By using an Android-based mobile health app to send users actionable activity prompts tailored for their current context, including the time of day, day of the week (weekday or weekend day), location (home, work, and other), and weather (suitable for walking outside or not), delivered at common times when individuals can participate in physical activity, such as during commute, midafternoon, or after dinner, researchers found that delivering activity walking suggestions (vs providing no suggestions) increased the subsequent 30 min step count by 14% to 24% on average. In light of these findings, delivering activity suggestions on wearable activity trackers may have a more substantial effect on physical activity, considering that individuals have more immediate and constant access to them. The HeartSteps study also found that the effects of contextually tailored activity suggestions decreased as the study duration increased. Nevertheless, the study was a 6-week intervention, which may not be enough time to keep users engaged and form the habits of participating in physical activity.

Most existing wearable activity trackers do not have built-in features that facilitate the establishment of consistent contextual cues for habit formation. Long-term users in our study established these contextual cues based on their routines. Designers of future wearable activity trackers or personal health informatics devices may consider adding in features to facilitate users to establish contextual cues to enable habit formation and later use an automatic process. For instance, GPS is available in many wearable activity trackers, and smart wearable activity

trackers can use location data to infer location cues. If wearable activity trackers send push notifications to invite users to annotate these contextual cues, future push notifications can be sent as reminders when the contextual cues are triggered. In addition, contextual cues are very much individualized. In other words, each individual may have different *when* and *where* cues that are meaningful for them. Wearable activity trackers may offer a list of cues to action for users to identify personally feasible cues.

Action Planning

Action planning uses the so-called *if-then* rules to establish the link between contexts and behaviors. It includes specific situational parameters (*when*, *where*) and a sequence of actions (*how*). When such links are repeated over time, behaviors might be elicited automatically when relevant situational cues are encountered. Our participants engaged in action planning to use creative contextual cues and reminders for prolonged use of wearable activity trackers, adding empirical evidence for the theory of the habit formation process [28]. There is already evidence showing the effectiveness of action planning for behavior change in multiple domains, including physical activity [37,56]. Action planning should be explored further as a strategy for the continued use of wearable activity trackers or other health informatics devices.

Action planning was done mainly by long-term users without any assistance from their wearable activity trackers, which is a gap where designers of future wearable activity trackers or health informatics devices can fill. The strategies used by those long-term users may be programmed into wearable activity trackers to facilitate habituation and continued use. For instance, wearable activity trackers may send push notifications suggesting planned behavioral enactment if the *when* and *where* contextual cues are triggered. For instance, one example of action planning among our participants was to get up and walk some steps when television commercials were on. A television commercial is a *when* contextual cue that triggers the *if-then* rule for walking behavior. Future wearable activity trackers may prompt the users to choose a few from a list of *if-then* rules that are personally relevant to them for action planning. As evidenced by the HeartSteps study, along with similar just-in-time interventions delivered at critical decision points that are most suitable for facilitating individuals' health behavior change [57], there is great potential for integrating action planning with contextual cues in initiating and maintaining the process of habit formation as we move toward the future of precision health [58].

Next-generation wearable activity trackers or health informatics may even be equipped with smart sensors to detect the *when* and *where* contextual cues and provide just-in-time information to nudge the users to engage in the planned behavior. For instance, a smartwatch equipped with sound sensors can detect television sounds and send just-in-time prompts that it is time to walk when the commercial is on. It should be noted here that these suggested action planning scenarios are not about wearing wearable activity trackers but about increasing physical activity. Nevertheless, the primary goal of any wearable activity tracker is to increase physical activity, and the discussed scenario cannot be implemented if the user is not wearing a wearable activity tracker.

Coping Planning

Coping planning is a mental simulation of overcoming anticipated barriers to action. Good intentions or well-set goals are more likely to be translated into action when people develop success scenarios and preparatory strategies for approaching difficult tasks. A systematic review shows that coping planning is an effective technique for promoting health-related behavior change, especially in combination with action planning [59]. Relapse and barriers to action are always expected. The key is how to deal with them when they happen. Our participants stressed that disruptions in wearable activity tracker use should not be a reason to stop using wearable activity trackers. They noted the importance of continuing to use wearable activity trackers even if step levels were lower than desired or individuals had forgotten to use the wearable activity tracker for a while. This insight suggests that behavioral disruptions are to be expected; however, *getting back on the horse* and resuming wearable activity tracker use are integral to continued long-term wearable activity tracker use. This finding is also consistent with previous evidence that occasional deviation of behavior does not seriously impair habit formation as long as additional repetition continues [41].

Nevertheless, coping planning before disruption or relapse is essential for habit formation. Coping planning ensures the continuance of the habit formation process and the effectiveness of other elements, including goal setting, environmental cues, and action planning, for habit formation. The lack of premeditations or deliberate mental efforts inherent in the coping planning process may constitute a critical reason that people fail to stick with the usage of wearable activity trackers. Nevertheless, coping planning is also the most often neglected component among studies on the habitual use of wearable activity trackers [27,60]. Considering its significance in helping participants bridge the gap between their intentions and actions to engage in health behaviors [59], lack of support for coping planning is a weak spot of existing wearable activity trackers and health informatics devices. Designers may consider incorporating coping planning, especially coupling coping planning with different contexts that can be sensed by wearable activity trackers (eg, lack of data for several days) or contextual data (eg, weather information) to better facilitate habituation. For instance, when the weather forecast indicates snow for the next 5 days and the GPS of the wearable activity tracker shows that the user is mostly walking outside, a push notification may be programmed to send to the user to plan for indoor walking before snowfall.

It is sometimes difficult for individuals to anticipate barriers and devise coping planning ahead of time. Another approach might be to monitor behaviors and obstacles and plan for coping strategies when the same obstacles occur in the future. For example, a wearable activity tracker might detect that the user did not wear a wearable activity tracker for several days or did not log the targeted steps for several days. The wearable activity tracker could then send a push notification to the user to annotate

why this happened and how to deal with this if this happens again in the future.

Limitations

Although the results of our study provide suggestions for enhancing long-term wearable activity tracker use habit formation, our study is not without limitations. First, we interviewed 20 older adult long-term wearable activity tracker users from across the United States to discern how they formed long-term wearable activity tracker use habits. Although the participants were geographically dispersed across the United States, they were skewed to *younger* older adults, with an average age of approximately 68 years. Older adults in their 80s and beyond are likely to be different from those in their upper 60s in terms of wearable activity tracker use and long-term wearable activity tracker use. In addition, 18 out of 20 participants of our sample were White. We know little about how the patterns identified in these interviews might differ among a more diverse racial and ethnic sample. Future research is needed that follows a diverse sample of older adults over time as they begin to use wearable activity trackers to determine how habits are formed in wearable activity tracker use. Second, our study relied on self-report, which may be subject to recall bias given that the average time of wearable activity tracker use for our long-term wearable activity tracker users was almost 3 years, although we implemented a photo diary to facilitate recall, especially recall of habitual behaviors. Third, we relied on habit formation literature to examine the continued use of wearable activity trackers. However, there are other models available that predict continued use of technology, such as the expectation-confirmation model [61], which may be adopted in future research regarding the continued use of wearable activity trackers and other health informatics devices. Fourth, the focus of this study is on the continued use of wearable activity trackers, not necessarily habitual physical activity. Nevertheless, some participants equated wearing wearable activity trackers with physical activity, and some participants noted that wearing a wearable activity tracker was the first step in behavior change. We deliberately only included interview results about wearing wearable activity trackers; however, wearing wearable activity trackers and engaging in physical activities were intertwined in some interviews. Fifth, further advancement of wearable technology in the population of older adults may uncover additional barriers to continuous technology use. For example, advanced wearable activity trackers may require older adults to share personal information with the technology owners, which may create privacy-related concerns. Sixth, although the participation incentive was US $80 in total, it matched with the level of time commitment and effort because the participants needed to engage in 2 interviews that lasted for 60 to 90 min each and they also needed to keep a photo diary of wearable activity tracker use. The possibility that the financial incentives would introduce a potential bias toward keeping up the appearance of long-term users is low because only long-term users identified by the Qualtrics survey received recruitment information for the interview study. In other words, participants disclosed their length of wearable activity tracker use in the Qualtrics survey before they were aware of the interview study and its incentives. Finally, as this study focuses mainly on the behavioral factors that are conducive for forming habits of using wearable activity trackers, we did not consider the personality and motivational aspects of habit formation. Research has found that self-control, self-determination [62-64], conscientiousness, and a growth mindset [65] are predictive of healthy habit formation, which, in turn, is predictive of physical activity or exercise behaviors. For example, the positive relationship between trait self-control and physical activity is mediated through behavioral automaticity, a key component of habit [62]. The positive relationships between self-control and positive life outcomes, such as healthy eating and consistent sleep, are also mediated through beneficial habits [66]. Another study found that self-determined motivations are more predictive of behavioral automaticity than nonself-determined motivations across 12 common daily behaviors [63]. Even though we did not directly study physical activity, it is arguable that these personality and motivational factors may also have similar effects on the habit of wearing wearable activity trackers, which is a necessary tool for facilitating behavior change. We suggest that these relationships should be explored in more detail in future research.

Conclusions

The results of this study shed light on a range of factors that may contribute to long-term wearable activity tracker use among older adults. Although contextual factors, such as where the wearable activity tracker is placed and how often it is charged, and technological factors, such as build-in goals and reminders, are important. Participants also noted the importance of self-initiated actions and coping planning for forming wearable activity tracker use habits. The results of this study provide practical recommendations for behavior change interventions. In summary, the results of this qualitative study of long-term *power users* of wearable activity trackers suggest specific ways to enhance long-term habit formation for wearable activity tracker use among older adults. The findings add rich qualitative insights into the emerging literature on habit formation. Strategies identified by long-term *power users* could be useful in the design of next-generation wearable activity trackers or technological devices to facilitate behavior change.

Acknowledgments

The study was supported in part by the Michigan State University S3 grant.

Conflicts of Interest

None declared.

References

1. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
2. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304 [FREE Full text] [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
3. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
4. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. J Med Internet Res Mhealth Uhealth 2015 Nov 19;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
5. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. J Med Internet Res Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
6. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit use and activity levels throughout a physical activity intervention: Exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 05;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
7. Tang LM, Kay J. Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2017 Jun 30 Presented at: Proc ACM Interact Mob Wearable Ubiquitous Technol; Jun 30, 2017; - p. 1-28 URL: https://dl.acm.org/doi/10.1145/3090091 [doi: 10.1145/3090091]
8. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
9. Liu JY, Kor PP, Chan CP, Kwan RY, Sze-Ki D. The effectiveness of a wearable activity tracker (WAT)-based intervention to improve physical activity levels in sedentary older adults: A systematic review and meta-analysis. Archives of Gerontology and Geriatrics 2020;91:104211 [FREE Full text] [doi: 10.1016/j.archger.2020.104211]
10. Building a better tracker: older consumers weigh in on activity and sleep monitoring devices internet. American Association of Retired Persons. URL: http://www.aarp.org/content/dam/aarp/home-and-family/personal-technology/2015-07/innovation-50-project-catalyst-tracker-study-AARP.pdf [accessed 2020-12-22]
11. Cordeiro F, Epstein DA, Thomaz E, Bales E, Jagannathan AK, Abowd GD, et al. Barriers and negative nudges: exploring challenges in food journaling. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:1159-1162 [FREE Full text] [doi: 10.1145/2702123.2702155] [Medline: 26894233]
12. Inside wearables Part 1: how behavior change unlocks long-term engagement internet. Endeavour Partners. 2014 Jan. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2020-12-22]
13. Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. New York: ACM Digital Library; 2015 Sep Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan p. 647-658. [doi: 10.1145/2750858.2807554]
14. Epstein DA, Kang JH, Pina LR, Fogarty J, Munson SA. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. In: Proc ACM Int Conf Ubiquitous Comput. New York: ACM Digital Library; 2016 Sep 12 Presented at: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September, 2016; Heidelberg Germany p. 829-840 URL: http://europepmc.org/abstract/MED/28516173 [doi: 10.1145/2971648.2971656]
15. Epstein D, Ping A, Fogarty J, Munson S. A lived informatics model of personal informatics. In: UbiComp'15. New York: ACM Digital Library; 2015 Sep Presented at: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sept 2015; Osaka, Japan p. 731-742 URL: https://dl.acm.org/doi/10.1145/2750858.2804250 [doi: 10.1145/2750858.2804250]
16. Lazar A, Koehler C, Tanenbaum TJ, Nguyen DH. Why we use and abandon smart devices. In: UbiComp'15. New York: ACM Digital Library; 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sept 2015; New York p. 635-646 URL: https://dl.acm.org/doi/10.1145/2750858.2804288 [doi: 10.1145/2750858.2804288]
17. Chuah SH. You inspire me and make my life better: investigating a multiple sequential mediation model of smartwatch continuance intention. Telematics and Informatics 2019 Oct;43:101245 [FREE Full text] [doi: 10.1016/j.tele.2019.101245]
18. Harrison D, Marshall P, Berthouze N, Bird J. Tracking physical activity: problems related to running longitudinal studies with commercial devices. In: UbiComp '14 Adjunct. New York: ACM Digital Library; 2014 Sep Presented at: Proceedings

of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication; 2012; New York p. 699-702 URL: https://dl.acm.org/doi/10.1145/2638728.2641320 [doi: 10.1145/2638728.2641320]
19. Shih PC, Han K, Poole ES, Rosson MB, Carroll JM. Use and adoption challenges of wearable activity trackers. In: iConference 2015 Papers. 2015 Mar 15 Presented at: IConference Proc Internet; 2015; - p. 164 URL: https://www.ideals.illinois.edu/handle/2142/73649
20. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. J Med Internet Res Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
21. Fritz T, Huang E, Murphy G, Zimmermann T. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2014 Apr Presented at: SIGCHI Conference on Human Factors in Computing Systems; April, 2014; Toronto p. A URL: https://dl.acm.org/doi/proceedings/10.1145/2556288 [doi: 10.1145/2556288.2557383]
22. Li L, Peng W, Kononova A, Bowen M, Cotten S. Factors associated with older adults' long-term use of wearable activity trackers. Telemed E-Health 2020 Jun;26(6):769-765 [FREE Full text] [doi: 10.1089/tmj.2019.0052] [Medline: 31553281]
23. Hermsen S, Moons J, Kerkhof P, Wiekens C, De GM. Determinants for sustained use of an activity tracker: observational study. J Med Internet Res Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
24. Verplanken B, Aarts H, van Knippenberg A, Moonen A. Habit versus planned behaviour: a field experiment. Br J Soc Psychol 1998 Mar;37((Pt 1)):111-128 [FREE Full text] [doi: 10.1111/j.2044-8309.1998.tb01160.x] [Medline: 9554090]
25. Verplanken B, Aarts H. Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? Eur Rev of Soc Psych 2011 Apr 15;10(1999):101-134 [FREE Full text] [doi: 10.1080/14792779943000035]
26. Wood W, Neal DT. The habitual consumer. J of Cons Psych 2009 Oct 2;19(4):579-592 [FREE Full text] [doi: 10.1016/j.jcps.2009.08.003]
27. Ellingson LD, Lansing JE, DeShaw KJ, Peyer KL, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. J Med Internet Res Mhealth Uhealth 2019 Feb 14;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
28. Lally P, Gardner B. Promoting habit formation. Health Psych Rev 2011 Oct 11;7(sup1):S137-SS58 [FREE Full text] [doi: 10.1080/17437199.2011.603640]
29. Bargh JA. The four horsemen of automaticity: awareness, intention, efficiency, control in social cognition. In: Wyer RS, Srull TK, editors. Handbook of Social Cognition. Hillsdale, NJ: Lawrence Erlbaun Associates Publishers; 1994:1-40.
30. Lally P, Wardle J, Gardner B. Experiences of habit formation: a qualitative study. Psychol Health Med 2011 Aug;16(4):484-489 [FREE Full text] [doi: 10.1080/13548506.2011.555774] [Medline: 21749245]
31. Gardner B, de Bruijn G, Lally P. A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. Ann Behav Med 2011 Oct;42(2):174-187 [FREE Full text] [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]
32. Verplanken B. Beyond frequency: habit as mental construct. Br J Soc Psychol 2006 Sep;45(Pt 3):639-656 [FREE Full text] [doi: 10.1348/014466605X49122] [Medline: 16984725]
33. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev 2007 Oct;114(4):843-863 [FREE Full text] [doi: 10.1037/0033-295X.114.4.843] [Medline: 17907866]
34. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of 'habit-formation' and general practice. Br J Gen Pract 2012 Dec;62(605):664-666 [FREE Full text] [doi: 10.3399/bjgp12X659466] [Medline: 23211256]
35. Gollwitzer P. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-403 [FREE Full text] [doi: 10.1037/0003-066X.54.7.493]
36. Einstein GO, McDaniel MA. Normal aging and prospective memory. J Exp Psychol Learn Mem Cogn 1990 Jul;16(4):717-716 [FREE Full text] [doi: 10.1037//0278-7393.16.4.717] [Medline: 2142956]
37. Sniehotta F, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul 15;35(4):565-576 [FREE Full text] [doi: 10.1002/ejsp.258]
38. O'Brien T, Troutman-Jordan M, Hathaway D, Armstrong S, Moore M. Acceptability of wristband activity trackers among community dwelling older adults. Geriatr Nurs 2015 Mar;36(2 Suppl):S21-S25 [FREE Full text] [doi: 10.1016/j.gerinurse.2015.02.019] [Medline: 25771957]
39. Sookhai L, Coppola J, Gaur C. Intergenerational activity tracker program: impact with health related outcomes on older adults. 2015 Long Isl Syst Appl Technol 2015:7 [FREE Full text] [doi: 10.1109/LISAT.2015.7160218]
40. Valenti G, Bonomi A, Westerterp K. Walking as a contributor to physical activity in healthy older adults: 2 week longitudinal study using accelerometry and the doubly labeled water method. J Med Internet Res Mhealth Uhealth 2016 Jun 07;4(2):e56 [FREE Full text] [doi: 10.2196/mhealth.5445] [Medline: 27268471]

41. Lally P, van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-909 [FREE Full text] [doi: 10.1002/ejsp.674]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res in Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
43. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. Washington DC: American Psychological Association; 2012:57-71.
44. Boyatzis RE. Transforming qualitative information: thematic analysis and code development. In: Qual Res. Thousand Oaks: SAGE Publications, Inc; 1998.
45. Guest G, MacQueen KM, Namey EE. Applied thematic analysis. In: Quali Res. Thousand Oaks: SAGE Publications; 2011.
46. QSR International Pty Ltd. NVivo. 2014. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-12-22]
47. QSR International Pty Ltd. NVivo. 2015. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-12-22]
48. Prochaska JO. Decision making in the transtheoretical model of behavior change. Med Decis Making 2008;28(6):845-849 [FREE Full text] [doi: 10.1177/0272989X08327068] [Medline: 19015286]
49. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J of Biomed Informatics 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153]
50. Spil T, Kijl B, Romijnders V. The adoption and diffusion of wearables. In: CT Unbounded, Social Impact of Bright ICT Adoption: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT, TDIT 2019, Accra, Ghana, June 21–22, 2019, Proceedings. Switzerland: Springer; May 16, 2019:31-47.
51. Elsawy B, Higgins K. Physical activity guidelines for older adults. Am Fam Physician 2010 Jan 1;81(1):55-59 [FREE Full text]
52. 2008 Physical Activity Guidelines Advisory Report. U.S Department of Health and Human Services. 2008. URL: https://health.gov/paguidelines/2008/report/ [accessed 2020-12-22]
53. Murnane E, Huffaker D, Kossinets G. Mobile health apps: Adoption, adherence, and abandonment. : ACM Press; 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable; 2015; Osaka, Japan URL: http://stanford.edu/~emurnane/files/UbiComp15_Mobile.pdf [doi: 10.1145/2800835.2800943]
54. Carstensen LL, Isaacowitz DM, Charles ST. Taking time seriously. A theory of socioemotional selectivity. Am Psychol 1999 Mar;54(3):165-181 [FREE Full text] [doi: 10.1037//0003-066x.54.3.165] [Medline: 10199217]
55. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartsteps. Ann Behav Med 2019 May 03;53(6):573-572 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
56. Carraro N, Gaudreau P. Spontaneous and experimentally induced action planning and coping planning for physical activity: a meta-analysis. Psychol of Sport and Exercise 2013 Mar;14(2):228-248 [FREE Full text] [doi: 10.1016/j.psychsport.2012.10.004]
57. Hiremath S, Amiri A, Thapa-Chhetry B, Snethen G, Schmidt-Read M, Ramos-Lamboy M, et al. Mobile health-based physical activity intervention for individuals with spinal cord injury in the community: a pilot study. PLoS One 2019;14(10):e0223762 [FREE Full text] [doi: 10.1371/journal.pone.0223762] [Medline: 31613909]
58. Hekler E, Tiro J, Hunter C, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. Ann Behav Med 2020 Nov 01;54(11):805-826 [FREE Full text] [doi: 10.1093/abm/kaaa018] [Medline: 32338719]
59. Kwasnicka D, Presseau J, White M, Sniehotta FF. Does planning how to cope with anticipated barriers facilitate health-related behaviour change? A systematic review. Health Psychol Rev 2013 Feb 19;7(2):129-145 [FREE Full text] [doi: 10.1080/17437199.2013.766832]
60. Rieder A, Lehrer C, Jung R. Understanding the habitual use of wearable activity trackers. In: Conf Wirtsch Siegen, Germany. 2019 Presented at: 14th Int Conf Wirtsch Siegen, Germany; 2019; Siegen p. 1002-1016 URL: https://research.cbs.dk/en/publications/understanding-the-habitual-use-of-wearable-activity-trackers
61. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001 Sep;25(3):351 [FREE Full text] [doi: 10.2307/3250921]
62. Pfeffer I, Strobach T. Behavioural automaticity moderates and mediates the relationship of trait self-control and physical activity behaviour. Psychol Health 2018 Jul;33(7):925-940 [FREE Full text] [doi: 10.1080/08870446.2018.1436176] [Medline: 29431514]
63. Radel R, Pelletier L, Pjevac D, Cheval B. The links between self-determined motivations and behavioral automaticity in a variety of real-life behaviors. Motiv Emot 2017 May 22;41(4):443-454 [FREE Full text] [doi: 10.1007/s11031-017-9618-6]
64. Gardner B, Lally P. Does intrinsic motivation strengthen physical activity habit? Modeling relationships between self-determination, past behaviour, and habit strength. J Behav Med 2013 Oct;36(5):488-487 [FREE Full text] [doi: 10.1007/s10865-012-9442-0] [Medline: 22760451]

65. Orvidas K, Burnette J, Russell V. Mindsets applied to fitness: growth beliefs predict exercise efficacy, value and frequency. Psychol of Sport and Exercise 2018 May;36:156-151 [FREE Full text] [doi: 10.1016/j.psychsport.2018.02.006]
66. Galla BM, Duckworth AL. More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. J Pers Soc Psychol 2015 Sep;109(3):508-525 [FREE Full text] [doi: 10.1037/pspp0000026] [Medline: 25643222]

Abbreviations

RQ: research question

Edited by L Buis; submitted 13.07.20; peer-reviewed by S Six, J Li, L Genaro, R Krukowski, V Shah, A McLean; comments to author 11.08.20; revised version received 03.11.20; accepted 30.11.20; published 19.01.21.

Please cite as:
Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M
Habit Formation in Wearable Activity Tracker Use Among Older Adults: Qualitative Study
JMIR Mhealth Uhealth 2021;9(1):e22488
URL: http://mhealth.jmir.org/2021/1/e22488/
doi:10.2196/22488
PMID:33464216

Corrigenda and Addenda

Correction: The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study

Michael Sobolev[1,2], PhD; Rachel Vitale[3], BA; Hongyi Wen[1], BSc, MSc; James Kizer[1], BSc, MSc; Robert Leeman[4], PhD; J P Pollak[1], PhD; Amit Baumel[5], PhD; Nehal P Vadhan[2], PhD; Deborah Estrin[1], PhD; Frederick Muench[3], PhD

[1]Cornell Tech, Cornell University, New York City, NY, United States

[2]Feinstein Institute for Medical Research, Northwell Health, Great Neck, NY, United States

[3]The Partnership to End Addiction, New York, NY, United States

[4]College of Health and Human Performance, Department of Health Education and Behavior, University of Florida, Gainsville, FL, United States

[5]University of Haifa, Haifa, Israel

Corresponding Author:
Frederick Muench, PhD
The Partnership to End Addiction
485 Lexington Avenue
3rd Floor
New York, NY, 10017
United States
Phone: 1 9175320623
Email: fmuench@toendaddiction.org

Related Article:

Correction of: https://mhealth.jmir.org/2021/1/e25018/

(JMIR Mhealth Uhealth 2021;9(1):e27439) doi:10.2196/27439

In "The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study" (JMIR Mhealth Uhealth 2021;9(1):e25018) the authors noted one error.

In the originally published article, in the subsection "Measurement of Impulsive Behavior," a URL was inadvertently inserted following the sentence "At the same time, trait-based studies can be confounded by other factors, including environment, mood, cognition, and social setting [2,28,29], and are heavily influenced by current state and context." This URL was unrelated to the article and has been removed from the corrected version.

The correction will appear in the online version of the paper on the JMIR Publications website on January 26, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.01.21; this is a non–peer-reviewed article;accepted 25.01.21; published 26.01.21.

Please cite as:
Sobolev M, Vitale R, Wen H, Kizer J, Leeman R, Pollak JP, Baumel A, Vadhan NP, Estrin D, Muench F
Correction: The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study
JMIR Mhealth Uhealth 2021;9(1):e27439
URL: http://mhealth.jmir.org/2021/1/e27439/
doi:10.2196/27439
PMID:33497353

Original Paper

Examining the Correlation Between Depression and Social Behavior on Smartphones Through Usage Metadata: Empirical Study

Yameng Wang[1,2], MSc; Xiaotong Ren[3], BA; Xiaoqian Liu[1,4], PhD; Tingshao Zhu[1,4], Prof Dr

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[2]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China

[3]College of Foreign Languages, Capital Normal University, Beijing, China

[4]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Corresponding Author:
Tingshao Zhu, Prof Dr
CAS Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
16 Lincui Road, Chaoyang District
Beijing, 100101
China
Phone: 86 15010965509
Email: tszhu@psych.ac.cn

Abstract

Background: As smartphone has been widely used, understanding how depression correlates with social behavior on smartphones can be beneficial for early diagnosis of depression. An enormous amount of research relied on self-report questionnaires, which is not objective. Only recently the increased availability of rich data about human behavior in digital space has provided new perspectives for the investigation of individual differences.

Objective: The objective of this study was to explore depressed Chinese individuals' social behavior in digital space through metadata collected via smartphones.

Methods: A total of 120 participants were recruited to carry a smartphone with a metadata collection app (MobileSens). At the end of metadata collection, they were instructed to complete the Center for Epidemiological Studies-Depression Scale (CES-D). We then separated participants into nondepressed and depressed groups based on their scores on CES-D. From the metadata of smartphone usage, we extracted 44 features, including traditional social behaviors such as making calls and sending SMS text messages, and the usage of social apps (eg, WeChat and Sina Weibo, 2 popular social apps in China). The 2-way ANOVA (nondepressed vs depressed × male vs female) and multiple logistic regression analysis were conducted to investigate differences in social behaviors on smartphones among users.

Results: The results found depressed users received less calls from contacts (all day: $F_{1,116}$=3.995, P=.048, η_2=0.033; afternoon: $F_{1,116}$=5.278, P=.02, η_2=0.044), and used social apps more frequently (all day: $F_{1,116}$=6.801, P=.01, η_2=0.055; evening: $F_{1,116}$=6.902, P=.01, η_2=0.056) than nondepressed ones. In the depressed group, females used Weibo more frequently than males (all day: $F_{1,116}$=11.744, P=.001, η_2=0.092; morning: $F_{1,116}$=9.105, P=.003, η_2=0.073; afternoon: $F_{1,116}$=14.224, P<.001, η_2=0.109; evening: $F_{1,116}$=9.052, P=.003, η_2=0.072). Moreover, usage of social apps in the evening emerged as a predictor of depressive symptoms for all participants (odds ratio [OR] 1.007, 95% CI 1.001-1.013; P=.02) and male (OR 1.013, 95% CI 1.003-1.022; P=.01), and usage of Weibo in the morning emerged as a predictor for female (OR 1.183, 95% CI 1.015-1.378; P=.03).

Conclusions: This paper finds that there exists a certain correlation between depression and social behavior on smartphones. The result may be useful to improve social interaction for depressed individuals in the daily lives and may be insightful for early diagnosis of depression.

(JMIR Mhealth Uhealth 2021;9(1):e19046) doi:10.2196/19046

KEYWORDS

depression; digital phenotyping; social behavior; smartphone usage; mobile sensing

Introduction

Depression is a mental disorder that is widespread in the world, with more than 300 million people affected [1]. Subthreshold depressive symptoms are far more common, causing significant impairment in people's lives and putting them at risk for future mental health concerns [2]. Depression is not only associated with decreased quality of life [3], decreased work productivity [4], and physical illnesses such as cardiovascular problems [5] and Alzheimer disease [6], but also increases the mortality and suicide rate [7,8]. The World Health Organization estimates that depression will be the leading cause of disease burden by 2030 [9].

Although depression is treatable using different methods, including antidepressants and psychotherapy, only fewer than half of those eligible received treatment [10,11]. One of the main obstacles is the difficulty of diagnosing depression. For instance, guideline-concordant care for depressed patients seeking treatment is typically initiated by primary care physicians [12,13], whereas primary care physicians might fail to identify most patients with depressive symptoms [14,15]. Because early treatment and intervention are shown to be associated with a better prognosis [16], more efficient methods of identifying depression could significantly improve the delivery of services to those in need.

Currently, smartphone has led to an enormous increase in personal convenience and effectiveness, with more than 3 billion users worldwide [17]. Despite the many uses and advantages of smartphones, they also bring many negative effects on the mental health of individuals (eg, depression). There has been evidence that general and problematic smartphone usage commonly co-occur with depression. On the one hand, some studies found that depression can cause smartphone addiction, because depressed individuals take their smartphones as a coping method to deal with their depressive emotion [18,19]. On the other hand, some studies found that high smartphone usage is associated with subsequent stress, sleep di culties, and depression [20,21]. Furthermore, some studies found a bidirectional relationship between smartphone usage and depression [22]. Specifically, depressed individuals may be driven to excessively use their smartphone to get rid of negative emotions, but this excessive smartphone usage consequently elicits more sleep problems, depression, irritability, and stress.

Communication and interpersonal interaction are among the most important functions of smartphones, and much research has been conducted on the relationship between depression and social behaviors on smartphones. While a smartphone is essential for communication and interpersonal interaction, it may make people less engaged with their real-life social environment [23-25]. Several studies found that depressed users have been reported to make fewer calls, but send more SMS text messages [26-28]. Kim et al [22] found that depressed users may rely on smartphones to alleviate their negative feelings and spend more time on communication, which in turn can deteriorate into problematic outcomes. Tamura et al [29] found that smartphone usage of 2 hours or more per day for social network services and online chats is associated with a higher risk of depression.

However, the majority of relevant studies measured smartphone usage using a self-report methodology. And self-reported smartphone usage did not correlate robustly with objectively measured smartphone usage [30,31]. Only recently, the increased availability of smartphone-based passive sensing data has provided new perspectives for the investigation of physical, individual differences, or mental health [32-34]. For instance, Saeb et al [35] found that both smartphone usage duration and frequency are positively associated with depression by analyzing daily life behavioral markers obtained using smartphone GPS and usage sensors. Using a smartphone-based self-monitoring system, Faurholt et al [36] found that depressive state in patients with bipolar disorder was related to more screen time, while manic state was related to frequent smartphone usage. Furthermore, it has been demonstrated that behavior data collected via smartphone can be used for monitoring individual depressive states [37], detecting social rhythms in bipolar disorder [38], and mobile intervention for depression [39]. Therefore, we aim to conduct research on depressed users' social behavior in digital space through metadata collected via smartphones.

To gain deeper insights, we include a broad range of behaviors based on previous studies which have reported on behavioral manifestations of depression in various types of behavior. For example, traditional communication behaviors (eg, making calls, sending SMS text messages) have been shown to be associated with depressive symptoms in many studies [23,27,28]. Moreover, as Baker et al [40] reviewed, there is a complex relationship between online social networking and depressive symptoms. Given that 99.1% of internet users are mobile internet users in China [41], usage of social apps on smartphones may be also associated with depression.

In addition, gender differences exist not only in depression but also in smartphone usage. Specifically, males with depression show impairment at lower symptom levels than females [42], and report consistently fewer symptoms than females [43]. As for gender differences in smartphone usage, several studies found that female users' smartphone usage is typically related to sociability, interpersonal relationships, and the creation of new relationships [44,45], but for male users, smartphone usage is simultaneously based on SMS text messages and voice conversations [46,47]. Furthermore, virtually all the studies indicate that females have higher levels of dependence and problematic usage than males [48,49]. Therefore, gender differences may also exist in the social behaviors on smartphones of depressed users.

By analyzing smartphone usage, this paper aims to examine the correlation between depression and social behavior on smartphones through metadata collected via smartphones. Specifically, we hypothesized that nondepressed and depressed

users may have differences in social behaviors on smartphones, and there are gender differences in social behavior among depressed users.

Methods

Study Procedure

In this study, a 2-step procedure was conducted: (1) data collection and (2) data analysis. In order to collect metadata of smartphone usage automatically, we developed an Android app named MobileSens [50,51]. MobileSens consists of 2 modules, one is for collecting metadata of smartphone usage, and the other is for filling the questionnaire, which can get smartphone usage data and the corresponding questionnaire results at the same time.

Data Collection

Participants

We recruited participants in Beijing, China, who own Android smartphones and also use smartphones in daily life by advertising on social networks. During this study, all participants were instructed to install MobileSens on their smartphones, and use it for more than 1 month. Given psychological measurement scales are composed of retrospective questions, participants were asked to fill out the corresponding questionnaires on the last day of metadata collection to ensure consistency between metadata and participants' psychological state. To encourage participants to take part in our study, we offered them a monetary reward depending on the number of days the smartphone usage data was uploaded. In particular, if participants finished uploading 30 days' data and completed the questionnaires, we rewarded them with RMB 200 (US $30). However, if participants dropped out of the study after finishing questionnaires and uploading data for less than 30 days, we provided experimental rewards based on the number of days the data were uploaded. This study was approved by the Review Board of Institute of Psychology, Chinese Academy of Sciences, H09036. Signed informed consent was obtained from each participant. Finally, we acquired data from a total of 120 participants (73 males, 47 females) with a mean age of 23.57 (SD 3.09).

Metadata of Smartphone Usage

MobileSens ran in the background to collect designated metadata. After participants installed MobileSens into their smartphone, most of their interactions with the smartphone were recorded and no sensitive personal data, such as the specific content of the SMS text messages, were collected. This app sensed a total of 15 categories of smartphone interaction events, as listed in Table 1. All the data would be uploaded via the internet to a remote central data server.

Table 1. Details on smartphone log data.

Category	Record content
App activity log	App creation/resume/launch/stop/exit
App package log	App package uninstallation/installation/replacement/change/data cleanup/restart
Call log	Number, contact name, and direction of call
Camera log	Use of camera button
Contact log	Contacts addition/deletion
Date change log	Change system date and time
GPS log	User's locale, altitude, latitude, longitude, and direction of movement
Headset log	Plug in/off headset
Location changed log	Change location
Power connected log	Connect/disconnect the power
Power log	Power on/off
Screen log	Screen on/off
Service app log	Service app creation/resume/launch/stop/exit
SMS text message log	Number, contact name, and direction of SMS text message
Wallpaper log	Change wallpaper

Questionnaires

On the last day of the study, each participant was required to complete an online assessment using MobileSens, consisting of a demographic questionnaire and a depression measurement scale. Depressive symptoms (minor depression) were evaluated using the Center for Epidemiological Studies-Depression Scale (CES-D) [52], which is widely used to measure depressive symptoms severity in the general population and is applicable to different ethnic contexts [53]. The CES-D measures the frequency of 20 depressive symptoms over the past week on a 4-point Likert scale. The total score may range from 0 to 60, with higher scores indicating higher levels of depression. In this study the Chinese version of questionnaire was implemented, and their validity and reliability have been proved previously [54].

Data Analysis

Feature Extraction

Among the metadata of smartphone usage sensed by MobileSens, the analyses of this work focused on call log, SMS text message log, and app activity log. These metadata are directly or, in some cases, indirectly related to individuals' social behavior on smartphones. For instance, call log and SMS text message log are the direct output of individuals' communication behavior; app activity log belonging to social apps can reflect individuals' online social activities, although social apps also have entertainment and information acquisition functions. Specifically, we designed 2 categories of social behaviors on metadata:

Traditional Communication Behavior

We mainly focused on the frequency of usage of calls and SMS text messages. In particular, we designed 4 kinds of behaviors related to traditional communication behavior from the metadata (making calls, receiving calls, sending SMS text messages, and receiving SMS text messages). In order to analyze the communication behavior between individuals and acquaintances (eg, contacts on smartphones), we designed another 4 kinds of communication behaviors that occur between contacts, which included making calls to contacts, receiving calls from contacts, sending SMS text messages to contacts, and receiving SMS text messages from contacts. Finally, we obtained a total of 8 behaviors.

Usage of Social Apps

We analyzed the frequency of usage of social apps from the following 2 aspects. First, we analyzed the frequency of usage of all social apps. Specifically, using a web crawler, we crawled an app classification framework from the Wandoujia app store (one of the most popular Android app stores in China), and got a list of all the social apps. Then, we extracted from metadata the usage records of all social apps for social apps' usage behavior. Second, we analyzed the frequency of usage of WeChat and Sina Weibo, 2 popular social apps in China, respectively. Finally, we designed a total of 3 behaviors.

After designing the above 2 categories of social behaviors, we extracted features from each category of behavior. Specifically, after calculating the frequency of all behaviors per day for each participant, we used the mean of the frequency of these behaviors as the feature of social behavior on smartphone all day (0:00 to 24:00). In addition, social activity represents the most important factor disrupting circadian rhythms after alternation of light/dark cycle [55], meaning that individuals' social behavior may differ between day and night. Moreover, circadian rhythms have repeatedly been linked to individual depression [56,57]. Thus, it may make sense to analyze individuals' social behaviors on smartphones during the day and at night. Further, some studies have found that many people with depression show a regular daily pattern of symptoms, usually with more severe symptoms in the morning [58,59]. To explore this difference, after dividing a day into day and night, we divided the day into morning and afternoon (morning: 6:00 to 12:00; afternoon: 12:00 to 18:00; evening: 18:00 to 6: 00). And then we calculated the average frequency of smartphone social behavior in different periods as the final features. Finally, we extracted a total of (8 + 3) × (1 + 3) = 44 behavior features.

Depressive Symptoms

To examine differences in social behaviors on smartphones between depressed and nondepressed users, we divided participants into 2 groups based on their scores on CES-D. The score of 16 has been traditionally used as a cutoff point for determining whether a person has symptoms of depression not only in Western countries [60], but also in China [54]. In this study, the cutoff point of 16 was used to distinguish individuals considered to be depressive from those classified as nondepressed.

Statistical Analysis

Descriptive statistics were computed as the mean and SD for continuous variables and absolute frequencies and percentage for categorical variables. The chi-square test and independent *t*-test were used to compare demographics between 2 groups. The 2-way ANOVA of depressive symptom (nondepressed group vs depressed group) × gender (male vs female) was conducted to examine the differences in social behaviors. Multiple logistic regression analysis with dichotomous depressive symptoms (nondepressed vs depressed) as dependent variables was used to find predictors of depressive symptoms. In this analysis, variable selection was performed using stepwise forward selection, subsequently including one by one the variables that were not statistically significant (α=.05). To improve the chances of retaining meaningful predictor variables, demographics that differed between the groups at P<.1, and behavior features that differed in the main effect of depressive symptoms in 2-way ANOVA at P<.1 were analyzed as independent variables via multiple logistic regression analysis [61]. For further data exploration of gender differences, multiple logistic regression models adjusted for gender were conducted. All statistical tests were performed using SPSS version 23 for Windows (IBM). The level of significance was set at .05.

Results

Demographics and Questionnaire Scores of CES-D

Of the 120 participants who formed the study sample, 71 participants were divided into the nondepressed group and 49 participants were divided into the depressed group. Demographics of participants are summarized in Table 2. The results of the chi-square test and independent *t*-test on demographics between the nondepressed and depressed groups showed no significance (gender: P=.76; education: P=.34; living place: P=.99; occupation: P=.75; marital status: P=.36; and age: P=.31).

Table 2. Demographics of participant versus depressive symptoms.

Characteristic	Nondepressed group (N=71)	Depressed group (N=49)	χ^2 (*df*) or *t* (*df*)	*P*-value
Gender, n (%)			0.10 (1)	.76[a]
Male	27 (38)	20 (41)		
Female	44 (62)	29 (59)		
Education, n (%)			2.16 (2)	.34[a]
Below university diploma	3 (4)	0 (0)		
University diploma	28 (39)	21 (43)		
Above university diploma	40 (56)	28 (57)		
Living place, n (%)			0.02 (2)	.99[a]
City	37 (52)	26 (53)		
Town	13 (18)	9 (18)		
Country	21 (30)	14 (29)		
Occupation, n (%)			0.10 (1)	.75[a]
Student	64 (90)	45 (92)		
Others	7 (10)	4 (8)		
Marital status, n (%)			0.83 (1)	.36[a]
Single/windowed	32 (45)	18 (37)		
Married/in a relationship	39 (55)	31 (63)		
Age, mean (SD)	23.34 (2.74)	23.92 (3.54)	1.01 (118)	.31[b]

[a]Chi-square test.

[b]Independent *t*-test.

Differences in Social Behaviors on Smartphone

To examine differences in social behavior features between the nondepressed and depressed groups, we conducted 2-way ANOVA of depressive symptom (nondepressed group vs depressed group) × gender (male vs female). All social behavior features with significant results are shown in Table 3.

For "usage of traditional communication functions," 2 behavioral features had significant differences. Specifically, in terms of "receiving calls from contacts all day," the main effect of depressive symptom was significant ($F_{1,116}$=3.995, *P*=.048,η^2=0.033). The frequency of receiving calls from contacts all day was significantly higher in the nondepressed group than in the depressed group (*P*=.048). By contrast, the main effect of gender and the interaction effect of depressive symptom and gender were not significant ($F_{1,116}$=0.005, *P*=.94, η^2=0.000; $F_{1,116}$=0.010, *P*=.92, η^2=0.000). As for "receiving calls from contacts in the afternoon," the main effect of depressive symptom was significant ($F_{1,116}$=5.278, *P*=.02, η^2=0.044). The frequency of receiving calls from contacts in the afternoon was significantly higher in the nondepressed group than in the depressed group (*P*=.02). However, the main effect of gender and the interaction effect of depressive symptom and gender were not significant ($F_{1,116}$=0.004, *P*=.95, η^2=0.000; $F_{1,116}$=0.006, *P*=.94, η^2=0.000).

In "usage of social apps," 6 behavioral features had significant results. Specifically, in terms of "usage of social apps all day," the main effect of depressive symptom was significant ($F_{1,116}$=6.801, *P*=.01, η^2=0.055). The frequency of usage of social apps was significantly higher in the depressed group than in the nondepressed group (*P*=.01). By contrast, the main effect of gender and the interaction effect of depressive symptom and gender were not significant ($F_{1,116}$=0.675, *P*=.41, η^2=0.006; $F_{1,116}$=1.654, *P*=.20, η^2=0.014). As for "usage of social apps in the evening," the main effect of depressive symptom was significant ($F_{1,116}$=6.902, *P*=.01, η^2=0.056). People in the depressed group used social apps more frequently than people in the nondepressed group in the evening. However, the main effect of gender and the interaction effect of depressive symptom and gender were not significant ($F_{1,116}$=0.095, *P*=.76, η^2=0.001; $F_{1,116}$=1.964, *P*=.16, η^2=0.017).

Table 3. Differences in smartphone usage among participants with different depressive symptoms.[a]

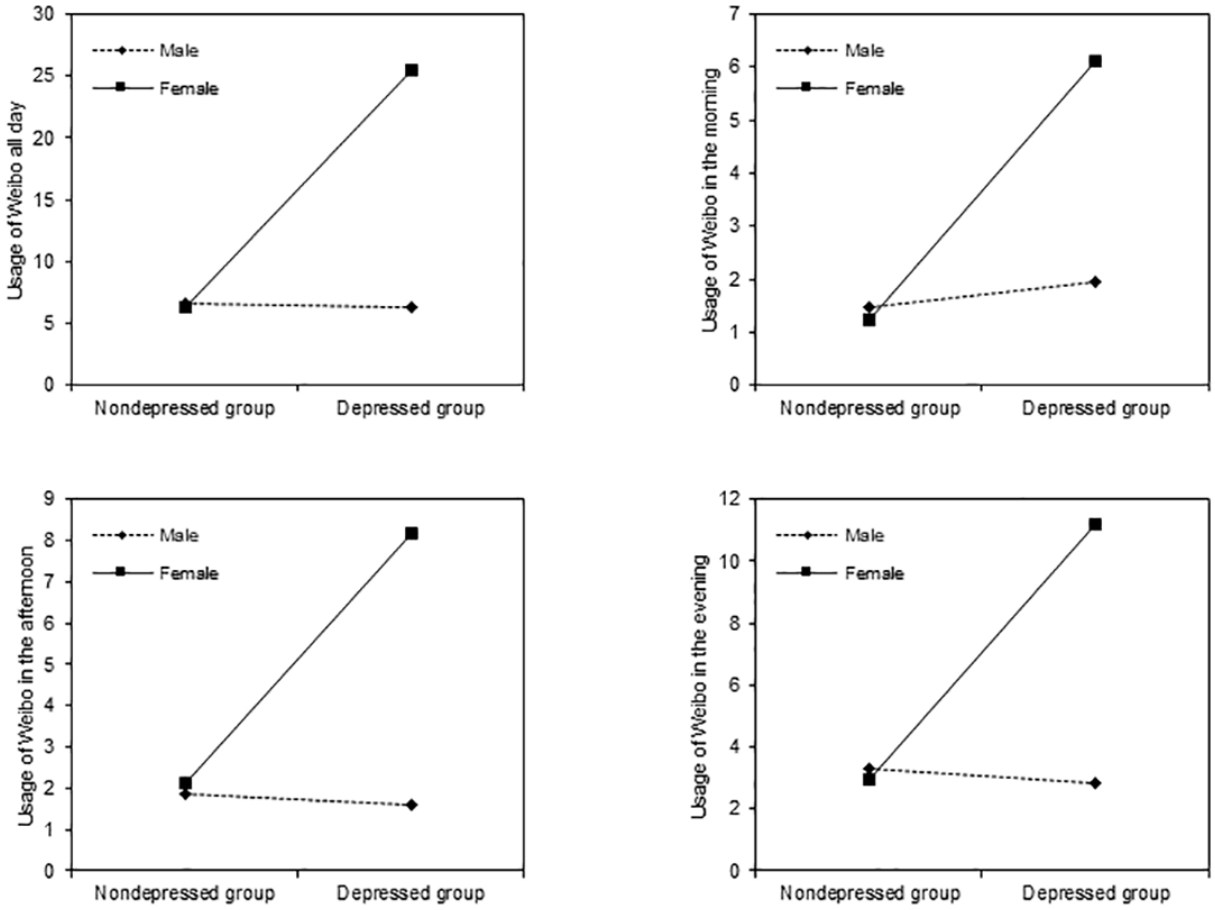
Behavioral feature	Nondepressed group		Depressed group		F_{group} (*df*)	P_{group} value	F_{gender} (*df*)	P_{gender} value	$F_{interaction}$ (*df*)	$P_{interaction}$ value
	Male	Female	Male	Female						
Traditional communication behavior, mean (SD)										
Receiving calls from contacts all day	0.49 (1.38)	0.48 (1.61)	0.05 (0.20)	0.09 (0.40)	3.995 (1, 116)	.048	0.005 (1, 116)	.94	0.010 (1, 116)	.92
Receiving calls from contacts in the afternoon	0.20 (0.53)	0.20 (0.64)	0.01 (0.02)	0.02 (0.07)	5.278 (1, 116)	.02	0.004 (1, 116)	.95	0.006 (1, 116)	.94
Usage of social apps, mean (SD)										
Usage of social apps all day	62.34 (114.24)	73.84 (107.88)	158.72 (172.41)	105.56 (120.28)	6.801 (1, 116)	.01	0.675 (1, 116)	.41	1.654 (1, 116)	.20
Usage of social apps in the evening	26.30 (40.08)	42.61 (69.62)	86.42 (115.56)	60.90 (88.05)	6.902 (1, 116)	.01	0.095 (1, 116)	.76	1.964 (1, 116)	.16
Usage of Weibo all day	6.64 (14.79)	6.31 (14.79)	6.34 (13.08)	25.49 (34.78)	6.689 (1, 116)	.01	6.642 (1, 116)	.01	7.118 (1, 116)	.009
Usage of Weibo in the morning	1.46 (3.27)	1.23 (2.90)	1.94 (4.40)	6.11 (8.58)	8.838 (1, 116)	.004	4.765 (1, 116)	.03	5.932 (1, 116)	.02
Usage of Weibo in the afternoon	1.88 (4.59)	2.11 (5.57)	1.58 (3.17)	8.19 (10.81)	6.370 (1, 116)	.01	8.918 (1, 116)	.003	7.764 (1, 116)	.006
Usage of Weibo in the evening	3.30 (7.40)	2.97 (6.73)	2.82 (5.97)	11.19 (17.93)	4.531 (1, 116)	.04	4.889 (1, 116)	.03	5.731 (1, 116)	.02

[a]Data for descriptive statistics are the mean (SD) of the frequency of social behavior on smartphone. *F* values of 2-way ANOVA are represented by F_{group}, F_{gender}, and $F_{interaction}$. *P* values of 2-way ANOVA are represented by P_{group} value, P_{gender} value, and $P_{interaction}$ value.

In addition, Weibo was the only app with significant results among the 2 popular social apps. Under "usage of Weibo all day," the main effect of depressive symptom, the main effect of gender, and the interaction effect of depressive symptom and gender were significant ($F_{1,116}$=6.689, *P*=.01, η^2=0.055; $F_{1,116}$=6.642, *P*=.01, η^2=0.054; $F_{1,116}$=7.118, *P*=.009, η^2=0.058, respectively). Simple effect analyses showed that in the depressed group, females used Weibo more frequently all day than males ($F_{1,116}$=11.744, *P*=.001, η^2=0.092; Figure 1). However, in the nondepressed group, there were no significant differences ($F_{1,116}$=0.005, *P*=.94, η^2=0.000). As for "usage of Weibo in the morning," the main effect of depressive symptom, the main effect of gender, and the interaction effect of depressive symptom and gender were significant ($F_{1,116}$=8.838, *P*=.004, η^2=0.071; $F_{1,116}$=4.765, *P*=.03, η^2=0.039; $F_{1,116}$=5.932, *P*=.02, η^2=0.049, respectively). Simple effect analyses showed that in the depressed group, the frequency of usage of Weibo among females in the morning was significantly higher than that of males ($F_{1,116}$=9.105, *P*=.003, η^2=0.073; Figure 1). However, in the nondepressed group, no significant gender differences existed ($F_{1,116}$=0.039, *P*=.85, η^2=0.000).

Concerning "usage of Weibo in the afternoon," the main effect of depressive symptom, the main effect of gender, and the interaction effect of depressive symptom and gender were significant ($F_{1,116}$=6.370, *P*=.01, η^2=0.052; $F_{1,116}$=8.918, *P*=.003, η^2=0.071; $F_{1,116}$=7.764, *P*=.006, η^2=0.063, respectively). Simple effect analyses showed that in the depressed group, females used Weibo more frequently in the afternoon than males ($F_{1,116}$=14.224, *P*<.001, η^2=0.109; Figure 1). However, in the nondepressed group, no significant gender differences existed ($F_{1,116}$=0.024, *P*=.88, η^2=0.000). In terms of "usage of Weibo in the evening," the main effect of gender and the interaction effect of depressive symptom and gender were not significant ($F_{1,116}$=4.531, *P*=.04, η^2=0.038; $F_{1,116}$=4.889, *P*=.03, η^2=0.040; $F_{1,116}$=5.731, *P*=.02, η^2=0.047, respectively). Simple effect analyses showed that in the depressed group, females used Weibo more frequently than males in the evening ($F_{1,116}$=9.052, *P*=.003, η^2=0.072; Figure 1). However, in the nondepressed group, no significant gender differences existed ($F_{1,116}$=.020, *P*=.89, η^2=0.000).

Figure 1. Interactions between depressive symptom and gender on usage of Weibo in different periods.



Predictors of Depressive Symptoms

The following 11 variables were entered into the multiple logistic regression model: “receiving calls from contacts all day,” “receiving calls from contacts in the afternoon,” “receiving calls from contacts in the evening,” “usage of social apps all day,” “usage of social apps in the morning,” “usage of social apps in the evening,” “usage of Weibo all day,” “usage of Weibo in the morning,” “usage of Weibo in the afternoon,” “usage of Weibo in the evening,” and “usage of WeChat in the evening.” Of the 11 variables, “usage of social apps in the evening” remained a significant independent predictor (*P*=.02) of depressive symptoms in the model established on the whole sample (Table 4). In addition, “usage of social apps in the evening” emerged as a significant factor for male (*P*=.01), and “usage of Weibo in the morning” emerged as a significant factor for female (*P*=.03).

Table 4. Multiple logistic models of depressive symptoms.

Behavioral feature	Odds ratio (95% CI)	*P* value
Total		
Usage of Weibo in the morning	1.089 (0.996-1.191)	.06
Usage of social apps in the evening	1.007 (1.001-1.013)	.02
Receiving calls from contacts in the afternoon	0.032 (0.000-2.705)	.13
Male		
Usage of social apps in the evening	1.013 (1.003-1.022)	.01
Female		
Usage of Weibo in the morning	1.183 (1.015-1.378)	.03

Discussion

Principal Findings

Our results supported the hypothesis that nondepressed and depressed users have differences in social behaviors on smartphones, and there were gender differences in social behavior among depressed users. Specifically, we used MobileSens to obtain metadata of smartphone usage on 120 participants, and extracted social behavior features to investigate the social behavior of nondepressed and depressed users. The results found depressed users received less calls from contacts (all day, in the afternoon) and used social apps more frequently (all day, in the evening) than nondepressed users. In the depressed group, females used Weibo more frequently than males (all day, in the morning, in the afternoon, in the evening).

Traditional Communication Behavior of Depressed Users

The results show that depressed users receive fewer calls from contacts throughout the day than nondepressed users. Consistent with our results, a cross-sectional study involving 6105 adults aged 18-84 years found that low frequency of contacts with friends and parents was a significant factor associated with depression [62]. In particular, social interactions with strong ties (eg, family, friends) play an important role in satisfying people's need for social connectedness [63-66]. Furthermore, meta-analyses have also shown that interventions for depression by addressing social relationships, including couples therapy and peer support, may be effective in reducing depressive symptoms [67,68]. Therefore, it is plausible that social interactions with strong ties via calls can satisfy psychological needs essential for reduction of depression and human flourishing. In addition, the result can be supplemented by one of the major assumptions of the behavioral theory of depression, namely, that a low rate of response-contingent positive reinforcement acts as an eliciting stimulus for depression [69-71]. Response-contingent positive reinforcement is defined as pleasurable or positive outcomes that follow an individual's behavior within his or her environment and increase the likelihood of those behaviors. Under such an assumption, social interactions with friends through calls could be perceived as a positive response contingent, whereas lack of social interaction with friends or family will increase depression.

Notably, only in the afternoon did the 2 groups differ significantly in the frequency of calls received from contacts over all periods (P=.02). Previous studies have found a circadian rhythm of positive affect in healthy individuals, peaking in the afternoon [72,73]. Moreover, this positive affect was positively related to the number of interactions in which participants engaged, and the amount of time spent engaged in social contact [74]. According to the social enhancement model [75], those who already have good mental state enjoy more benefits from social interactions, while problems are only compounded for those who have psychological deficits. Thus, people who are already in good mental states have more willingness to engage in social interactions, which may encourage others to call them. However, unlike healthy individuals, the diurnal variation of positive affect in depressed patients showed an inverse-U shape with a steeper overall increase from morning lows to evening highs [76]. Thus, it is not surprising that difference exists in the number of receiving calls in the afternoon between the 2 groups.

Usage of Social Apps of Depressed Users

We found that depressed users use social apps throughout the day more frequently than nondepressed ones. There might be several reasons. First, depressed people expect that interpersonal interaction via the internet can alleviate their psychological problems, because they believe that it is less risky and easier to get support and build relationships online than face to face [77,78]. Given that 99.1% of internet users are mobile internet users in China [41], the positive association between depression and internet overuse would be replicated for a smartphone. Second, deficiency in self-regulation is a common manifestation among those with depression [79], which makes it difficult for them to maintain a healthy level and amount of smartphone use. Therefore, depressed users use social apps more frequently than nondepressed ones.

As for "usage of social apps in the evening," we found that it is a predictor for depression on both the overall data (all participants) and male data (male participants), and the usage frequency of depressed users is higher than nondepressed ones. Previous studies have found that depression symptoms tend to be relatively more active during the evening and night [59,80-82], which may exacerbate the usage of social apps. Additionally, people prefer to spend time on social media rather than face-to-face activities [83-85], but the quality of communication over social media has been highlighted as a potential limiting factor in building strong, emotionally intense relationships [86-88]. Furthermore, frequent smartphone usage can make people vulnerable to negative outcomes, such as interpersonal isolation [89]. Therefore, the usage of social apps not only fails to lessen their negative emotions, but also actually worsens these outcomes.

In the usage of top 2 social apps, at any time, the depressed ones used Weibo more frequently than the nondepressed ones. In the depressed group, females used Weibo more frequently than males. As we discussed above, the depressed ones use social apps more than the nondepressed ones, so this conclusion applies equally to Weibo. What deserves our attention is the gender differences in Weibo usage in the depressed group. Similar to Twitter, but unlike WeChat, Weibo focuses on sharing of opinions and information exchange rather than on social interaction [90], and offers some anonymity in online communication [91]. A previous study found that females are more likely to report Weibo usage for communicating with peers, passing time, and entertainment, whereas male users are more likely than females to report Weibo use for social compensation and social identity gratifications [92]. So depressed females are more likely to overuse Weibo than depressed males. Given a worsening of symptoms in the morning hours and early morning waking are characteristics of depression [58], depressed females may try to eliminate psychological problems by using Weibo. This could explain the usage of Weibo in the morning, which could serve as a predictor of depression in females.

Limitations and Future Work

One limitation of this study is that we only use questionnaire-based scales to measure depression. Although the validity of the questionnaires as a screening tool in accessing depression severity has been well proven in the literature, the questionnaire score itself cannot be used as diagnostic. Second, the sample in this study was composed of healthy individuals rather than clinical patients, which means that there were few "real" patients with depression disorders in our sample. So our results are applicable to individuals with minor depression, and their applicability to patients with confirmed moderate to severe depression has not been verified. Moreover, given the convenience sampling method of this study and the majority of participants were students, the extent to which sampling bias affected the results cannot be determined. Therefore, replication and extension of this research using larger, more representative samples are desirable in future work.

Although the results found that depressed users used social apps more frequently than nondepressed ones, the psychological mechanism behind it is still unclear. For instance, because deficiency in self-regulation is a common manifestation among those with depression [79], future work could explore the mediating effect of self-regulation in the relationship between depression and social behavior on smartphones. Additionally, our results showed that compared with nondepressed users, depressed users received fewer calls from acquaintances. Future studies may need to pay closer attention to whether it is the reason that depressed users use social apps frequently. Furthermore, our result found gender differences in some social behaviors among depressed users (eg, the usage of Weibo), which suggests that social-based features should be treated as a personalized feature that should be assessed in a within-subject analysis in future studies.

Conclusions

In summary, by analyzing metadata of smartphone usage acquired by MobileSens, we found that relationships exist between depression and social behavior on smartphones, which means we may be able to implement an early diagnosis system of depression through smartphone usage. Besides, the results of this study provide useful suggestions to those who are depressive in daily lives.

Acknowledgments

This study was partially supported by the Key Research Program of the Chinese Academy of Sciences (No. ZDRW-XH-2019-4). The authors thank everyone who participated in the study.

Conflicts of Interest

None declared.

References

1. Friedrich M. Depression Is the Leading Cause of Disability Around the World. JAMA 2017 Apr 18;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28418490]
2. Lewinsohn PM, Solomon A, Seeley JR, Zeiss A. Clinical implications of "subthreshold" depressive symptoms. J Abnorm Psychol 2000 May;109(2):345-351. [Medline: 10895574]
3. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life: the Heart and Soul Study. JAMA 2003 Jul 09;290(2):215-221 [FREE Full text] [doi: 10.1001/jama.290.2.215] [Medline: 12851276]
4. Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity? J Occup Environ Med 2008 Apr;50(4):401-410. [doi: 10.1097/JOM.0b013e31816bae50] [Medline: 18404013]
5. Musselman DL, Evans DL, Nemeroff CB. The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. Arch Gen Psychiatry 1998 Jul 01;55(7):580-592. [doi: 10.1001/archpsyc.55.7.580] [Medline: 9672048]
6. Wragg RE, Jeste DV. Overview of depression and psychosis in Alzheimer's disease. Am J Psychiatry 1989 May;146(5):577-587. [doi: 10.1176/ajp.146.5.577] [Medline: 2653053]
7. Minkoff K, Bergman E, Beck AT, Beck R. Hopelessness, depression, and attempted suicide. Am J Psychiatry 1973 Apr;130(4):455-459. [doi: 10.1176/ajp.130.4.455] [Medline: 4691303]
8. Hawton K, Casañas ICC, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord 2013 May;147(1-3):17-28. [doi: 10.1016/j.jad.2013.01.004] [Medline: 23411024]
9. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
10. Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. Arch Gen Psychiatry 2001 Jan 01;58(1):55-61. [doi: 10.1001/archpsyc.58.1.55] [Medline: 11146758]
11. Harman JS, Edlund MJ, Fortney JC. Disparities in the adequacy of depression treatment in the United States. Psychiatr Serv 2004 Dec;55(12):1379-1385. [doi: 10.1176/appi.ps.55.12.1379] [Medline: 15572565]
12. Simon G, VonKorff M. Recognition, management, and outcomes of depression in primary care. Arch Fam Med 1995 Feb;4(2):99-105. [doi: 10.1001/archfami.4.2.99] [Medline: 7842160]

13. van Weel C, van Weel-Baumgarten E, van Rijswijk E. Treatment of depression in primary care. BMJ 2009 Mar 19;338(mar19 1):b934. [doi: 10.1136/bmj.b934] [Medline: 19299476]
14. Schulberg HC, Saul M, McClelland M, Ganguli M, Christy W, Frank R. Assessing depression in primary medical and psychiatric practices. Arch Gen Psychiatry 1985 Dec 01;42(12):1164-1170. [doi: 10.1001/archpsyc.1985.01790350038008] [Medline: 4074109]
15. Wells KB, Hays RD, Burnam MA, Rogers W, Greenfield S, Ware JE. Detection of depressive disorder for patients receiving prepaid or fee-for-service care. Results from the Medical Outcomes Study. JAMA 1989 Dec 15;262(23):3298-3302. [doi: 10.1001/jama.262.23.3298] [Medline: 2585674]
16. Kamphuis MH, Stegenga BT, Zuithoff NPA, King M, Nazareth I, de Wit NJ, et al. Does recognition of depression in primary care affect outcome? The PREDICT-NL study. Fam Pract 2012 Feb 22;29(1):16-23. [doi: 10.1093/fampra/cmr049] [Medline: 21859837]
17. Kemp S. DIGITAL 2019: Global Internet Use Accelerates. 2019 Jan. URL: https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates [accessed 2020-03-01]
18. Kim M, Kim H, Kim K, Ju S, Choi J, Mi Y. Smartphone Addiction: (Focused Depression, Aggression and Impulsion) among College Students. Indian Journal of Science and Technology 2015;8(25):1-8 [FREE Full text] [doi: 10.17485/ijst/2015/v8i25/80215]
19. Matar Boumosleh J, Jaalouk D. Depression, anxiety, and smartphone addiction in university students- A cross sectional study. PLoS One 2017 Aug 4;12(8):e0182239 [FREE Full text] [doi: 10.1371/journal.pone.0182239] [Medline: 28777828]
20. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
21. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: A systematic review and meta-analysis. Psychiatry Res 2020 Feb;284:112686. [doi: 10.1016/j.psychres.2019.112686] [Medline: 31757638]
22. Kim J, Seo M, David P. Alleviating depression only to become problematic mobile phone users: Can face-to-face communication be the antidote? Computers in Human Behavior 2015 Oct;51:440-447. [doi: 10.1016/j.chb.2015.05.030]
23. Alter A. Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked. London, England: Penguin; 2017.
24. Misra S, Cheng L, Genevie J, Yuan M. The iPhone Effect: The Quality of In-Person Social Interactions in the Presence of Mobile Devices. Environment and Behavior 2014 Jul;48(2):275-298. [doi: 10.1177/0013916514539755]
25. Przybylski AK, Weinstein N. A Large-Scale Test of the Goldilocks Hypothesis. Psychol Sci 2017 Feb;28(2):204-215. [doi: 10.1177/0956797616678438] [Medline: 28085574]
26. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. J Am Med Inform Assoc 2020 Apr 01;27(4):522-530. [doi: 10.1093/jamia/ocz221] [Medline: 31977041]
27. Višnjić A, Veličković V, Sokolović D, Stanković M, Mijatović K, Stojanović M, et al. Relationship between the Manner of Mobile Phone Use and Depression, Anxiety, and Stress in University Students. Int J Environ Res Public Health 2018 Apr 08;15(4):697 [FREE Full text] [doi: 10.3390/ijerph15040697] [Medline: 29642471]
28. Thomée S, Dellve L, Härenstam A, Hagberg M. Perceived connections between information and communication technology use and mental symptoms among young adults - a qualitative study. BMC Public Health 2010 Feb 12;10(1):66 [FREE Full text] [doi: 10.1186/1471-2458-10-66] [Medline: 20152023]
29. Tamura H, Nishida T, Tsuji A, Sakakibara H. Association between Excessive Use of Mobile Phone and Insomnia and Depression among Japanese Adolescents. Int J Environ Res Public Health 2017 Jun 29;14(7):701 [FREE Full text] [doi: 10.3390/ijerph14070701] [Medline: 28661428]
30. Boase J, Ling R. Measuring Mobile Phone Use: Self-Report Versus Log Data. J Comput-Mediat Comm 2013 Jun 10;18(4):508-519. [doi: 10.1111/jcc4.12021]
31. Kobayashi T, Boase J. No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use. Communication Methods and Measures 2012 Apr;6(2):126-143. [doi: 10.1080/19312458.2012.679243]
32. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
33. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
34. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
35. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]

36. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE, et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. Int J Methods Psychiatr Res 2016 Dec 01;25(4):309-323 [FREE Full text] [doi: 10.1002/mpr.1502] [Medline: 27038019]
37. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. New York, NY: Association for Computing Machinery; 2015 Sep Presented at: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 2015; Osaka, Japan. [doi: 10.1145/2750858.2805845]
38. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
39. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
40. Baker DA, Algorta GP. The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies. Cyberpsychol Behav Soc Netw 2016 Nov;19(11):638-648. [doi: 10.1089/cyber.2016.0206] [Medline: 27732062]
41. China Internet Network Information Center. The 44th China Statistical Report on Internet Development Dictionary. 2019 Aug. URL: http://www.cac.gov.cn/2019-08/30/c_1124938750.htm [accessed 2020-03-01]
42. Bebbington P, Brugha T, Meltzer H, Farrell M, Ceresa C, Jenkins R, et al. Psychiatric disorder and dysfunction in the UK National Survey of Psychiatric Morbidity. Soc Psychiatry Psychiatr Epidemiol 2000 May 16;35(5):191-197. [doi: 10.1007/s001270050227] [Medline: 10941993]
43. Wilhelm K, Parker G, Asghari A. Sex differences in the experience of depressed mood state over fifteen years. Soc Psychiatry Psychiatr Epidemiol 1998 Jan 9;33(1):16-20. [doi: 10.1007/s001270050016] [Medline: 9448440]
44. Bianchi A, Phillips JG. Psychological predictors of problem mobile phone use. Cyberpsychol Behav 2005 Feb;8(1):39-51. [doi: 10.1089/cpb.2005.8.39] [Medline: 15738692]
45. Toda M, Monden K, Kubo K, Morimoto K. MOBILE PHONE DEPENDENCE AND HEALTH-RELATED LIFESTYLE OF UNIVERSITY STUDENTS. Soc Behav Pers 2006 Jan 01;34(10):1277-1284. [doi: 10.2224/sbp.2006.34.10.1277]
46. Billieux J, Van der Linden M, Rochat L. The role of impulsivity in actual and problematic use of the mobile phone. Appl. Cognit. Psychol 2008 Dec;22(9):1195-1210. [doi: 10.1002/acp.1429]
47. Igarashi T, Takai J, Yoshida T. Gender differences in social network development via mobile phone text messages: A longitudinal study. Journal of Social and Personal Relationships 2016 Jun 30;22(5):691-713. [doi: 10.1177/0265407505056492]
48. Jenaro C, Flores N, Gómez-Vela M, González-Gil F, Caballo C. Problematic internet and cell-phone use: Psychological, behavioral, and health correlates. Addiction Research & Theory 2009 Jul 11;15(3):309-320. [doi: 10.1080/16066350701350247]
49. Sánchez-Martínez M, Otero A. Factors associated with cell phone use in adolescents in the community of Madrid (Spain). Cyberpsychol Behav 2009 Apr;12(2):131-137. [doi: 10.1089/cpb.2008.0164] [Medline: 19072078]
50. Guo R, Zhu T, Wang Y, Xu X. MobileSens: A framework of behavior logger on Andriod mobile device. New York, NY: IEEE; 2011 Presented at: 2011 6th International Conference on Pervasive Computing and Applications; October 2011; Port Elizabeth, South Africa p. 281-286. [doi: 10.1109/ICPCA.2011.6106518]
51. Li A, Li H, Guo R, Zhu T. MobileSens: a ubiquitous psychological laboratory based on mobile device. International Journal of Cyber Behavior, Psychology and Learning (IJCBPL) 2013;3(2):47-55. [doi: 10.4018/ijcbpl.2013040104]
52. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
53. Roberts RE. Reliability of the CES-D scale in different ethnic contexts. Psychiatry Research 1980 May;2(2):125-134. [doi: 10.1016/0165-1781(80)90069-4]
54. Jie Z, Zhen-Yun W, Ge F, Juan L, Bu-Xin H, Zhi-Yan C. Development of the Chinesr age norms of CES-D in urban area. Chinese Mental Health Journal 2010 Apr 13;24(2):139-143. [doi: 10.3969/j.issn.1000-6729.2010.02.015]
55. Salgado-Delgado R, Tapia Osorio A, Saderi N, Escobar C. Disruption of circadian rhythms: a crucial factor in the etiology of depression. Depress Res Treat 2011;2011:839743 [FREE Full text] [doi: 10.1155/2011/839743] [Medline: 21845223]
56. Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. Hum Psychopharmacol 2008 Oct;23(7):571-585 [FREE Full text] [doi: 10.1002/hup.964] [Medline: 18680211]
57. Goodwin F, Jamison K. Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, Second Edition. Oxford, UK: Oxford University Press; 2007.
58. Battle DE. Diagnostic and Statistical Manual of Mental Disorders (DSM). Codas 2013;25(2):191-192 [FREE Full text] [doi: 10.1590/s2317-17822013000200017] [Medline: 24413388]
59. Gordijn M, Beersma D, Bouhuys A, Reinink E, Van den Hoofdakker R. A longitudinal study of diurnal mood variation in depression; characteristics and significance. Journal of Affective Disorders 1994 Aug;31(4):261-273. [doi: 10.1016/0165-0327(94)90102-3]

60. Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. Int J Psychiatry Med 1990;20(3):259-277. [doi: 10.2190/LYKR-7VHP-YJEM-MKM2] [Medline: 2265888]
61. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. Perspect Clin Res 2017;8(3):148-151 [FREE Full text] [doi: 10.4103/picr.PICR_87_17] [Medline: 28828311]
62. Kleinberg A, Aluoja A, Vasar V. Social support in depression: structural and functional factors, perceived control and help-seeking. Epidemiol Psychiatr Sci 2013 Sep 24;22(4):345-353. [doi: 10.1017/s2045796013000504]
63. Mehl MR, Vazire S, Holleran SE, Clark CS. Eavesdropping on happiness: well-being is related to having less small talk and more substantive conversations. Psychol Sci 2010 Apr 18;21(4):539-541 [FREE Full text] [doi: 10.1177/0956797610362675] [Medline: 20424097]
64. Reis HT, Sheldon KM, Gable SL, Roscoe J, Ryan RM. Daily Well-Being: The Role of Autonomy, Competence, and Relatedness. Pers Soc Psychol Bull 2016 Jul 02;26(4):419-435. [doi: 10.1177/0146167200266002]
65. Vittengl JR, Holt CS. Positive and Negative Affect in Social Interactions as a Function of Partner Familiarity, Quality of Communication, and Social Anxiety. Journal of Social and Clinical Psychology 1998 Jun;17(2):196-208. [doi: 10.1521/jscp.1998.17.2.196]
66. Wheeler L, Reis H, Nezlek J. Loneliness, social interaction, and sex roles. J Pers Soc Psychol 1983 Oct;45(4):943-953. [doi: 10.1037//0022-3514.45.4.943] [Medline: 6631669]
67. Pfeiffer PN, Heisler M, Piette JD, Rogers MA, Valenstein M. Efficacy of peer support interventions for depression: a meta-analysis. Gen Hosp Psychiatry 2011 Jan;33(1):29-36 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.10.002] [Medline: 21353125]
68. Barbato A, D'Avanzo B. Efficacy of couple therapy as a treatment for depression: a meta-analysis. Psychiatr Q 2008 Jun 8;79(2):121-132. [doi: 10.1007/s11126-008-9068-0] [Medline: 18259866]
69. Lewinsohn P. A Behavioral Approach to Depression: Essential Papers on Depression. Manhattan, NY: NYU Press; 1974:150-172.
70. Lewinsohn PM, Graf M. Pleasant activities and depression. J Consult Clin Psychol 1973 Oct;41(2):261-268. [doi: 10.1037/h0035142] [Medline: 4147832]
71. Lewinsohn PM, Libet J. Pleasant events, activity schedules, and depressions. J Abnorm Psychol 1972 Jun;79(3):291-295. [doi: 10.1037/h0033207] [Medline: 5033370]
72. Boivin DB, Czeisler CA, Dijk DJ, Duffy JF, Folkard S, Minors DS, et al. Complex interaction of the sleep-wake cycle and circadian phase modulates mood in healthy subjects. Arch Gen Psychiatry 1997 Feb 01;54(2):145-152. [doi: 10.1001/archpsyc.1997.01830140055010] [Medline: 9040282]
73. Murray G, Nicholas CL, Kleiman J, Dwyer R, Carrington MJ, Allen NB, et al. Nature's clocks and human mood: the circadian system modulates reward motivation. Emotion 2009 Oct;9(5):705-716. [doi: 10.1037/a0017080] [Medline: 19803592]
74. Berry DS, Hansen JS. Positive affect, negative affect, and social interaction. Journal of Personality and Social Psychology 1996 Oct;71(4):796-809. [doi: 10.1037/0022-3514.71.4.796]
75. Kraut R, Kiesler S, Boneva B, Cummings J, Helgeson V, Crawford A. Internet Paradox Revisited. J Social Isssues 2002 Jan;58(1):49-74. [doi: 10.1111/1540-4560.00248]
76. Crowe E, Daly M, Delaney L, Carroll S, Malone KM. The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in Major Depression. Psychiatry Res 2019 Sep;279:98-108. [doi: 10.1016/j.psychres.2018.02.032] [Medline: 29661498]
77. Morahan-Martin J. Internet use and abuse and psychological problems. In: Adam J, Katelyn M, Tom P, Ulf-Dietrich R, editors. Oxford Handbook of Internet Psychology. Oxford, UK: Oxford University Press; 2007.
78. Morahan-Martin J, Schumacher P. Incidence and correlates of pathological Internet use among college students. Computers in Human Behavior 2000 Jan;16(1):13-29. [doi: 10.1016/s0747-5632(99)00049-7]
79. Bandura A, Pastorelli C, Barbaranelli C, Caprara GV. Self-efficacy pathways to childhood depression. Journal of Personality and Social Psychology 1999;76(2):258-269. [doi: 10.1037/0022-3514.76.2.258]
80. Rusting CL, Larsen RJ. Diurnal patterns of unpleasant mood: associations with neuroticism, depression, and anxiety. J Pers 1998 Feb 28;66(1):85-103. [doi: 10.1111/1467-6494.00004] [Medline: 9457771]
81. Lustberg L, Reynolds CF. Depression and insomnia: questions of cause and effect. Sleep Med Rev 2000 Jun;4(3):253-262. [doi: 10.1053/smrv.1999.0075] [Medline: 12531168]
82. Gellman M, Turner J. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013.
83. NIE NH. Sociability, Interpersonal Relations, and the Internet. American Behavioral Scientist 2016 Jul 27;45(3):420-435. [doi: 10.1177/00027640121957277]
84. Mesch GS. Family Relations and the Internet: Exploring a Family Boundaries Approach. Journal of Family Communication 2006 Apr;6(2):119-138. [doi: 10.1207/s15327698jfc0602_2]
85. Nie N, Hillygus D. The impact of Internet use on sociability: Time-diary findings. It & Society 2002;1(1):1-20 [FREE Full text]
86. Schiffrin H, Edelman A, Falkenstern M, Stewart C. The associations among computer-mediated communication, relationships, and well-being. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):299-306. [doi: 10.1089/cyber.2009.0173] [Medline: 20557249]

87. Cummings JN, Butler B, Kraut R. The quality of online social relationships. Commun. ACM 2002 Jul;45(7):103-108. [doi: 10.1145/514236.514242]
88. Gross EF, Juvonen J, Gable SL. Internet Use and Well-Being in Adolescence. J Social Isssues 2002 Jan;58(1):75-90. [doi: 10.1111/1540-4560.00249]
89. Turkle S. Reclaiming Conversation: The Power of Talk in the Digital Age. New York, NY: Penguin Press; Oct 6, 2015.
90. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web. New York, NY: Association for Computing Machinery; 2010 Apr Presented at: WWW '10: The 19th International Conference on World Wide Web; April 2010; Raleigh, NC p. 591-600. [doi: 10.1145/1772690.1772751]
91. Huberman BA, Romero DM, Wu F. Social Networks that Matter: Twitter Under the Microscope. arXiv 2008 [FREE Full text] [doi: 10.2139/ssrn.1313405]
92. Chan M, Wu X, Hao Y, Xi R, Jin T. Microblogging, online expression, and political efficacy among young chinese citizens: the moderating role of information and entertainment needs in the use of weibo. Cyberpsychol Behav Soc Netw 2012 Jul;15(7):345-349. [doi: 10.1089/cyber.2012.0109] [Medline: 22780993]

Abbreviations

CES-D: The Center for Epidemiological Studies-Depression Scale
OR: odds ratio

Edited by L Buis; submitted 02.04.20; peer-reviewed by JH Kim, L Castro; comments to author 01.09.20; revised version received 26.10.20; accepted 30.11.20; published 06.01.21.

Please cite as:
Wang Y, Ren X, Liu X, Zhu T
Examining the Correlation Between Depression and Social Behavior on Smartphones Through Usage Metadata: Empirical Study
JMIR Mhealth Uhealth 2021;9(1):e19046
URL: https://mhealth.jmir.org/2021/1/e19046
doi:10.2196/19046
PMID:33404512

Original Paper

The Temporal Relationship Between Ecological Pain and Life-Space Mobility in Older Adults With Knee Osteoarthritis: A Smartwatch-Based Demonstration Study

Mamoun T Mardini[1,2], MSc, PhD; Subhash Nerella[3], MSc; Matin Kheirkhahan[4], MSc, PhD; Sanjay Ranka[5], PhD; Roger B Fillingim[6], PhD; Yujie Hu[7], PhD; Duane B Corbett[1], PhD; Erta Cenko[8], MSc; Eric Weber[6], BSc; Parisa Rashidi[3], PhD; Todd M Manini[1], PhD

[1]Department of Aging and Geriatric Research, University of Florida, Gainesville, FL, United States

[2]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

[3]Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States

[4]Google LLC, Mountain View, CA, United States

[5]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States

[6]Department of Community Dentistry and Behavioral Science, University of Florida, Gainesville, FL, United States

[7]Department of Geography, University of Florida, Gainesville, FL, United States

[8]Department of Epidemiology, University of Florida, Gainesville, FL, United States

Corresponding Author:
Mamoun T Mardini, MSc, PhD
Department of Aging and Geriatric Research
University of Florida
2004 Mowry Road
PO Box 112610
Gainesville, FL, 32610
United States
Phone: 1 352 273 8962
Email: malmardini@ufl.edu

Abstract

Background: Older adults who experience pain are more likely to reduce their community and life-space mobility (ie, the usual range of places in an environment in which a person engages). However, there is significant day-to-day variability in pain experiences that offer unique insights into the consequences on life-space mobility, which are not well understood. This variability is complex and cannot be captured with traditional recall-based pain surveys. As a solution, ecological momentary assessments record repeated pain experiences throughout the day in the natural environment.

Objective: The aim of this study was to examine the temporal association between ecological momentary assessments of pain and GPS metrics in older adults with symptomatic knee osteoarthritis by using a smartwatch platform called Real-time Online Assessment and Mobility Monitor.

Methods: Participants (n=19, mean 73.1 years, SD 4.8; female: 13/19, 68%; male: 6/19, 32%) wore a smartwatch for a mean period of 13.16 days (SD 2.94). Participants were prompted in their natural environment about their pain intensity (range 0-10) at random time windows in the morning, afternoon, and evening. GPS coordinates were collected at 15-minute intervals and aggregated each day into excursion, ellipsoid, clustering, and trip frequency features. Pain intensity ratings were averaged across time windows for each day. A random effects model was used to investigate the within and between-person effects.

Results: The daily mean pain intensities reported by participants ranged between 0 and 8 with 40% reporting intensities ≥2. The within-person associations between pain intensity and GPS features were more likely to be statistically significant than those observed between persons. Within-person pain intensity was significantly associated with excursion size, and others (excursion span, total distance, and ellipse major axis) showed a statistical trend (excursion span: P=.08; total distance: P=.07; ellipse major axis: P=.07). Each point increase in the mean pain intensity was associated with a 3.06 km decrease in excursion size, 2.89 km decrease in excursion span, 5.71 km decrease total distance travelled per day, 31.4 km^2 decrease in ellipse area, 0.47 km decrease ellipse minor axis, and 3.64 km decrease in ellipse major axis. While not statistically significant, the point estimates for number

of clusters (P=.73), frequency of trips (P=.81), and homestay (P=.15) were positively associated with pain intensity, and entropy (P=.99) was negatively associated with pain intensity.

Conclusions: In this demonstration study, higher intensity knee pain in older adults was associated with lower life-space mobility. Results demonstrate that a custom-designed smartwatch platform is effective at simultaneously collecting rich information about ecological pain and life-space mobility. Such smart tools are expected to be important for remote health interventions that harness the variability in pain symptoms while understanding their impact on life-space mobility.

(JMIR Mhealth Uhealth 2021;9(1):e19609) doi:10.2196/19609

KEYWORDS

ecological momentary assessment; smartwatch app; life-space mobility; pain; knee osteoarthritis; global positioning system

Introduction

The world population of adults aged 65 years or older is rapidly growing [1]. This phenomenon, unprecedented in history, highlights a need to maintain and promote programs that manage chronic diseases and symptoms causing increased risk of loss of mobility and disability. The National Center for Health Statistics [2] reports that, of adults 65-75 years old, 30% and 14.3% have physical impairments and difficulty walking one quarter-mile (approximately 400 m), respectively. Rates are higher in those older than 75 years—48.6% have physical impairments, and 27.7% have difficulty walking one quarter-mile (approximately 400 m). These impairments have a significant negative impact on life-space mobility—the daily activities and geographical area in which people engage. As a result, many older adults anchor to their houses [3]. Osteoarthritis is the most common age-related joint disease in the United States, affecting over 30 million US adults [4]. Pain associated with osteoarthritis is accompanied by a reduction in daily functioning, limitations in walking, and increased risk of overt disability. Pain experiences have within- and between-person variability due to physiological, medical, behavioral, and environmental differences [5]. This variability is complex and cannot be captured with traditional recall-based pain surveys. As a solution, ecological momentary assessments (EMA) record repeated pain experiences throughout the day in a person's natural environment. It minimizes retrospective [6,7] and recent-experience bias [8,9]. However, there are drawbacks as the EMA tools that utilize paper surveys or dedicated digital boxes tend to be intrusive, cannot be easily customized, and are not wearable. In prior work [10], microinteraction EMAs—in which people are prompted with questions, similar to those of ROAMM, that can be understood at a glance and answered in a few seconds—were developed on smart watches and compared to less frequent EMA prompts on smartphones; researchers found that although prompts on the watch were 8 times more frequent than those on the phone, participants adhered 35% more to microinteraction EMAs on the watch. Participants also responded to EMAs in less time and reported the EMAs to be less distracting on the watch than those on the phone [11]. Therefore, EMAs on a smartwatch might serve as an excellent approach for enhancing adherence.

Mobility within the perspective of life-space can be described as the habitual movement of individuals [12-14]. Life-space mobility includes spatial size and frequency of interaction with the surrounding environment. The construct is influenced by physical function and spatial extent of movement, but also the cognitive, psychological, social, and environmental disposition of an individual. Life-space mobility has been measured using various methods [12,13,15,16]. Life-Space Diary, introduced by May et al [13] in 1985, was the first measure. It asked participants to report daily their zone out of 5 predefined concentric zones, referenced to their bedrooms. Similarly, Life-Space Questionnaire, introduced by Stalvey et al [14] in 1999, consisted of 9 yes or no questions asking whether a participant was in a certain region within their environment in the last 3 days. Life-Space Assessment, introduced by Baker et al [12] in 2003, added another perspective by documenting how far and how often an individual travels to predefined regions within their environment, while also considering any assistance needed during mobility. However, there remain issues with life-space mobility assessment—paper-based and recall of information are an added burden on participants and introduce more challenges related to adherence and recall bias.

The use of personal devices such as smartphones and smartwatches is growing rapidly in both young and older adult population groups. According to the International Data Corporation Worldwide Quarterly Wearable Device Tracker, smartwatches accounted for 44.2% of the wearable market in 2018; this is expected to rise to 47.1% by 2023 [17]. The widespread use of wearables and their high computational and sensory capabilities provide a platform to reach and interact with a large share of population, particularly individuals with medical conditions. This is highly significant due to the ability to monitor individuals continuously and intervene whenever and wherever medical conditions occur [18]. It also opens new opportunities to link complex states in a temporal manner.

In this demonstration study, we used a custom-designed smartwatch platform called Real-time Online Assessment and Mobility Monitor (ROAMM) that synchronizes EMA of pain experiences with GPS data to examine their temporal associations in older adults with symptomatic knee osteoarthritis. We hypothesized that higher pain experiences would be associated with lower life-space mobility features.

Methods

Study Population

This study was approved by the University of Florida institutional review board (UFIRB 201601858), and written informed consent was obtained from all participants. We enrolled 19 older adults. Inclusion criteria were age ≥65 years

and diagnosis of unilateral or bilateral symptomatic knee osteoarthritis. Exclusion criteria were failure or inability to provide informed consent; diagnosis of dementia; and being unable to communicate because of severe hearing loss or speech disorder. A convenience sample was drawn from a population of older adults with knee osteoarthritis. Two participants were not interested in participating after being informed about the study. Each participant received compensation of a US $50 gift card. Table 1 shows the descriptive characteristics of participants.

Table 1. Participants' descriptive characteristics.

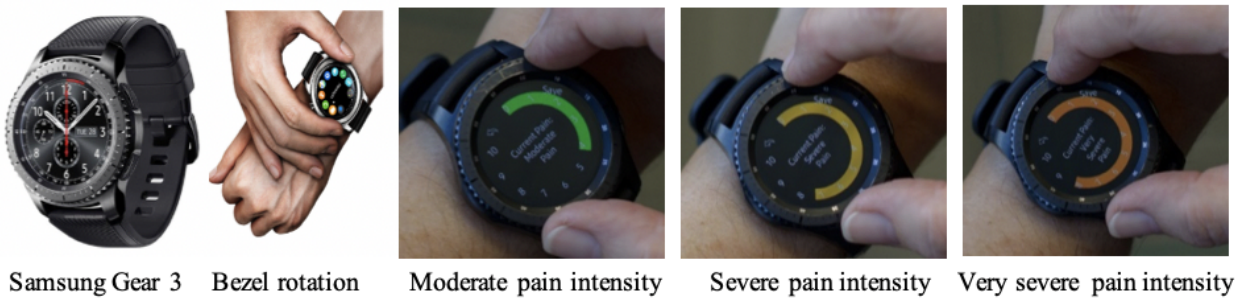
Characteristics	Value (n=19)
Age (years), mean (SD)	73.1 (4.8)
Gender, n (%)	
Male	6 (32)
Female	13 (68)
BMI (kg/m^2), mean (SD)	28.23 (4)
Ethnicity, n (%)	
White	15 (79)
African American	3 (16)
Asian	1 (5)
Education, n (%)	
Graduate	10 (53)
College	6 (32)
High school	2 (10)
Declined to respond	1 (5)
Live alone, n (%)	
Yes	4 (21)
No	15 (79)
Housing, n (%)	
Single Family Home	16 (84)
Other	1 (5)
Other (mobile home, boat)	2 (11)

Ecological Momentary Assessment of Pain Using ROAMM

ROAMM was developed at the University of Florida to enable real-time capture of patient-generated information—wearable sensor data collected simultaneously with symptom EMAs. For this study, EMA of pain was evaluated using the 11-point Box Scale (0=no pain, 1-2=mild pain, 3-5=moderate pain, 6=severe pain, 7-9=very severe pain, 10=worst possible pain), a valid and reliable numerical rating scale [19,20]. Participants were instructed about the anchors.

Participants were prompted about their pain intensity at random times in the morning (8 AM to noon), afternoon (noon to 4 PM), and evening (4 PM to 8 PM). The smartwatch application also captures GPS coordinates (latitude and longitude) every 15 minutes throughout the day. Data were transferred every 15 seconds and stored securely in a remote server. The application interface was developed after holding a focus group as explained by Manini and colleagues [21]. ROAMM architecture is explained in detail in our published papers [22,23] (Figure 1). shows ROAMM app for answering a pain prompt on a smartwatch (Samsung Gear 3, Samsung Group).

Figure 1. ROAMM app on Samsung Gear 3. Participants rotate the bezel on the watch to select the intensity on a scale from 0 to 10 and the color changes accordingly.



GPS Data Collection and Feature Extraction

In this demonstration study, there were some technical difficulties during the initial phase of data collection. These difficulties included weak GPS signal coverages in some places and data transmission problems. The watch required manual checks on mobile networks, roaming, and location services. These issues were discovered and solved during the data collection process. However, data quality checks during the analysis revelated that insufficient GPS data for 9 out of 28 participants. Participants with missing data were similar age (mean 73.3, SD 6.1 years old) and female proportion (7/9, 77.8%) compared to the 19 participants in our paper. Thus, we believe the missing data were randomly lost and did not cause a selection bias.

GPS, the global positioning system, is a navigation utility that furnishes the position of a receiver by measuring its distance from a number of satellites. GPS has been used in the health care domain in behavior [24,25] and gerontological research [26].

Excursion features included excursion size, excursion span, and total distance. Excursion size is the farthest distance an individual travels from home within a specific time window. Excursion span is the farthest distance between all locations away from home. These features provide an individual's travel pattern that can be generally described as (1) compact and away from home; (2) sparse and away from home; (3) compact and close to home; and (4) sparse and close to home (Figure 2). Total distance provides overall view of mobility by summing between all the location points.

Ellipsoid features used a spanning ellipsoid (or ellipsoid hull), which is defined as the minimum area that encompasses all points in 2 dimensions. We used this method to draw an ellipse such that all GPS coordinates lie inside or on the boundary of the ellipse. We aggregated 3 features from the ellipse: (1) ellipse minor axis, which is the shortest diameter passing through the center of the ellipse; (2) ellipse major axis, which is the longest diameter passing through the center of the ellipse, and (3) area of the ellipse (Figure 3).

Figure 2. Illustration of possible travel patterns using excursion size and span features. The solid line represents excursion size, and the dashed line represents excursion span.

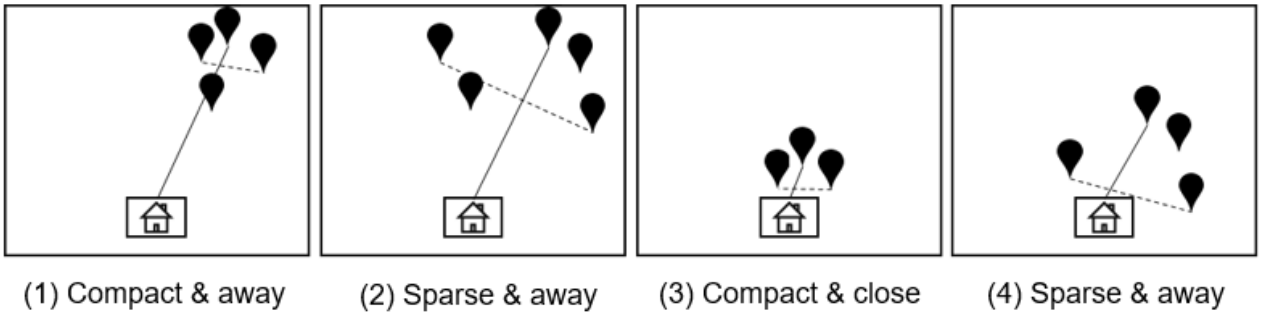
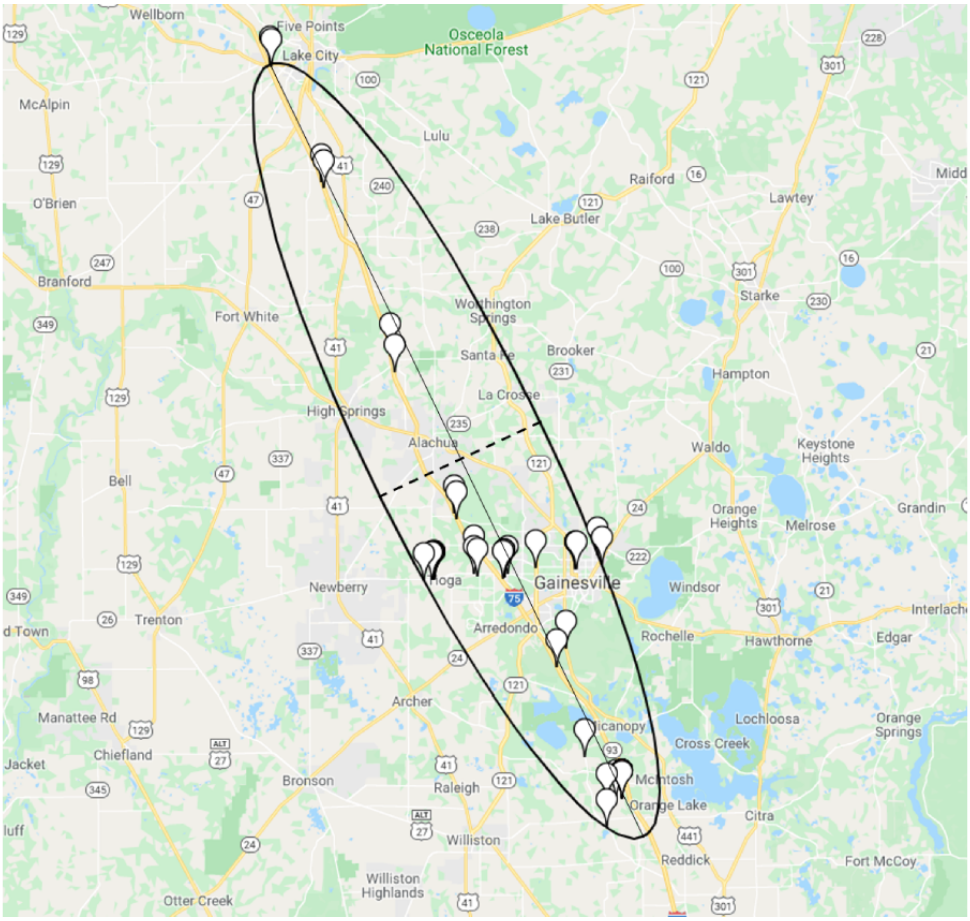


Figure 3. An illustration of the ellipse encompassing all the GPS coordinates for a participant during a 1-day time frame. The dashed line passing through the center represents the ellipse minor axis, and the solid line passing through the center represents the ellipse major axis.



Clustering features provided information on where individuals spend most of their time. This is essential to understand the variability in locations. We used a distance-based clustering mechanism, where nearby locations are clustered together. Each cluster has a centroid, and the distance from the centroid determines the membership of a coordinate in that cluster. We used an adaptive *k*-mean algorithm to cluster locations, which does not require a predefined number of clusters that the conventional *k*-mean algorithm requires. Before providing GPS coordinates to the adaptive *k*-mean clustering algorithm, we classified them into stationary and moving coordinates by calculating the time derivative at each location. When the time derivative was <1 km/h), the GPS coordinate was considered stationary. Only stationary points were considered as input to the clustering algorithm. We ran a simulation to find the optimal number of clusters, with a threshold of 500 m from the cluster's centroid as an inclusion criterion for each cluster. We started with one cluster and gradually increased the number of clusters until all GPS coordinates were assigned to a specific cluster. After clustering all points, we aggregated 2 relevant features: number of clusters and entropy. The number of clusters is simply a count of the generated clusters. The entropy provides information on the distribution of time in different clusters. Entropy measures the degree of disorder or the level of uncertainty in the information theory. In our analysis, entropy was used to measure the level of uncertainty in the time spent in different clusters. Entropy is calculated using the following formula:

where p_i is the percentage of time a participant spends at cluster *i*, and p_i is between (0,1].

A low entropy value means a lower level of uncertainty and that the participant spent most of the time at one location, which is an indication of lower life-space mobility (Figure 4).

Figure 4. Illustration of clustering for a participant's coordinates within a 1-day time frame. A total of 4 clusters are shown on the map, where each cluster contains a collection of GPS coordinates.

Frequency of trips and homestay percentage provided information about the number of trips away from home and time spent at home, respectively. First, we classified the GPS coordinates as home or away from home, then calculated the 2 features accordingly. Homestay is represented as a fraction between 0 and 1. It is the ratio of the number of GPS coordinates within the home radius (ie, 100 m) to the total number of coordinates. Homestay is considered 0 (or 0% when all the GPS coordinates are outside home in a given time period of interest, and 1 (or 100%) when all the GPS coordinates are within the home radius. A trip is calculated when a sequence of GPS coordinates—home, away from home, home—occurs chronologically. The number of trips occurring within a specific time window are summed to yield the frequency of trips.

Statistical Analysis

We evaluated the relationship between EMA of pain (predictor) and the measures of life-space mobility using GPS features (outcomes). Pain intensity ratings were averaged across day-windows. This was done to better connect to the day-based frequency of measurement for the GPS features. In addition to that, we graphically expressed pain intensities into 2 groups: low pain (<2) and high pain (≥2), but statistical comparison was not performed.

A 2-level random effects model (participant and day) was used to account for repeated measurements. The model was fit after disaggregating the within and between-persons effects. Parceling these effects allows a more in-depth understanding about the association between GPS features and EMA pain. The approach used person-mean centering around the grand mean (between-person effect) and the within person effect (each person-specific mean for the time varying covariate) [27-29]. We used the xtcenter command (Stata/MP; version 16.0 for Windows; StataCorp LLC) and entered terms for within-person and between-person effects into the model. The model was also adjusted for age, living alone, and gender covariates as fixed effects. An independent-covariance structure, which was confirmed as the most efficient without loss of model fit using the Akaike information criterion, was used in all models. All analyses were conducted using Stata/MP. Statistical significance was confirmed at the $P \leq .05$ level. Because this study is a demonstration project, $P \leq .10$ was considered as a trend effect.

Results

Participants wore the smartwatch for a mean of 13.16 (SD 2.94) days and responded to a mean of 82% of pain prompts. Figure 5 shows the distribution of reported pain intensities by all participants. A pain intensity rating of 0 was the most common intensity and the highest was 8. The mean pain intensity for the low pain group was 0.26 (SD 0.44) and for the high pain group was 2.78 (SD 0.93). Descriptive characteristics of the life-space mobility features are listed in Table 2. Multimedia Appendix 1 shows that some GPS features were intercorrelated—there were strong correlations between excursion features and ellipsoid features and weak correlations between the remaining features.

Figure 5. Histogram showing pain distribution.

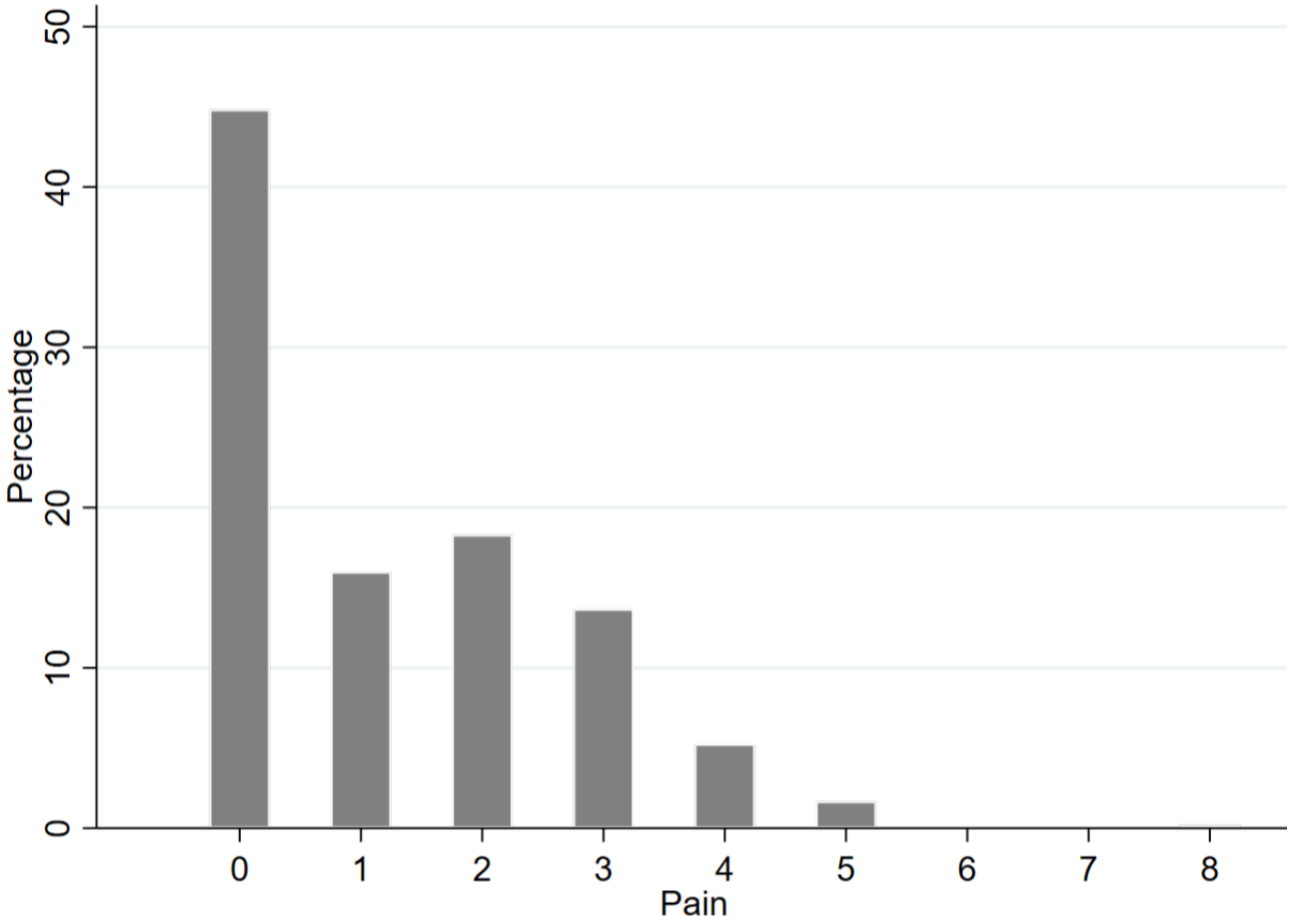


Table 2. Descriptive characteristics of life-space mobility features.

Features	Mean	SD	Median	Kurtosis
Excursion size (km)	11.35	18.97	18.99	17.87
Excursion span (km)	11.33	19.73	4.72	15.34
Total distance (km)	23.14	38.21	8.67	10.43
Ellipse area (km^2)	104.89	342.11	8.00	42.39
Ellipse minor axis (km)	3.41	4.65	1.49	5.87
Ellipse major axis (km)	14.59	23.42	7.34	16.75
Frequency of trips	2.88	4.20	2	8.46
Homestay percentage	0.66	0.31	0.74	–0.31
Number of clusters	2.02	1.08	2	0.86
Entropy	0.30	0.34	0.20	0.05

The median and mean of the daily GPS features of the low pain group were generally higher than those of the high pain group for excursion features (Figure 6) and ellipsoid features (Figure 7). The results of the mixed-effect model are shown in Table 3. There were no between-person effects of pain intensity on GPS features (Figure 8 and Figure 9), but within-persons association were predominant. The majority of GPS features (7 out of 10) indicated that having high pain was associated with a lower value (ie, life-space mobility). Among these GPS features, within-person pain intensity was significantly associated with excursion size, and others (excursion span, total distance, and ellipse major axis) showed a statistical trend ($P<.10$).

Figure 6. Distance features including excursion size, excursion span, and total distance for each pain group.

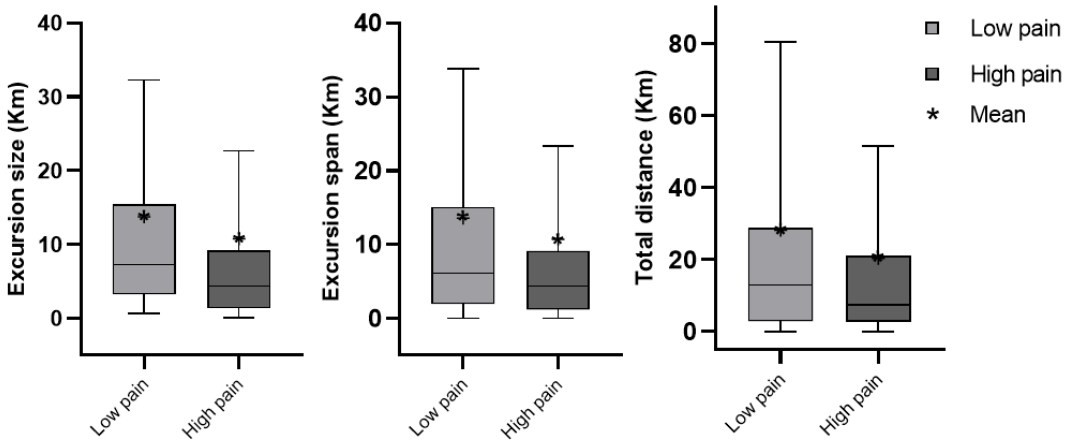


Figure 7. Ellipse minor axis, ellipse major axis, and ellipse area for each pain group.

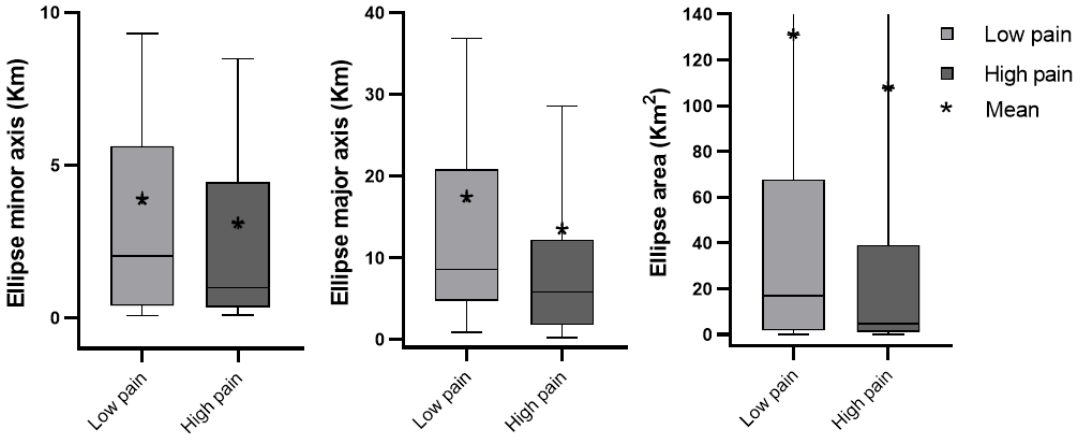


Table 3. Mixed effect association between pain and GPS features adjusted for age, living alone, and gender covariates.

GPS features and mixed-model effects[a]	Coefficient	SE	*P*>|*z*|	95% CI
Excursion size				
Between	–3.79	4.52	.40	–12.65, 5.06
Within	–3.06	1.58	.05	–6.16, 0.04
Excursion span				
Between	–3.18	4.71	.50	–12.41, 6.05
Within	–2.89	1.65	.08	–6.11, 0.35
Total distance				
Between	–11.92	9.16	.19	–29.94, 5.95
Within	–5.71	3.17	.07	–11.93, 0.51
Ellipse area				
Between	–58.23	86.66	.50	–228.07, 111.62
Within	–31.42	27.67	.26	–85.65, 22.81
Ellipse minor axis				
Between	–1.31	1.08	.23	–3.43, 0.82
Within	–0.47	0.38	.22	–1.21, 0.28
Ellipse major axis				
Between	–4.86	5.57	.38	–15.78, 6.06
Within	–3.64	1.97	.07	–7.50, 0.22
Frequency of trips				
Between	0.51	0.50	.31	–0.47, 1.49
Within	0.05	0.22	.81	–0.38, 0.49
Homestay percentage				
Between	–0.003	0.04	.94	–0.08, 0.08
Within	0.03	0.02	.15	–0.01, 0.06
Number of clusters				
Between	0.11	0.24	.63	–0.35, 0.58
Within	0.03	0.08	.73	–0.14, 0.19
Entropy				
Between	–0.04	0.07	.53	–0.18, 0.10
Within	–0.0003	0.03	.99	–0.05, 0.05

[a]Values represent the within- and between-person effect.

Figure 8. Clustering features including number of clusters and entropy for each pain group.

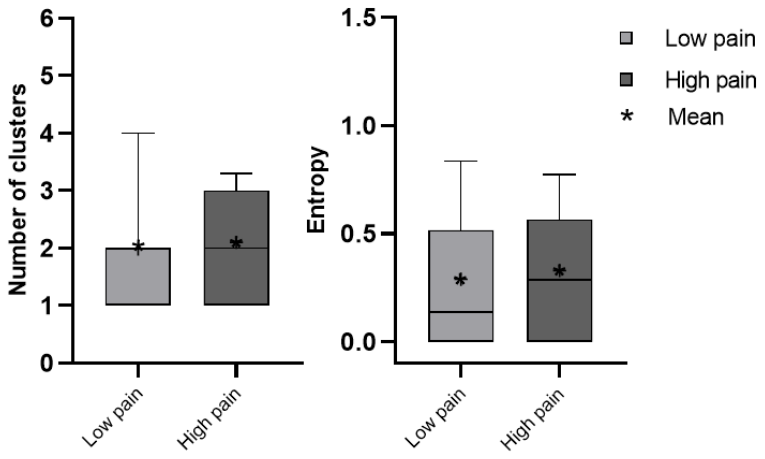
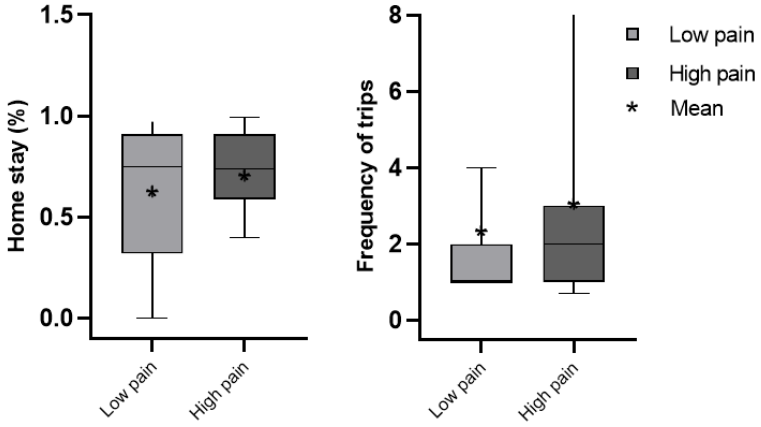


Figure 9. Frequency of trips and homestay percentages for each pain group.



Discussion

This study used a customized smartwatch app for EMA of pain and life-space mobility as a demonstration project. Previously, these constructs have not been coupled into a single platform that permits synchronizing of symptoms with objective measures of mobility in the natural environment. The results suggest that EMA of pain is negatively associated with most but not all life-space mobility features. Importantly, within-person effects, but not between-person effects, were more likely to be statistically significant. In general, older adults with confirmed knee osteoarthritis had lower life-space mobility, when pain intensity exceeded 2 out of 10. The results confirm the feasibility and analytic procedures for using smartwatch technology to harnesses sensor data alongside EMA of clinically relevant symptoms.

Chronic pain, such as pain from symptomatic knee osteoarthritis, is dynamic [30]. The variability of pain within and between days makes it hard to fully capture pain experience [5]. There has been long-standing interest in understanding daily pain [5,31]. Earlier endeavors relied on patient recall of pain, which is susceptible to recall bias and lack of ecological validity of the assessment [32]. EMA is an alternative tool to allow researchers to capture and assess a person's pain multiple times in the person's natural environments. Electronic handheld devices have provided additional features to EMA research via their ability to capture moment-by-moment data generated by their built-in sensors, allowing an in-depth understanding of the impact that pain experiences have on mobility patterns (eg, life-space mobility). Studies [33-36] have used handheld devices (eg, smartphones and iPod) for EMA of pain; however, the use of these devices was limited to electronically record participants' diaries without utilizing the built-in sensors, and no studies have used smartwatches for data collection.

In our study, 10 semantically meaningful features were extracted from the GPS coordinates according to previous work [24-26] and redefined as life-space mobility metrics. We chose to separate within- and between-person effects to study the associations of pain intensity ratings on GPS features. This was done because typical coefficients from random effects models represent a blend of both [37]. The within-person associations demonstrate that pain and life-space mobility relate to each other on an individual level. This is important because previous research demonstrates that life-space mobility is lower in people reporting higher levels of pain [37]. The within-person associations found in this study not only support the previous

between-person findings, they also support the notion that pain-related interventions are likely to have an impact on an individuals' life-space mobility. Specifically, excursion features were negatively associated with pain intensity. Among these features, within-person pain intensity was significantly associated with excursion size, and showed a statistical trend ($P<.10$) with excursion span and total distance. Each point increase in the mean pain intensity was associated with a 3.06 km decrease in excursion size, 2.89 km decrease in excursion span, and 5.71 km decrease in total distance. This suggests overall travel patterns are closer to home and more compact when older adults are experiencing a higher mean pain intensity.

The spanning ellipsoid, which summarizes the GPS coordinators into 2 dimensions, was negatively associated with within-pain intensity. The ellipsoid features represent a close approximation of the life-space concept (ie, reaching circular levels away from home). Ellipse major axis, which indicates the maximum distance across life-space, was significantly associated with pain intensity. Each point increase in the mean pain intensity was associated with 31.4 km^2 decrease in ellipse area, 0.47 km decrease in ellipse minor axis, and 3.64 km decrease ellipse major axis. Notably, the ellipse tends to have a smaller area, length, and width with higher pain intensity, which is similar to our observation about excursion features. Point estimates suggest that higher intensities of pain may constrain individuals to their home and limit the number of places they can visit.

Location clustering provides information about the distribution of places individuals spend outside of their homes. The number of clusters and entropy both contribute to understanding the variability of places visited by participants. While not statistically significant, the directionality of the coefficients indicated that higher pain intensities were associated with a higher number of places an individual stays at (stationary places). In other words, higher pain appears to be associated with spending more time at a lower number of locations, but this needs to be confirmed in larger samples.

The directionality of the point estimates demonstrated that the frequency of trips was higher when pain intensity was high. Although this may seem counterintuitive, coupled with the other results, it appears that these frequent trips were close to home. Similarly, point estimates for homestay percentage were positively associated with pain intensity. Given the weak associations of these features, trip frequency and homestay percentage may not be useful features for understanding the impact of pain on life-space mobility.

The association between pain and life-space mobility has not been widely studied, and more research is needed in this regard [38,39]. Despite the lack of relevant research, our results agree with those of Rantakokko et al [38] and Liddle et al [40], where life-space mobility was found to be negatively associated with pain. Rantakokko et al [38] examined the association between life-space mobility and multiple outcomes, including pain, in patients with Parkinson disease. They followed a paper-based questionnaire and assessed life-space mobility using life-space assessment. They found that life-space mobility is negatively associated with pain [38]. Similarly, Liddle et al [40] examined life-space mobility among patients with Parkinson disease using GPS on smartphones. They found that people with more symptoms spend more time at home and travel shorter distances.

Other studies have found strong associations between life-space mobility and depression [25], visual impairment [41], and personal and social characteristics [42,43] using GPS. Among these studies, only Cornwell et al [42] used smartphones for GPS tracking and EMA collection to examine the social environments relevant to older adults' everyday lives, where they found that certain activities such as exercising, shopping, socializing, and social activities were likely to take place outside of residential tracts. These studies [25,41-43] show the important role GPS features play, when coupled with other outcomes, in understanding individuals' behavior and their experience in natural environment and the importance of wearables in linking complex states in a temporal manner.

This demonstration study provided insights on the potential relationship between life-space mobility and pain in older adults with symptomatic knee osteoarthritis by utilizing smartwatches. Our results demonstrate that a custom-designed smartwatch platform was effective at simultaneously collecting rich information about ecological pain and life-space mobility. ROAMM could potentially help clinicians in assessing pain or other patient-reported outcomes in patients' natural environments, while continuously collecting relevant sensory data. Though our results point to interesting insights in understanding the relationship between life-space mobility and EMA of pain, our study had limitations. The sample size was not large enough to generalize and infer causality between life-space mobility and pain. Additionally, the overall pain reported by participants was low, with the majority reporting a pain intensity less than 4. In the future, we aim to recruit a larger sample size with more diversity in terms of pain intensity.

The major goal of this study was to demonstrate that a smartwatch platform—ROAMM—could be used to collect EMA of pain with concurrent mobility tracking via GPS for life-space mobility assessment in older adults with symptomatic knee osteoarthritis. Point estimates from other life-space mobility features confirm that the directionality of associations is plausible and provides initial evidence for their utility in future studies. In general, it appears that higher intensities of pain intensity tend to limit their life-space mobility by either constraining them to their residence or limiting their excursion lengths. This area of research is still in its infancy, but with apps similar to ROAMM, the demand for these tools is expected to increase remote health endeavors that are gaining significant momentum in health care. Such connected technologies have a potentially important role giving practitioners information about their patients' behaviors, symptoms, and health condition sequalae in their patients' natural environments.

Acknowledgments

Research reported in this publication was supported by the Data Science and Applied Technology Core of the Claude D Pepper Older Americans Independence Center at the University of Florida (P30 AG028740), the University of Florida Informatics Institute Seed Funds, and the University of Florida Clinical and Translational Science Institute, which is supported in part by the US National Institutes of Health National Center for Advancing Translational Sciences (award number UL1 TR001427). PR was partially supported by a National Science Foundation Career Award (NSF-IIS 1750192). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health; partial funding from R21 AG059207 was used to support staff and faculty during the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Feature correlation matrix.
[DOCX File , 33 KB - mhealth_v9i1e19609_app1.docx]

References

1. Population projections. US Census Bureau. URL: https://www.census.gov/programs-surveys/popproj.html [accessed 2020-12-26]
2. Summary health statistics: difficulties in physical functioning, 2018. National Center for Health Statistics. URL: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2018_SHS_Table_A-10.pdf [accessed 2020-12-26]
3. Inagami S, Cohen DA, Finch BK. Non-residential neighborhood exposures suppress neighborhood effects on self-rated health. Soc Sci Med 2007 Oct;65(8):1779-1791. [doi: 10.1016/j.socscimed.2007.05.051] [Medline: 17614175]
4. Osteoarthritis. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/arthritis/basics/osteoarthritis.htm [accessed 2020-12-26]
5. Bartley EJ, Robinson ME, Staud R. Pain and fatigue variability patterns distinguish subgroups of fibromyalgia patients. J Pain 2018 Apr;19(4):372-381 [FREE Full text] [doi: 10.1016/j.jpain.2017.11.014] [Medline: 29253551]
6. Schwarz N. Self-reports: how the questions shape the answers. American Psychologist 1999;54(2):93-105. [doi: 10.1037/0003-066X.54.2.93]
7. Bradburn N, Rips L, Shevell S. Answering autobiographical questions: the impact of memory and inference on surveys. Science 1987 Apr 10;236(4798):157-161. [doi: 10.1126/science.3563494] [Medline: 3563494]
8. van den Brink M, Bandell-Hoekstra EN, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. Headache 2001 Jan 20;41(1):11-20. [doi: 10.1046/j.1526-4610.2001.111006011.x] [Medline: 11168599]
9. Stinson JN, Jibb LA, Lalloo C, Feldman BM, McGrath PJ, Petroz GC, et al. Comparison of average weekly pain using recalled paper and momentary assessment electronic diary reports in children with arthritis. Clin J Pain 2014 Dec;30(12):1044-1050. [doi: 10.1097/AJP.0000000000000072] [Medline: 24535055]
10. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: microinteraction-based ecological momentary assessment (EMA) using a smartwatch. 2016 Sep Presented at: ACM Conference on Ubiquitous Computing; September 14-16; Heidelberg, Germany p. 1124-1128 URL: http://europepmc.org/abstract/MED/30238088 [doi: 10.1145/2971648.2971717]
11. Ponnada A, Haynes C, Maniar D, Manjourides J, Intille S. Microinteraction Ecological Momentary Assessment Response Rates. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-16. [doi: 10.1145/3130957]
12. Baker PS, Bodner EV, Allman RM. Measuring life-space mobility in community-dwelling older adults. J Am Geriatr Soc 2003 Nov;51(11):1610-1614. [doi: 10.1046/j.1532-5415.2003.51512.x] [Medline: 14687391]
13. May D, Nayak USL, Isaacs B. The life-space diary: a measure of mobility in old people at home. Int Rehabil Med 1985 Jul 28;7(4):182-186. [doi: 10.3109/03790798509165993] [Medline: 4093250]
14. Stalvey BT, Owsley C, Sloane ME, Ball K. The Life Space Questionnaire: a measure of the extent of mobility of older adults. J Appl Gerontol 2016 Jun 29;18(4):460-478. [doi: 10.1177/073346489901800404]
15. Peel C, Sawyer Baker P, Roth DL, Brown CJ, Bodner EV. Assessing mobility in older adults: the UAB study of aging life-space assessment. Phys Ther 2005;85(10):1008. [doi: 10.1093/ptj/85.10.1008]
16. Tinetti ME, Ginter SF. The nursing home life-space diameter. a measure of extent and frequency of mobility among nursing home residents. J Am Geriatr Soc 1990 Dec;38(12):1311-1315. [doi: 10.1111/j.1532-5415.1990.tb03453.x] [Medline: 2254569]
17. IDC forecasts steady double-digit growth for wearables as new capabilities and use cases expand the market opportunities. Business Wire. 2019 Mar 18. URL: https://www.businesswire.com/news/home/20190318005571/en/IDC-Forecasts-Steady-Double-Digit-Growth-for-Wearables-as-New-Capabilities-and-Use-Cases-Expand-the-Market-Opportunities [accessed 2019-11-04]

18. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
19. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [doi: 10.1016/s0304-3959(01)00349-9] [Medline: 11690728]
20. Jamison RN, Gracely RH, Raymond SA, Levine JG, Marino B, Herrmann TJ, et al. Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. Pain 2002 Sep;99(1-2):341-347. [doi: 10.1016/s0304-3959(02)00178-1] [Medline: 12237213]
21. Manini TM, Mendoza T, Battula M, Davoudi A, Kheirkhahan M, Young ME, et al. Perception of older adults toward smartwatch technology for assessing pain and related patient-reported outcomes: pilot study. JMIR Mhealth Uhealth 2019 Mar 26;7(3):e10044 [FREE Full text] [doi: 10.2196/10044] [Medline: 30912756]
22. Nair S, Kheirkhahan M, Davoudi A, Rashidi P, Wanigatunga AA, Corbett DB, et al. ROAMM: a software infrastructure for real-time monitoring of personal health. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services; September 14-16; Munich, Germany. [doi: 10.1109/healthcom.2016.7749479]
23. Kheirkhahan M, Nair S, Davoudi A, Rashidi P, Wanigatunga AA, Corbett DB, et al. A smartwatch-based framework for real-time and online assessment and mobility monitoring. J Biomed Inform 2019 Jan;89:29-40 [FREE Full text] [doi: 10.1016/j.jbi.2018.11.003] [Medline: 30414474]
24. Wolf PSA, Jacobs WJ. GPS technology and human psychological research: a methodological proposal. JMMSS 2010 Oct 11;1(1):1. [doi: 10.2458/azu_jmmss_v1i1_1-7]
25. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
26. Shoval N, Auslander GK, Freytag T, Landau R, Oswald F, Seidl U, et al. The use of advanced tracking technologies for the analysis of mobility in Alzheimer's disease and related cognitive diseases. BMC Geriatr 2008 Mar 26;8(1):7 [FREE Full text] [doi: 10.1186/1471-2318-8-7] [Medline: 18366748]
27. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford, England: Oxford University Press; 2009.
28. Raudenbush SW. Toward a coherent framework for comparing trajectories of individual change. In: Collins LM, Sayers AG, editors. Decade of Behavior. New Methods for the Analysis of Change. Washington, DC: American Psychological Association; 2001:35-64.
29. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 2011;62:583-619 [FREE Full text] [doi: 10.1146/annurev.psych.093008.100356] [Medline: 19575624]
30. Stone A, Shiffman S. Ecological momentary assessment (Ema) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
31. Allen K, Coffman C, Golightly Y, Stechuchak K, Keefe F. Daily pain variations among patients with hand, hip, and knee osteoarthritis. Osteoarthritis Cartilage 2009 Oct;17(10):1275-1282 [FREE Full text] [doi: 10.1016/j.joca.2009.03.021] [Medline: 19410670]
32. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
33. Fischer S, Doerr JM, Strahler J, Mewes R, Thieme K, Nater UM. Stress exacerbates pain in the everyday lives of women with fibromyalgia syndrome--the role of cortisol and alpha-amylase. Psychoneuroendocrinology 2016 Jan;63:68-77. [doi: 10.1016/j.psyneuen.2015.09.018] [Medline: 26431802]
34. Garcia-Palacios A, Herrero R, Belmonte M, Castilla D, Guixeres J, Molinari G, et al. Ecological momentary assessment for chronic pain in fibromyalgia using a smartphone: a randomized crossover study. Eur J Pain 2014 Jul 22;18(6):862-872. [doi: 10.1002/j.1532-2149.2013.00425.x] [Medline: 24921074]
35. Hallman DM, Ekman AH, Lyskov E. Changes in physical activity and heart rate variability in chronic neck-shoulder pain: monitoring during work and leisure time. Int Arch Occup Environ Health 2014 Oct 26;87(7):735-744. [doi: 10.1007/s00420-013-0917-2] [Medline: 24162088]
36. Houtveen JH, Sorbi MJ. Prodromal functioning of migraine patients relative to their interictal state--an ecological momentary assessment study. PLoS One 2013;8(8):e72827 [FREE Full text] [doi: 10.1371/journal.pone.0072827] [Medline: 23977358]
37. Kuspinar A, Verschoor C, Beauchamp M, Dushoff J, Ma J, Amster E, et al. Modifiable factors related to life-space mobility in community-dwelling older adults: results from the Canadian Longitudinal Study on Aging. BMC Geriatr 2020 Jan 31;20(1):35 [FREE Full text] [doi: 10.1186/s12877-020-1431-5] [Medline: 32005107]
38. Rantakokko M, Iwarsson S, Slaug B, Nilsson MH. Life-space mobility in Parkinson's disease: associations with motor and non-motor symptoms. J Gerontol A Biol Sci Med Sci 2019 Mar 14;74(4):507-512. [doi: 10.1093/gerona/gly074] [Medline: 29648576]
39. Mackichan F, Adamson J, Gooberman-Hill R. 'Living within your limits': activity restriction in older people experiencing chronic pain. Age Ageing 2013 Nov;42(6):702-708. [doi: 10.1093/ageing/aft119] [Medline: 23978405]

40. Liddle J, Ireland D, McBride SJ, Brauer SG, Hall LM, Ding H, et al. Measuring the lifespace of people with Parkinson's disease using smartphones: proof of principle. JMIR Mhealth Uhealth 2014 Mar 12;2(1):e13 [FREE Full text] [doi: 10.2196/mhealth.2799] [Medline: 25100206]
41. Curriero FC, Pinchoff J, van Landingham SW, Ferrucci L, Friedman DS, Ramulu PY. Alteration of travel patterns with vision loss from glaucoma and macular degeneration. JAMA Ophthalmol 2013 Nov 01;131(11):1420-1426 [FREE Full text] [doi: 10.1001/jamaophthalmol.2013.4471] [Medline: 24030033]
42. York Cornwell E, Cagney KA. Aging in activity space: results from smartphone-based GPS-tracking of urban seniors. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):864-875 [FREE Full text] [doi: 10.1093/geronb/gbx063] [Medline: 28586475]
43. Boissy P, Blamoutier M, Brière S, Duval C. Quantification of free-living community mobility in healthy older adults using wearable sensors. Front Public Health 2018 Aug 13;6:216 [FREE Full text] [doi: 10.3389/fpubh.2018.00216] [Medline: 30151357]

Abbreviations

EMA: ecological momentary assessment
GPS: global positioning system
ROAMM: Real-time Online Assessment and Mobility Monitor

Edited by G Eysenbach, L Buis; submitted 24.04.20; peer-reviewed by S Schneider, S Difrancesco; comments to author 25.06.20; revised version received 18.09.20; accepted 20.10.20; published 13.01.21.

Please cite as:
Mardini MT, Nerella S, Kheirkhahan M, Ranka S, Fillingim RB, Hu Y, Corbett DB, Cenko E, Weber E, Rashidi P, Manini TM
The Temporal Relationship Between Ecological Pain and Life-Space Mobility in Older Adults With Knee Osteoarthritis: A Smartwatch-Based Demonstration Study
JMIR Mhealth Uhealth 2021;9(1):e19609
URL: http://mhealth.jmir.org/2021/1/e19609/
doi:10.2196/19609
PMID:33439135

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org