JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 7 (2019), Issue 8 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Original Papers

Viewpoints

Reviews

Original Paper

An App to Improve Eating Habits of Adolescents and Young Adults (Challenge to Go): Systematic Development of a Theory-Based and Target Group–Adapted Mobile App Intervention

Anna Rohde[1,2,3], MSc; Anja Duensing[1,2], MSc; Christine Dawczynski[1,2], PhD; Jasmin Godemann[3], PhD; Stefan Lorkowski[1,2], PhD; Christine Brombach[1,4], PhD

[1]Institute of Nutritional Sciences, Faculty of Biological Sciences, Friedrich Schiller University Jena, Jena, Germany

[2]Competence Cluster for Nutrition and Cardiovascular Health, Halle-Jena-Leipzig, Germany

[3]Department of Communication and Engagement in Agricultural, Nutritional and Environmental Sciences, Justus Liebig University Giessen, Giessen, Germany

[4]Institute of Food and Beverage Innovation, Zurich University of Applied Sciences, Zurich, Switzerland

Corresponding Author:
Anna Rohde, MSc
Institute of Nutritional Sciences
Faculty of Biological Sciences
Friedrich Schiller University Jena
2nd Fl
Dornburger Straße 25
Jena, 07743
Germany
Phone: 49 3641 9 49710
Fax: 49 3641 9 49712
Email: at.anna.rohde@gmail.com

Abstract

Background: Due to the widespread use of mobile phones, dietary mobile apps are promising tools for preventing diet-related noncommunicable diseases early in life. However, most of the currently available nutrition apps lack scientific evaluation and user acceptance.

Objective: The objective of this study was the systematic design of a theory-driven and target group–adapted dietary mobile app concept to promote healthy eating habits with a focus on drinking habits as well as consumption of fruits and vegetables in adolescents and young adults, especially from disadvantaged backgrounds.

Methods: The design process was guided by the behavior change wheel (BCW). The development process comprised 3 stages. In stage 1, the target behavior was specified, and facilitators and barriers were identified. Furthermore, important insights into target group interests, needs, and values in the field of nutrition and apps were revealed. To this end, 2 empirical studies were conducted with the target group. In stage 2, results of stage 1 were translated into behavior change techniques (BCTs) and, finally, into app functionalities and features. Consequently, in stage 3, the concept was evaluated and optimized through expert interviews.

Results: Facilitators and barriers for achieving the target behavior were psychological capabilities (eg, self-efficacy), reflective motivation (eg, fitness), automatic motivation, social support, and physical opportunity (eg, time). Target group interests, needs, and values in the field of nutrition were translated into target group preferences for app usage, for example, low usage effort, visual feedback, or recipes. Education, training, incentives, persuasion, and enablement were identified as relevant intervention functions. Together with the target group preferences, these were translated via 14 BCTs, such as rewards, graded tasks, or self-monitoring into the app concept Challenge to go (C2go). The expert evaluation suggested changes of some app features for improving adherence, positive health effects, and technical feasibility. The C2go concept comprises 3 worlds: the (1) drinking, (2) vegetable, and (3) fruit worlds. In each world, the users are faced with challenges including feedback and a quiz. Tips were developed based on the health action process approach and to help users gain challenges and, thereby, achieve the target behavior. Challenges can be played alone or against someone in the community. Due to different activities, points can be collected, and levels can be achieved. Collected points open access to an Infothek (information section), where users can choose content that interests them. An avatar guides user through the app.

Conclusions: C2go is aimed at adolescents and young adults and aims to improve their fruit and vegetable consumption as well as drinking habits. It is a theory-driven and target group–adapted dietary mobile intervention concept that uses gamification and was systematically developed using the BCW.

(JMIR Mhealth Uhealth 2019;7(8):e11575) doi:10.2196/11575

KEYWORDS

adolescents; young adults; mobile phone; mobile apps; mHealth; health behavior; healthy eating; motivation

Introduction

Background

Globally, diet-related noncommunicable diseases (NCDs) are the leading cause of death and disease burden [1,2]. Numerous studies emphasize the association between a suboptimal diet and deaths due to NCDs such as stroke, heart disease, or type 2 diabetes [3-6]. Among dietary risk factors for NCDs are the low consumption of fruits and vegetables [7-9] and the high consumption of sugar-sweetened beverages [10-14].

German survey data highlights the high prevalence of overweight and obesity. Almost 60% of the population is overweight or obese (women 51% and men 66%) [15]. Among the younger population, about 16% of girls and 18.5% of boys in the age between 14 and 17 years are overweight or obese [16], likely due to a more sedentary lifestyle characterized by decreased physical activity and unbalanced dietary behavior [15]. Only 7% of the girls and boys in Germany in the age between 14 and 17 years follow the dietary recommendations for the consumption of 2 portions of fruits and 3 portions of vegetables each day. On average, 0.9 portion each of vegetables and fruits per day are consumed [17]. Nearly, 23% of boys and 17% of girls drink sugar-sweetened beverages daily [18], whereas recommended beverages are water and unsweetened teas [19]. Data from the United States also reveal similar results. American guidelines recommend 2.5 cups of vegetables and 2 cups of fruits per day (for a calorie level of 2000 kcal) as well as drinking water and reducing consumption of sugar-sweetened beverages [20]. However, less than 50% of children and adolescents meet dietary recommendations for any food group [21]. The level of education positively influences food consumption and, thus, the quality of the diet [22,23]. Higher school education and higher income result in a lower body mass index [15,16].

To summarize, nutrition surveys highlight that adolescents and young adults have unbalanced diets [17,24,25], especially adolescents and young adults in disadvantaged circumstances [17]. As adolescence is characterized by cutting ties with the parents' household and the development of one's own lifestyle, this could be a reasonable stage of life for behavior change interventions [26], in particular for focusing on adolescents and young adults with lower education levels.

Digitalization and Mobile Health

State-of-the-art mobile phones provide new possibilities for dietary interventions. Mobile health (mHealth) is an emerging field and describes various health services offered on portable devices. These include health apps in various areas, such as nutrition, fitness, wellness, diagnostics, and therapy [27], but systematic studies in the area of mHealth are scarce [28-30]. Studies also highlight missing user acceptance of nutrition apps, for which the relatively high usage effort might be a reason [31,32]. Rohde et al concluded that app usage in the long term is influenced by user- and app-related acceptance factors [32]. The former highlights the importance of knowing the target group for designing accepted mHealth interventions; the latter emphasizes the importance of considering different app characteristics, for example, implementing instructions or motivators for engagement and adherence in app-based interventions [32].

In the context of long-term adherence and acceptance of mHealth interventions, *gamification* is an emerging field. Gamification means that playful elements such as points or leaderboards are used in a context that is normally not played (for example using a quiz, where one can earn points, instead of just giving information) [33,34]. Gamification can be a motivational component of digital behavior change interventions by playfully making uninteresting topics interesting and, thereby, engage users in the long term [34].

Theory Guidance for Intervention Development

The behavior change wheel (BCW) is a framework for developing health interventions [35]. It proposes a systematic design process of behavior change interventions that helps to translate theory into practice [30].

The health action process approach (HAPA) is a health behavior model and, as a stage model, an interesting template for the theory-based development of dietary health messages that can be adapted to persons at different stages of the behavior change process [36]. It was successfully applied in several nutrition behavior change interventions [37-39].

Objective

The aim of this study was to describe the iterative concept development and the final concept of a theory-based and target group–adapted mobile app for motivating adolescents and young adults (aged 14-25 years), especially from disadvantaged backgrounds, to improve their dietary habits with respect to the consumption of fruits and vegetables, as well as drinking behavior.

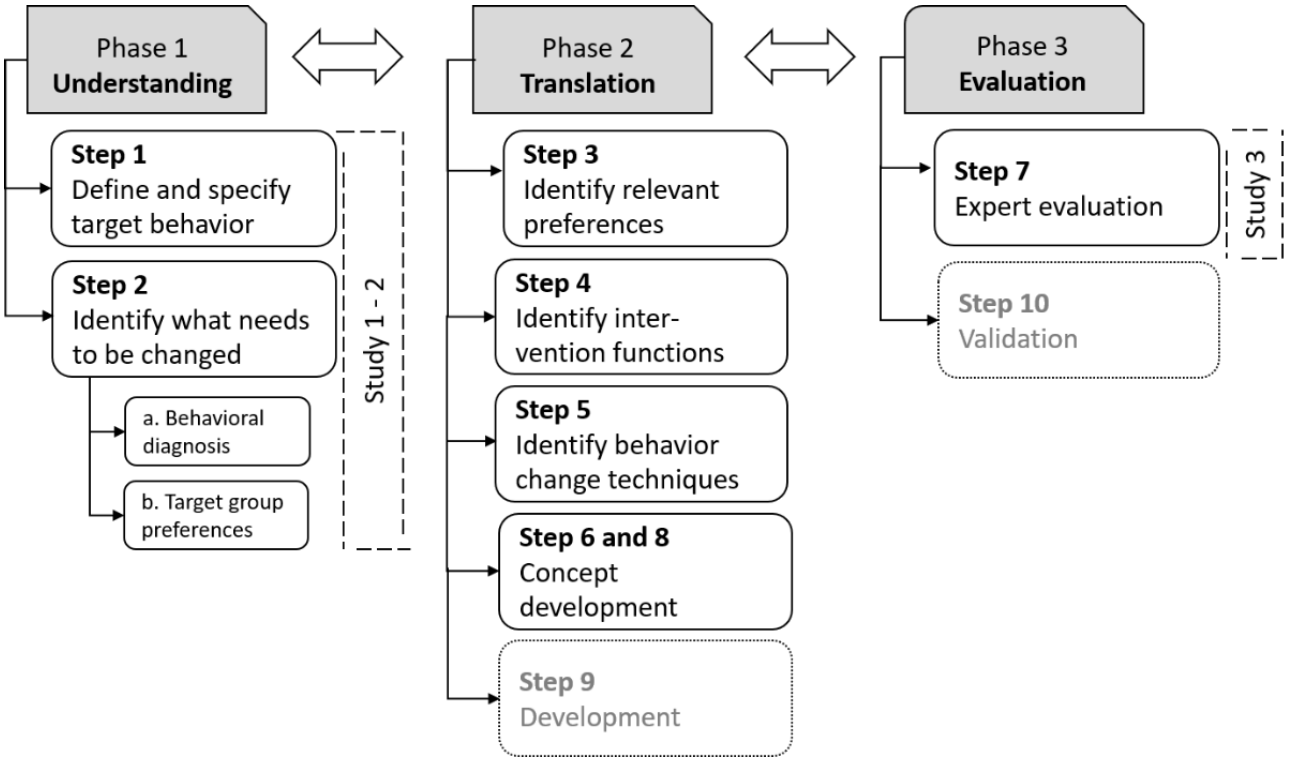
Methods

Overview

The app design process was guided by the BCW [40]. After defining the problem and the 3 target behaviors, the app design process followed 3 stages (Figure 1): Phase 1, specifying the target behavior and identifying what needs to change to achieve

it; Phase 2, translating results into app functionalities and features; and Phase 3, expert evaluation of the concept. In total, 3 empirical studies were conducted to derive relevant app features and content as well as to optimize the concept.

Figure 1. Systematic design process of the dietary mobile app for adolescents and young adults. Steps 1-7 (in black font) are discussed in the text; steps 9 and 10 (in grey font) are being or will be carried out.



Phase I: Understanding Behavior and Target Group Preferences

Step 1: Define and Specify Target Behavior

In total, 3 target behaviors were chosen by the author (AR) by identifying possible target behaviors through literature search and opinions of nutrition experts from the Friedrich Schiller University Jena. A list of potential target behaviors was shortened using the following criteria: Likely impact of the target behavior on outcome, ease of reaching target behavior, possible positive or negative spillover effects, and ease of measurement [40]. Each criterion was rated as unacceptable, unpromising but worth considering, promising, or very promising. Rating and selecting the target behavior were supported by empirical research (Step 2b). Upon selection of the most promising target behaviors, they were specified in detail and context: who, when, where, how often, and with whom will the target group perform the target behavior?

Step 2: Identify What Needs to Change to Reach Target Behavior

Behavioral Diagnosis

The capability-opportunity-motivation-behavior (COM-B) model was used to identify what needs to change for adolescents and young adults to achieve the target behavior. Evidence from the literature and empirical research informed this procedure by exploring target group's capabilities, motivation, and opportunity to achieve the target behaviors (eg, which physical capabilities are needed to eat 2 portions of fruit per day?) and the need for change. Next, each capability, motivation, and opportunity was evaluated as feasible or not for implementing in a dietary mobile app.

Target Groups Preferences (Sampling From Empirical Studies 1 and 2)

Besides informing behavioral diagnosis, study results were also analyzed to explore the target group's preferences for app characteristics, behavior change techniques (BCTs), features, and content. The ethics committee of the University noted no ethical concerns (processing number 4850-06/16).

Study 1: Nutrition and Apps From the Target Group's Perspective

The objective of this study, conducted in 2016, was to get insights into nutrition habits, values, and needs, and to develop ideas as to how nutritional behavior and health among adolescents and young adults could be improved through mobile apps. Study participants (n=11) tested the German dietary mobile app *Was ich esse* [41] for 1 week before face-to-face, semistructured interviews. The participants were between 14 and 21 years old (average age: 18 years, SD 2.4) and mostly women (n=8). A total of 5 participants went to secondary school at the time of the study. Others were university students (n=3), trainees (n=1), volunteers (n=1), or were looking for an apprenticeship (n=1). The audio-recorded material was transcribed and analyzed by means of content analysis [42]. Data segments were coded into the following topics: (1) mobile phone and app usage, (2) test app experiences, (3) nutritional habits, (4) nutrition values, (5) nutritional improvement wishes and strategies, and (6) understanding of health. Additional

information on recruitment, compensation, the interview guide, and the test app are provided in Multimedia Appendix 1.

Study 2: Nutrition and Mobile Phone Apps—Interests, Needs, and Values Among Adolescents and Young Adults

This study aimed for a better understanding of available mobile phone resources, app use, and needs as well as interests and values in the field of nutrition of adolescents and young adults. To this end, a questionnaire was developed and administered as a paper-pencil version. It included the following topics: mobile operating system, mobile phone rate, favorite apps, experiences with dietary mobile apps, importance of different app characteristics (eg, importance of customizability), and nutritional interests (eg, sports nutrition, health, and food waste) and values (eg, freshness of food and self-cooked meals). Data were analyzed descriptively. The inclusion criterion for participation was an age between 14 and 25 years. Teachers and social workers in Jena were contacted via mail to secure them as gatekeepers. In total, 210 participants from 5 different organizations took part (females n=99, males n=108, no information provided n=3). The average age was 18 years (n=208, SD 2.4; no information provided n=2). The youngest and oldest person was 15 and 25 years old, respectively. Participants went to vocational school (Berufs(fach)schule) (n=164). Others stated that they had participated in vocational preparation classes (n=27) or went to high school (Gymnasium) (n=11). One person each went to secondary school (Hauptschule) and regular school (Regelschule); 4 persons stated *other* (no information provided n=2). Additional information on recruitment, compensation, and the scales in the questionnaire are presented in Multimedia Appendix 1.

Phase II: Translation of Research Results Into App Content and Features

Step 3: Identify Relevant Target Group Preferences

Results from studies 1 and 2 were compared and merged in target group preferences, and an acceptance-rejection process followed (Figure 2). Decisions of the author (AR) for rejection or acceptance of target group preferences were based on APEASE criteria [40]: Is the respective preference *a* ffordable, *p* racticable, *e* ffective, and *a* cceptable, are *s* ide-effects expected or offense against *e* quity? If all criteria are met or at least rated probably, the preference was accepted and further implemented as app content, app features, BCTs, or considered as important app characteristics (Steps 5 and 6).

Step 4: Identify Intervention Functions

This step aimed to move from understanding the behavior to selecting intervention functions. This was supported by a matrix of links between COM-B and intervention functions [40]. Appropriate intervention functions were selected by using the APEASE criteria.

Step 5: Identify Behavior Change Techniques

To choose which BCTs can deliver the intervention functions, a linking was used [40]. The list of candidate BCTs (n=118) was narrowed by APEASE criteria. The rating was supported by the evidence of effectiveness for promoting healthy food choices, and on the basis of target group preferences.

Step 6: Concept Development (Prototype I)

Together with target group preferences, the BCTs were translated into app features, content, and characteristics. Content development of feedback was guided by HAPA for preintender and intenders and gamification aspects were implemented to enhance user engagement with the app [30,34].

Phase III: Evaluation

Step 7: Expert Evaluation

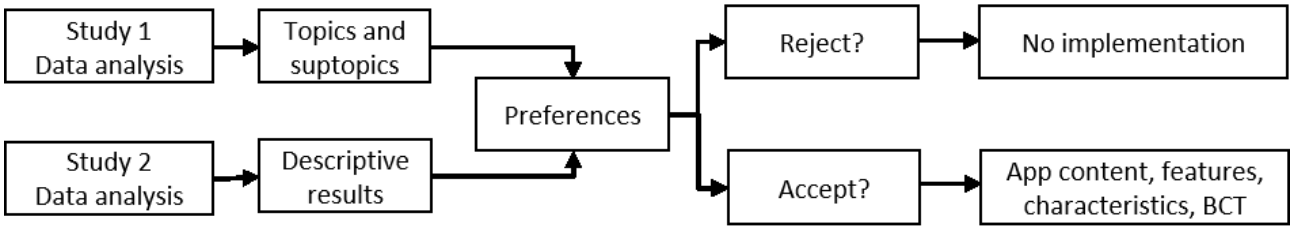
The concept was evaluated and optimized using 3 evaluation criteria: (1) acceptance among target group, (2) positive health effects due to app use, and (3) technical feasibility. To this end, professionals with knowledge of the target group, app development or nutrition behavior were recruited. Recruitment took place via emails and later personally by telephone. Ultimately, 8 face-to-face interviews were conducted with experts of the following professions: Marketing, 2 social workers/teachers, dietician, app development, media psychology, psychotherapist, and a person of the target group itself. The interviews started with the presentation of the concept using mock ups. The semistructured interviews were audio recorded and transcribed verbatim. Data were evaluated with structured qualitative content analyses [42].1 The following 5 topics were discussed: (1) important features and needs of a dietary mobile app among the target group, (2) advantages of the concept, (3) disadvantages of the concept, (4) suggestions to improve the concept, and (5) technical feasibility.

In the next step, the recommendations for improving the concept were rated using APEASE criteria and were either accepted and implemented, or rejected and not implemented.

Step 8: Final Intervention (Prototype II)

On the basis of the evaluation, the concept was adapted by defining the final features and functionalities of the app.

Figure 2. Process of identification of relevant target group preferences. BCT: behavior change technique.



Results

Phase I: Understanding Behavior and Target Group Preferences

Step 1: Defined and Specified Target Behavior

A range of target behaviors was listed, such as consumption of more vegetables, fruits, water, tea, fibers, or eating less saturated fatty acids. Following this, an important point in rating the ease of achieving the behavior was to avoid the feeling of waiver in nutritional terms, so that a target behavior does not forbid, but permits and increases, consumption of food [43].

Finally, 3 target behaviors, in line with German intake recommendations [19] were chosen: The consumption of (1) 2 portions of fruits per day, (2) 3 portions of vegetables per day, and (3) drinking 1.5 L or more of unsweetened beverages per day to decrease the consumption of sugar-sweetened beverages (only nonalcoholic beverages are considered). To achieve the target behaviors, adolescents and young adults have to eat and drink fruits, vegetables, and sugar-free drinks at mealtimes (or in between), and often enough to achieve the respective target behavior. They can do it anywhere and either by themselves or with others.

Step 2: Identified What Needs to Change to Reach Target Behavior

Behavioral Diagnosis

The results of the behavioral diagnosis revealed facilitators and barriers to the target behavior in the following COM-B components: psychological capabilities (eg, nutrition education, self-efficacy, and risk perception), reflective motivation (eg, weight loss, satiety, fitness, and illness prevention), automatic motivation, social support, and physical opportunity (eg, time and financial resources). An overview of the results with quotes from study participants and references is displayed in Multimedia Appendix 2.

Target Group Preferences: Empirical Study Results

Study 1

An excerpt of the results in 4 of the 6 main topics is presented in Table 1. In addition, Multimedia Appendix 3 presents a complete overview of the results.

Study 2

The operating system most used was Android (Google) and most of the participants used a mobile flat rate. Among the participants' favorite apps were communication and social media apps (WhatsApp, Facebook, and Instagram), video apps such as YouTube, and gaming apps such as Clash of Clans and Clash Royal. In all, 26% of the participants had experiences with apps in the area of nutrition, above all recipe apps. The most interesting subjects in the area of nutrition were health, cooking, and sports nutrition. Good taste, satiety, and freshness of food were the most important nutritional values. The most important app characteristics were free of charge, contact to friends/family, and fast use. Multimedia Appendix 3 gives a full insight into the results.

Phase II: Translation of Research Results Into App Contents and Features

Step 3: Identified Relevant Target Group Preferences

An excerpt of results of the process of the identification and selection of relevant target group preferences for app characteristics and features is presented in Table 2. For the derivation of the preferences, all subtopics of the topics (1) to (5) of study 1 were included as well as the most frequent answers of study 2. Results that were rated mostly *not important* and *partly true* with tendency to *not important* were not considered to derive preferences. For example, *Nutrition habits of other cultures* and *Nutrition and skin* were both mostly rated as *partly interesting* (n=91; n=95). The first was not considered to derive a preference because it shows a tendency toward *not interesting* (n=72). However, the latter showed a tendency toward *interesting* (n=65) and was therefore considered to derive a preference.

Step 4: Identified Intervention Functions

Candidate intervention functions were education, persuasion, incentivization, coercion, training, restriction, environmental restructuring, modeling, and enablement. The rating of theses for both fruits and vegetables, and drinking behavior led to the selection of education, persuasion, incentivization, training, and enablement.

Step 5: Identified Behavior Change Techniques

According to the APEASE criteria, 14 BCTs were derived to bring about behavior change (Table 3).

Table 1. Excerpt of results from study 1.

Main topic, subtopics	Quotes (translated)
Mobile phone and app usage	
Mobile phone is used for entertainment and when bored (eg, games and videos)	Jana: "I use my mobile phone when I'm bored or when I have to wait for the bus or something, then I play games."
Nutritional values	
Cooking stands for independency	Caro: "Yeah, for later, if maybe I have a family and I cannot cook, that would be a bit ... And cooking is also important to me, so I do not always depend on someone."
Spending on food should be kept low	Leon: "We like to eat exotic fruits. But you always have to see how much money you have at your disposal."
Nutritional improvement wishes and strategies	
Eating healthier	Jana: "I often wish that I ate healthier."
Test app experiences	
High usage effort through tracking	Daria: "So, I did not continue using the app because it was very time-consuming tracking everything and properly. At the beginning it was a lot of fun, but eventually it got harder, because sometimes you do not think about tracking."
Visual feedback is used as consumption orientation and promotes self-control	Tino: "Through the app I've noticed that I do not eat enough vegetables. That's why I bought some cucumbers or tomatoes."

Table 2. Excerpt of identified and selected target group preferences for app characteristics and features.

Study 1		Study 2	Target group preferences for app characteristics and features, based on findings from studies 1 and 2	Accept or reject
Topic	Subtopic	Results		
Mobile phone and app usage	Mobile phone is used for entertainment and when bored (eg, games)	Important app characteristics: Entertainment	Features for use when bored/individual time of usage	Accept[a]
Mobile phone and app usage	Listening to music	—[b]	Music	Reject[c,d]
Test app experiences	Disadvantage: High usage effort through tracking (as a result not everything was tracked)	Important app characteristics: Fast use	Supporting low user effort and fast use	Accept[a]
Test app experiences	Advantage: Test-app use for comparison of visual feedback with others	Favorite apps are mostly communications apps	Social comparison	Accept[a]
Test app experiences	Improvement suggestion: More feedback through additional evaluation charts	—	Different evaluation charts	Reject[c,e]

[a]Maintain suspense and adherence.

[b]Not applicable.

[c]Not relevant for target behavior/not in line with target behavior.

[d]Not affordable as incentive.

[e]Focus shall be kept on portions not on calorie intake.

Table 3. Intervention functions with capability-opportunity-motivation-behavior components and behavior change techniques (including evidence of effectiveness and target group preferences).

Intervention functions	BCTs[a] with evidence from literature	Target group preferences
Education	Self-monitoring of behavior [44-46]	Tracking for promoting awareness of eating behavior
Education	Feedback on behavior [47-50]	Tips are motivational (low cost and easy tips)
Education	Information about health consequences [47,51,52]	—[b]
Education	Prompts/cues [53]	Support through reminder
Persuasion	Information about health consequences [47,51,52]	—
Persuasion	Feedback on behavior [47-50]	Tips are motivational (low cost and easy tips)
Persuasion	Verbal persuasion about capability [34]	—
Persuasion	Social comparison [49,54,55]	Social comparison
Incentivization	Feedback on behavior [47-50]	Tips are motivational (low cost and easy tips)
Incentivization	Self-monitoring of behavior [44-46]	Tracking for promoting awareness of eating behavior
Incentivization	Nonspecific incentive/reward (includes positive reinforcement) [49,56,57]	Gamification
Training	Instruction on how to perform a behavior	Tips are motivational (low cost and easy tips)
Training	Feedback on behavior [47-50]	Tips are motivational (low cost and easy tips)
Training	Self-monitoring of behavior [44-46]	Tracking for promoting awareness of eating behavior
Training	Graded tasks [58]	—
Enablement	Action planning[c] [44]	—
Enablement	Coping planning[c] [44]	—
Enablement	Goal setting (behavior) [59]	Goal setting
Enablement	Discrepancy between current behavior and goal [46]	Tracking for promoting awareness of eating behavior
Enablement	Self-monitoring of behavior [44-46]	Tracking for promoting awareness of eating behavior
Enablement	Graded tasks [58]	—
Enablement	Social support (unspecified) [55]	—

[a]BCT: behavior change technique.

[b]Not applicable.

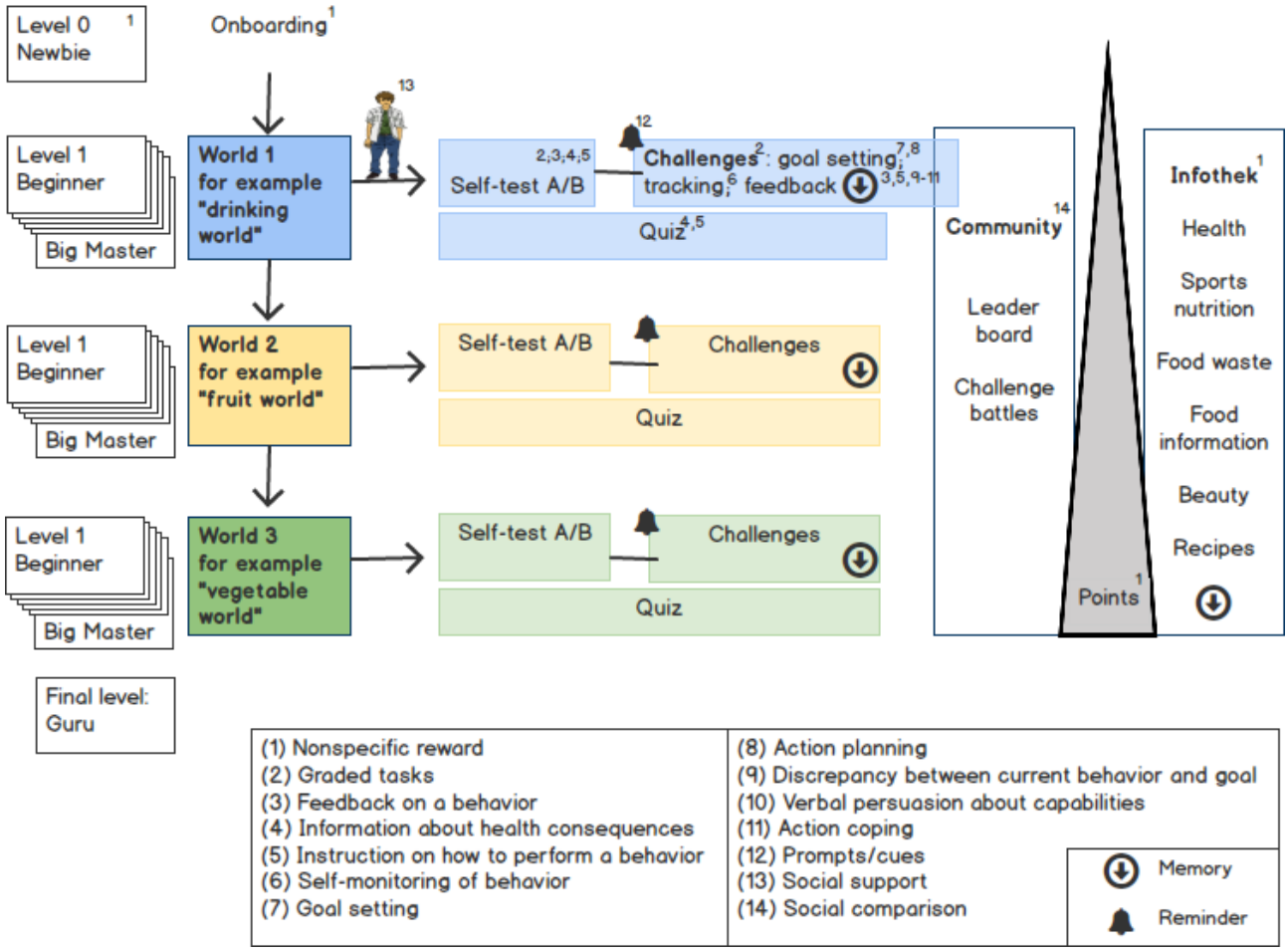
[c]Based on the health action process approach [36].

Step 6: Developed Preliminary Concept (Prototype I)

This step resulted in the development of the *Challenge to go* (*C2go*) app concept. Multimedia Appendix 4 shows how target group preferences and BCTs were matched and jointly translated into app features.

The following section gives an overview over the concept (Figure 3). After onboarding, the user can choose among 3 worlds: the *drinking*, the *vegetable*, or the *fruit world*. In each world, users can accept *challenges* and participate in a *quiz*. To get access to the challenges, users must go through self-tests. Consequently, in a challenge, the user must choose a behavioral goal from a list that he or she tries to achieve, for example, in the fruit world, one portion of fruit per day for 1 week. Challenges can be played alone or against someone else in the community. Different *feedback* is given to motivate and to empower the user to achieve his or her challenge goals, for example, the informative feedback after a lost challenge gives tips on how challenge goals can be reached. Each tip is stored and always accessible for the user. For further support, *reminders* can be set. Through different activities users earn *points* and achieve *levels*. The points open access to the *Infothek*, where users can choose content that is of their interest. The content received from the *Infothek* is stored and always accessible. A leaderboard compares user scores in the *community*, which is made up by other app users. Through passing challenges, users ascend in levels all the way up to *Big Master*. After completing a world, the next world can be selected. If the user has reached the highest level in every world, he or she becomes a *Guru*. Through the whole app, users are guided by an *avatar*.

Figure 3. Flowchart of *C2go*.



Phase III: Evaluation

Step 7: Expert Evaluation

Data from expert evaluation and rating results suggested changes for the following app features: worlds, challenges, feedback, *Infothek*, and quiz. Furthermore, the issue of usage motivation was discussed (Table 4).

Step 8: Final Intervention (Prototype II)

The following sections describe the app features of Prototype II in detail. Further information regarding interventions functions, BCTs, and COM-B components are given in Multimedia Appendix 4.

Onboarding

Onboarding, that is, the way a user is introduced into the app content, is important to motivate the user for adherence [33]. As such, the app starts with an introductory question (Textbox 1).

The introductory question is used to make users curious and motivate them to use the app through connecting his or her personal aim with app usage and selecting a question that can only be answered correctly, so that the user cannot lose and does not become demotivated [33]. Depending on the answer, a progress bar representing the guru status is implemented, and either titled *Health Guru*, *Wellbeing Guru*, or *Performance Guru*. A user reaches the *Guru* level after succeeding in all 3 worlds.

Consequently, a tutorial gives a brief overview of the app and how to use the app [32,33]. Afterwards, profile settings allow enhancement and customization of the app [33] and are associated with the selection of the avatar in terms of age and sex.

Self-Test

Each world starts with a self-test, which comprises 2 parts. The first part is a 24-hour dietary recall (eg, vegetables in the vegetable world). Results of this recall are used to divide users into HAPA stages—either intender, if dietary guidelines are not met, or actor, if guidelines are met [19]. The second part is a quiz, which serves as an introduction into the world concept and challenge rules and basics of nutrition education. The answer of each question is formulated depending on the HAPA stage of the user [36] and personal app aim (Textbox 2 for an example), which are stated by the user. After completing both parts of the self-tests, the user gets access to the challenges.

Dissolution: Drinking (at least) 1.5 L per day of sugar-free drinks is good for you and helps you to get closer to your goal: to live healthier.

Table 4. Concept changes after expert evaluation.

App feature/characteristic	Prototype I	Prototype II
Challenges	Strict defined rules to reach goals	Easing the rules: More jokers for the users of the drinking world in the *Big Master* level
Feedback	Visual feedback	More visual feedback, for example, a smiling avatar for rewarding consumption of water
Feedback	Informative feedback (tips)	Optional extra button for more information, the user can choose if he/she wants more information concerning feedback
Feedback	—[a]	Reflective questions in feedback, for example, *What are your best tips for a friend to encourage him to drink more water?*
Feedback	—	Goal-orientated feedback: Asking an introductory question, linking the tips to the answer and thereby to the user's goal
Usage motivation of app	Feedback development: Focus on preintender and intender	Informative feedback focuses on intenders only, preintenders are not considered. This decision was supported by literature, as health apps are more likely used by health-conscious people [31]
Worlds	—	Bonus worlds: Might be possible with an update of the app (menu button *bonus worlds*)
Worlds	Worlds are started one after the other	The worlds can be played simultaneously
Infothek	Point-related access or as a present	More frequently access: Through reduction of the distance of the points
Infothek	Information about healthy snacks/drinks	Including more information about healthy snacks and drinks
Infothek	Health, food waste, and beauty are topics	Content ideas for the following topics: (1) *Health***:** Where does enjoyment of taste start and where does it end (start of addiction); mechanism of resorption of nutrients; *foodwaste*: Shelf life of food, seasonality of fruits and vegetables, environmental pollution; (2) *Beauty*: Cosmetics without any animal testing
Quiz	Answering question with yes/no; assigning answers	Answering questions with drag and drop mechanisms

[a]Not applicable.

Textbox 1. Introductory question.

Where can Challenge 2 go support YOU the most?

- Live healthier
- Feeling good in my body
- More fitness and performance

Textbox 2. Example of a quiz question from the second part of the self-test. Example for a user whose personal aim is to live healthier.

How much should we drink at least a day?

- 0.5 L
- 1.5 L
- 1 L
- 3 L

Challenges

The challenges are implemented for the user to reach self-imposed consumption goals, which can be chosen out of a list, for example, 3 servings of vegetables per day in the vegetable world or 3 portions of unsweetened beverages in the drinking world (level *Adept* in Table 5).

In the drinking world, a sugar mountain builds up additionally, while tracking sugar-sweetened beverages, which users must reduce through answering quiz questions before attacking the next challenge. In addition, in the fruit and vegetable world, it is not only quantity, but also quality, that counts. Users are motivated to eat as many colors as possible, as recommended in other studies [39,60]. In every world, the aim is to get better in behavioral terms from challenge to challenge up to the highest

level, called *Big Master*, which meets the target behavior. To pass the challenges, users get support through feedback and tips. Before finishing a world successfully, every user must play the *Big Master* challenge, even if he or she is already carrying out the behavior and finishing the quiz.

Feedback and Community

Different feedback is implemented in the *C2go* app: visual, informative, motivational, evaluative, and competitive. An avatar, which is intended as social support [36], gives some visual and all informative, motivational, and evaluative feedback. The avatar rewards user's desirable behavior with a facial expression (smiling face). Other visual feedback is given through progress bars for different features (self-test, challenges, quiz, and overall progress in app that represents the *Guru Status*) and a graph for an actual versus target feedback for the target behavior. Informative feedback contains messages relevant for intenders (based on HAPA [44]) by encouraging action coping (overcoming barriers, including reflexive questions) and action planning (when, how, where implementing behavior, including instructions on how to perform a behavior) to close the gap between intention and actual behavior [36,39]. Motivational feedback is given through encouraging messages that should sustain intention and self-efficacy [34,48]. Similarly, self-efficacy will be boosted by evaluative feedback [34,39]. When creating feedback, care was taken that this was formulated in a positive way [32,48] and using colloquial language [39]. For examples, refer to Table 6. Furthermore, competitive feedback is given through a leaderboard in the *community* [34]. In addition, challenges can be played alone or against other app user in the community to promote motivation for behavior change through social comparison and competition [54,61]. The community consists of all app users.

Reminder

Reminders are push notifications and can be set to support users on the way to meeting their goals. Furthermore, they function as a re-engagement tool [53]. Types of different reminders are presented in Table 7. Every reminder can be switched off or on.

Table 5. Levels in the drinking world of the app.

Phase	Level
After selecting first world	Beginner
After completing self-test B	Climber
After the first challenge (24-hour-Challenge)	Adept
3 portion challenge	Adept pro
4 portion challenge	Expert
5 portion challenge	Expert pro
6 portion challenge	Master
6 portion challenge + no sugary drinks	Big Master

Table 6. Feedback examples.

Feedback type	Time point	Example
Motivational	During challenges	*Great first 7 days!*
Evaluative	After a challenge	*Try again: the next level is already waiting for you!*
Informative	After a challenge	*Drink a portion of unsweetened beverage with each meal, for example, (mineral) water or herbal/fruit tea (warm or cold)*

Table 7. App reminders and time points.

Reminder	Time point
Consumption of target behavior	From 9 am to 9 pm, every 3 hours
Tracking	9 pm
Start of challenge	9 am
End of challenge day	9 pm, if necessary next day 7 am und 12 pm
End of a challenge	Immediately, if necessary 36 hours and 48 hours later
Infothek	If access has been granted

Quiz

In terms of content, the quiz aims to provide knowledge about the health-related value of each target behavior and the national intake recommendations. To this end, each world has its own quiz, for example, the fruit world quiz contains questions regarding fruit intake recommendations and associated health

benefits and risks (Textbox 3 for an example). All questions must be answered correctly to complete the respective world. Wrongly answered questions will be repeated later.

Textbox 3. Example of quiz question.

How long can humans survive without liquid intake?

A. 2-4 days

B. 1 week

C. 1 day

D. 50 days

Dissolution: We (humans) can live without solid food for more than a month, but without drinking we die after 2-4 days!

Infothek

The *Infothek* is an information section where users get access to interesting information regarding 6 nutrition-related topics, which were derived from study results described above: health, food information, beauty, sports and food, food waste, and recipes. Users get access at certain scores to get motivated to app usage and reward it. Regardless of the score, once a week access to the *Infothek* is given to re-engage the user. Information is predominantly in written form (short messages). Also, short videos and podcasts are intended.

Gamification Approach

The *C2go* app implements severely playful elements, such as points, levels, leaderboard, challenges, onboarding, feedback, progress bars and customization, to engage users.

Discussion

Overview

Mobile phone-based interventions are increasingly used to promote a healthy lifestyle [30,62-65]. In this study, a mobile phone app was considered as a very acceptable tool for the delivery of a nutrition intervention for adolescents and young adults, as mobile phone usage is widespread across all education and income levels [66]. Furthermore, adolescents and young adults are at a stage of life where their own lifestyles, including *eating styles*, are developed and established. The empirical study results described above confirm the general openness for, and interest of adolescents and young adults in, a dietary mobile app. Next, other digital health interventions were rated as helpful and satisfactory by adolescents and young adults [57,67,68] and other age groups [69,70]. Studies by others have demonstrated that mobile phone apps are widely accepted by users for intervention delivery in the field of healthy eating [30,62,71]. Despite the growing interest in mHealth research, the development of a theory-driven app is not well described in the scientific literature [30]. Furthermore, to our knowledge, no dietary mobile app has been designed to meet the needs and interests of adolescents and young adults in Germany. Our study provides a step-by-step description of how evidence (eg, from empirical studies with the target group) and theory can be translated systematically into an app concept in the field of mHealth for contributing to the prevention of NCDs.

Systematic Design Process

The concept for the C2go app was designed along the BCW and based on the input of the target group as well as the literature analysis. The BCW has been used by other researchers to guide the development of mHealth interventions [30,72].

Effective interventions need theory guidance [36,73]: Using a framework helps design interventions systematically, deriving factors that need to be changed and avoiding intervention development based on personal experiences, favorite theories, or superficial analyses [40,73,74]. Besides this, theory-based interventions help to understand which, and how, techniques are effective, and results can be used to optimize theoretical concepts. Furthermore, using theory in research is also helpful for the communication between researches and disciplines [73]. Often, the use of underlying theories and concepts in intervention trials is not well described. Theories are only mentioned as frameworks but descriptions of how they were integrated into the scientific design process are often lacking [74]. In this study, the use of the BCW as a framework permitted the systematic and comprehensive design process, which was underpinned by a model of behavior change and a behavioral diagnosis of the target behavior, before starting the design process. However, it is necessary to expand the BCW regarding the derivation of empirical study results and its translation into BCTs and, finally, into mHealth app features [30]. This could minimize the influence of individual expertise, creativity, and reasonable decision of scientists on which features should actually be implemented in the app [30].

A major strength of this study was the examination of the behavior, interests, needs, and values of the target group, because digital interventions are most engaging when they are matched to the target group's characteristics, needs, expectations, and skills [75]. Following this, other studies revealed that involving the target group throughout all phases of intervention development is important to make it more relevant to their life [30,63,68]. Our study therefore aimed at focusing on adults from disadvantaged backgrounds. Therefore, attempts were made to recruit study participants in (public) places with lower educational background in particular, for example, vocational school.

In total, 3 target behaviors were initially chosen, because concentrating on many or unspecified target behaviors (eg, whole nutritional intake) is assumed to be less effective than considering only a few and specified target behaviors (but then

intensely) [76]. The selection of the target behaviors was indirectly confirmed by the target group, as beverages, fruit, and vegetables were voted as easy-to-track food. A further advantage of choosing the 3 target behaviors is that they focus on *promotin* g individual behaviors, for example, *more* fruits, instead of forbidding food (eg, *eat only 1 piece of candy per day*).

Altogether, study results along with literature searches were used to support the behavioral diagnosis, the identification of intervention functions, and BCTs. The behavioral diagnosis revealed that certain capabilities (eg, psychological capability: awareness of consumption), opportunities, and motivational aspects are needed to establish healthy eating habits. Furthermore, study results together with evidence from the scientific literature were used to identify 5 intervention functions and 14 BCTs. The latter were translated into features of the *C2go* app.

Final Intervention: The Challenge to Go App

The *C2go* app concept targets improved drinking habits as well as increased fruit and vegetable consumption among adolescents and young adults. To this end, users choose from 3 worlds: the drinking, the vegetable, or the fruit world. A core feature of the *C2go* app concept is the use of challenges. These consist of goal-setting and self-monitoring for target behavior. Both techniques were requested by participants and are supported by evidence from the literature [36,45,58,77,78]. Focusing on BCTs for effective behavior change interventions is important. Nevertheless, considering determinants of engagement is also crucial [75]. Therefore, the *C2go* app concept implements various game elements and process motivators, which reward the process of behavior change. Examples are points, levels for status gain, rankings, or engagement loops through challenges aiming at edutainment and loyalty [33,34]. Gamification approaches can provide motivation in settings where information only is not sufficient to bring about change [34]. Various other mHealth interventions used gamification for promoting user engagement successfully [30,43,63]. Concentrating on process motivators instead of long-term, logical outcome motivators (eg, prevention of NCDs) is proposed to influence self-efficacy for behavior change positively and more effectively [79].

The individual choice of worlds and, inter alia, the setting of individual goals support customization [33] to satisfy different needs and motivation for app use (engagement). Reminders were also implemented to increase user engagement [53]. Furthermore, the implementation of an avatar will improve user engagement and acceptance [80].

Different feedback was implemented to motivate app usage [54] and behavior change. Implemented informative feedback targets intenders. This serves to boost self-efficacy, thereby helping to overcome barriers and to achieve target behavior [36,51]. Visual feedback through progress bars was used to replace possible missing intrinsic motivation for behavior change [34]. Evaluative feedback, such as *Congratulations* if challenges are passed or encouraging feedback if challenges are not passed, were implemented to increase self-efficacy [43]. Motivating messages serve to increase self-efficacy in encouraging the idea that skills for succeeding are available [34]. Evaluative, informative, and motivating feedback is given through an avatar that was implemented for identification and positive learning effects [34]. When formulating this feedback, it was important to select positive language to increase the self-efficacy and satisfaction of the user [48]. Competitive feedback comes from the community and the leaderboard [34,43].

Limitations and Future Research

Several limitations must be considered when interpreting the findings of the present approach. First, the design, development, and implementation of mHealth concepts take time. Consequently, by the time of implementation, technology and target group interests may have evolved [30]. Second, regarding the target behaviors, the app focused on drinking and fruit and vegetable consumption only. Other food groups and nutrition behaviors (eg, snacking) could be targeted in the app at a later stage of development, along with physical activity. This provides opportunities for future research and extension of the app. Third, the participants who assisted in the app development were only from 1 region in Germany, and they may have had a bigger interest in nutrition or apps as nonparticipants. Future investigation should include a more diverse group of participants.

The next step is the validation of the *C2go* app concept to demonstrate its impact on drinking and fruit and vegetable consumption, as well as its usability in a controlled intervention trial. Moreover, financial opportunities for sustainable maintenance possibilities of scientific applications must be investigated.

Conclusions

C2go is a theory-based and target group–adapted mobile intervention that was systematically developed using the BCW. *C2go* aims to improve drinking habits and the consumption of vegetables and fruit among adolescents and young adults, especially from disadvantaged backgrounds, using a gamification approach.

Acknowledgments

This research was conducted within the Competence Cluster for Nutrition and Cardiovascular Health Halle-Jena-Leipzig, which is supported by the Federal Ministry of Education and Research under grant no. 01EA1411C.

Authors' Contributions

AR designed studies 1, 2, and 3, developed the intervention, managed data collection and analysis of studies 1 and 2, and wrote the manuscript. AD codeveloped study 3, managed data collection and analysis of study 3, wrote the respective part of the

manuscript, and provided feedback on the manuscript. CB, SL, JG, and CD contributed guidance and consultation throughout the studies and discussed study designs and results. They provided feedback on the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional information on studies 1 and 2.

[PDF File (Adobe PDF File), 203KB - mhealth_v7i8e11575_app1.pdf]

Multimedia Appendix 2

Behavioral diagnosis to derive what needs to be changed to achieve the target behavior and estimation of feasibility in a dietary mobile app.

[PDF File (Adobe PDF File), 78KB - mhealth_v7i8e11575_app2.pdf]

Multimedia Appendix 3

Results of studies 1 and 2.

[PDF File (Adobe PDF File), 117KB - mhealth_v7i8e11575_app3.pdf]

Multimedia Appendix 4

Bringing together behavior change techniques and target group preferences for derivation of app features.

[PDF File (Adobe PDF File), 54KB - mhealth_v7i8e11575_app4.pdf]

References

1. World Health Organization. 2014. Global Status Report On Noncommunicable Diseases 2014 URL: https://tinyurl.com/yyxm5ym6 [accessed 2019-07-18]
2. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe 2014: epidemiological update. Eur Heart J 2014 Nov 7;35(42):2950-2959. [doi: 10.1093/eurheartj/ehu299] [Medline: 25139896]
3. Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. J Am Med Assoc 2017 Dec 7;317(9):912-924 [FREE Full text] [doi: 10.1001/jama.2017.0947] [Medline: 28267855]
4. DGK – Leitlinien - Deutsche Gesellschaft für Kardiologie – Herz. 2012. ESC Pocket Guidelines: Prävention von Herz-Kreislauf-Erkrankungen URL: https://leitlinien.dgk.org/files/PL_Pr%c3%a4vention_Internet_13.pdf [accessed 2019-07-18]
5. Umer A, Kelley GA, Cottrell LE, Giacobbi P, Innes KE, Lilly CL. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. BMC Public Health 2017 Dec 29;17(1):683 [FREE Full text] [doi: 10.1186/s12889-017-4691-z] [Medline: 28851330]
6. Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization; 2011.
7. Bhupathiraju SN, Tucker KL. Coronary heart disease prevention: nutrients, foods, and dietary patterns. Clin Chim Acta 2011 Aug 17;412(17-18):1493-1514 [FREE Full text] [doi: 10.1016/j.cca.2011.04.038] [Medline: 21575619]
8. Gan Y, Tong X, Li L, Cao S, Yin X, Gao C, et al. Consumption of fruit and vegetable and risk of coronary heart disease: a meta-analysis of prospective cohort studies. Int J Cardiol 2015 Mar 15;183:129-137. [doi: 10.1016/j.ijcard.2015.01.077] [Medline: 25662075]
9. Bhupathiraju SN, Wedick NM, Pan A, Manson JE, Rexrode KM, Willett WC, et al. Quantity and variety in fruit and vegetable intake and risk of coronary heart disease. Am J Clin Nutr 2013 Dec;98(6):1514-1523 [FREE Full text] [doi: 10.3945/ajcn.113.066381] [Medline: 24088718]
10. Brown CM, Dulloo AG, Montani JP. Sugary drinks in the pathogenesis of obesity and cardiovascular diseases. Int J Obes (Lond) 2008 Dec;32(Suppl 6):S28-S34. [doi: 10.1038/ijo.2008.204] [Medline: 19079277]
11. Bray GA. Energy and fructose from beverages sweetened with sugar or high-fructose corn syrup pose a health risk for some people. Adv Nutr 2013 Mar 1;4(2):220-225 [FREE Full text] [doi: 10.3945/an.112.002816] [Medline: 23493538]
12. American Heart Association. 2017. Diet Drinks and Possible Association With Stroke and Dementia; Current Science Suggests Need for More Research: American Heart Association Stroke Journal Report URL: https://tinyurl.com/yxr65ups [accessed 2019-07-18] [WebCite Cache ID 70mPwgAMA]

13. Nseir W, Nassar F, Assy N. Soft drinks consumption and nonalcoholic fatty liver disease. World J Gastroenterol 2010 Jun 7;16(21):2579-2588 [FREE Full text] [doi: 10.3748/wjg.v16.i21.2579] [Medline: 20518077]
14. Vartanian LR, Schwartz MB, Brownell KD. Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. Am J Public Health 2007 Apr;97(4):667-675. [doi: 10.2105/AJPH.2005.083782] [Medline: 17329656]
15. Max Rubner-Institut. 2008. Ergebnisbericht Teil 1: Nationale Verzehrsstudie Ii: Die Bundesweite Befragung Zur Ernährung Von Jugendlichen Und Erwachsenen URL: https://tinyurl.com/yxg9adoh [WebCite Cache ID 70sFLZxY8]
16. Schienkiewitz AK, Brettschneider SD, Schaffrath RA. [Overweight and obesity in childhood and adolescence in Germany? Cross-section results from KiGGS Wave 2 and Trends]. J Health Monit 2018;3(1):16-23. [doi: 10.17886/RKI-GBE-2018-005]
17. Borrmann A, Mensink GB, KiGGS Study Group. [Fruit and vegetable consumption by children and adolescents in Germany: results of KiGGS wave 1]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Sep;58(9):1005-1014. [doi: 10.1007/s00103-015-2208-4] [Medline: 26141246]
18. Mensink GB, Schienkiewitz A, Rabenberg M, Borrmann A, Richter A, Haftenberger M. Consumption of sugary soft drinks among children and adolescents in Germany. Results of the cross-sectional KiGGS wave 2 study and trends. J Health Monit 2018;3(1). [doi: 10.17886/RKI-GBE-2018-007]
19. Deutsche Gesellschaft für Ernährung. 2018. Vollwertig Essen Und Trinken Nach Den 10 Regeln Der DGE URL: https://www.dge.de/ernaehrungspraxis/vollwertige-ernaehrung/10-regeln-der-dge/ [WebCite Cache ID 70nawdnGb]
20. Health.gov. 2015. Dietary Guidelines for Americans 2015-2020 URL: https://health.gov/dietaryguidelines/2015/resources/2015-2020_Dietary_Guidelines.pdf [WebCite Cache ID 74IMEvWH2]
21. Gu X, Tucker KL. Dietary intakes of the US child and adolescent population and their adherence to the current dietary guidelines: trends from 1999 to 2012. FASEB J 2017;31(1):29.1 [FREE Full text]
22. Darmon N, Drewnowski A. Does social class predict diet quality? Am J Clin Nutr 2008 May;87(5):1107-1117. [doi: 10.1093/ajcn/87.5.1107] [Medline: 18469226]
23. Heuer T, Krems C, Moon K, Brombach C, Hoffmann I. Food consumption of adults in Germany: results of the German national nutrition survey II based on diet history interviews. Br J Nutr 2015 May 28;113(10):1603-1614 [FREE Full text] [doi: 10.1017/S0007114515000744] [Medline: 25866161]
24. Mensink GB, Kleiser C, Richter A. [Food consumption of children and adolescents in Germany. Results of the German health interview and examination survey for children and adolescents (KiGGS)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2007;50(5-6):609-623. [doi: 10.1007/s00103-007-0222-x] [Medline: 17514445]
25. Mensink GB, Heseker H, Stal A, Richter A, Vohmann C, Koch R. Ernährungs-Umschau. 2007. Die Aktuelle Nährstoffversorgung Von Kindern Und Jugendlichen in Deutschland: Ergebnisse Aus Eskimo URL: https://www.ernaehrungs-umschau.de/fileadmin/Ernaehrungs-Umschau/pdfs/pdf_2007/11_07/EU11_636_646.qxd.pdf
26. Bartsch S. Ernährungs-Umschau. 2010. Jugendesskultur URL: https://www.ernaehrungs-umschau.de/fileadmin/Ernaehrungs-Umschau/pdfs/pdf_2010/08_10/EU08_2010_432_438.qxd.pdf
27. Albrecht UV. Federal Ministry of Health (Germany). 2016. Chancen und Risiken von Gesundheits-Apps (CHARISMHA) URL: https://tinyurl.com/y66aspjr [WebCite Cache ID 74JUDD5K9]
28. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 5;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
29. 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 Dec 4;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
30. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
31. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR Mhealth Uhealth 2018 Mar 13;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
32. Rohde A, Lorkowski S, Dawczynski C, Brombach C. Dietary mobile apps: acceptance among young adults: a qualitative study. Ernahrungs Umschau 2017;64(2):36-43. [doi: 10.4455/eu.2017.007]
33. Zichermann G, Cunningham C. Gamification By Design: Implementing Game Mechanics In Web And Mobile Apps. Sebastopol, California: O'Reilly Media; 2011.
34. Berger V, Schrader U. Fostering sustainable nutrition behavior through gamification. Sustainability 2016 Jan 12;8(1):67. [doi: 10.3390/su8010067]
35. 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]
36. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]

37. Lin CY, Scheerman JF, Yaseri M, Pakpour AH, Webb TL. A cluster randomised controlled trial of an intervention based on the Health Action Process Approach for increasing fruit and vegetable consumption in Iranian adolescents. Psychol Health 2017 Dec;32(12):1449-1468. [doi: 10.1080/08870446.2017.1341516] [Medline: 28681612]
38. Storm V, Dörenkämper J, Reinwand DA, Wienert J, de Vries H, Lippke S. Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. J Med Internet Res 2016 Apr 11;18(4):e78 [FREE Full text] [doi: 10.2196/jmir.5147] [Medline: 27068880]
39. Brookie KL, Mainvil LA, Carr AC, Vissers MC, Conner TS. The development and effectiveness of an ecological momentary intervention to increase daily fruit and vegetable consumption in low-consuming young adults. Appetite 2017 Dec 1;108:32-41. [doi: 10.1016/j.appet.2016.09.015] [Medline: 27642037]
40. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. Croydon, England: Silverback Publishing; 2014.
41. Google Play. 2018. Was ich esse App URL: https://play.google.com/store/apps/details?id=de.aid.android.ernaehrungspyramide&hl=de [accessed 2019-07-17] [WebCite Cache ID 75c4v0HNU]
42. Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. Dritte Edition. Weinheim, Basel: Beltz Juventa; 2014.
43. Mummah SA, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 8;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
44. Zhou G, Gan Y, Miao M, Hamilton K, Knoll N, Schwarzer R. The role of action control and action planning on fruit and vegetable consumption. Appetite 2015 Aug;91:64-68. [doi: 10.1016/j.appet.2015.03.022] [Medline: 25819607]
45. van Achterberg T, de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162 [FREE Full text] [doi: 10.1093/heapro/daq050] [Medline: 20739325]
46. Harkin B, Webb TL, Chang BP, Prestwich A, Conner M, Kellar I, et al. Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychol Bull 2016 Feb;142(2):198-229. [doi: 10.1037/bul0000025] [Medline: 26479070]
47. Dusseldorp E, van Genugten L, van Buuren S, Verheijden MW, van Empelen P. 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]
48. van de Ridder JM, Peters CM, Stokking KM, de Ru JA, ten Cate OT. Framing of feedback impacts student's satisfaction, self-efficacy and performance. Adv Health Sci Educ Theory Pract 2015 Aug;20(3):803-816. [doi: 10.1007/s10459-014-9567-8] [Medline: 25542198]
49. Prestwich A, Kellar I, Parker R, MacRae S, Learmonth M, Sykes B, et al. How can self-efficacy be increased? Meta-analysis of dietary interventions. Health Psychol Rev 2014;8(3):270-285. [doi: 10.1080/17437199.2013.813729] [Medline: 25053214]
50. Poínhos R, van der Lans IA, Rankin A, Fischer AR, Bunting B, Kuznesof S, et al. Psychological determinants of consumer acceptance of personalised nutrition in 9 European countries. PLoS One 2014;9(10):e110614 [FREE Full text] [doi: 10.1371/journal.pone.0110614] [Medline: 25334009]
51. Godinho CA, Alvarez M, Lima ML. Formative research on HAPA model determinants for fruit and vegetable intake: target beliefs for audiences at different stages of change. Health Educ Res 2013 Dec;28(6):1014-1028. [doi: 10.1093/her/cyt076] [Medline: 23856178]
52. Kessels LT, Ruiter RA, Brug J, Jansma BM. The effects of tailored and threatening nutrition information on message attention. Evidence from an event-related potential study. Appetite 2011 Feb;56(1):32-38. [doi: 10.1016/j.appet.2010.11.139] [Medline: 21094195]
53. Berkovsky S, Hendrie G, Freyne J, Noakes M, Usic K. The HealthierU portal for supporting behaviour change and diet programs. In: Georgiou A, Grain H, Schaper LK, editors. Driving Reform: Digital Health Is Everyone's Business: Selected Papers from the 23rd Australian National Health Informatics Conference. Amsterdam: IOS Press Incorporated; 2015:15-21.
54. Baranowski MT, Bower PK, Krebs P, Lamoth CJ, Lyons EJ. Effective feedback procedures in games for health. Games Health J 2013 Dec;2(6):320-326. [doi: 10.1089/g4h.2013.1328] [Medline: 26197072]
55. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 6;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
56. Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G. Financial incentive-based approaches for weight loss: a randomized trial. J Am Med Assoc 2008 Dec 10;300(22):2631-2637 [FREE Full text] [doi: 10.1001/jama.2008.804] [Medline: 19066383]
57. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to health: a novel mhealth intervention using FitBits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2017 Jan;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]

58. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Dec 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
59. Lara J, Evans EH, O'Brien N, Moynihan PJ, Meyer TD, Adamson AJ, et al. Association of behaviour change techniques with effectiveness of dietary interventions among adults of retirement age: a systematic review and meta-analysis of randomised controlled trials. BMC Med 2014 Oct 7;12:177 [FREE Full text] [doi: 10.1186/s12916-014-0177-3] [Medline: 25288375]
60. König LM, Renner B. Colourful = healthy? Exploring meal colour variety and its relation to food consumption. Food Qual Prefer 2018 Mar;64:66-71. [doi: 10.1016/j.foodqual.2017.10.011]
61. Werch CE, Moore MJ, Bian H, DiClemente CC, Ames SC, Weiler RM, et al. Efficacy of a brief image-based multiple-behavior intervention for college students. Ann Behav Med 2008 Oct;36(2):149-157 [FREE Full text] [doi: 10.1007/s12160-008-9055-6] [Medline: 18800217]
62. Mummah S, Robinson TN, Mathur M, Farzinkhou S, Sutton S, Gardner CD. Effect of a mobile app intervention on vegetable consumption in overweight adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Dec 15;14(1):125 [FREE Full text] [doi: 10.1186/s12966-017-0563-2] [Medline: 28915825]
63. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016 May 13;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
64. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug 22;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
65. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
66. Google Services. 2015. Global Connected Consumer Study - Ergebnisse Für Deutschlandine Studie Von TNS Infratest in Kooperation Mit Dem BVDW Und Google URL: https://services.google.com/fh/files/misc/global-connected-consumer-studie-deutschland.pdf [accessed 2019-07-19]
67. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
68. Nollen NL, Hutcheson T, Carlson S, Rapoff M, Goggin K, Mayfield C, et al. Development and functionality of a handheld computer program to improve fruit and vegetable intake among low-income youth. Health Educ Res 2013 Apr;28(2):249-264 [FREE Full text] [doi: 10.1093/her/cys099] [Medline: 22949499]
69. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
70. Mummah SA, Mathur M, King AC, Gardner CD, Sutton S. Mobile technology for vegetable consumption: a randomized controlled pilot study in overweight adults. JMIR Mhealth Uhealth 2016 May 18;4(2):e51 [FREE Full text] [doi: 10.2196/mhealth.5146] [Medline: 27193036]
71. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
72. Robinson E, Higgs S, Daley AJ, Jolly K, Lycett D, Lewis A, et al. Development and feasibility testing of a smart phone based attentive eating intervention. BMC Public Health 2013 Jul 9;13:639 [FREE Full text] [doi: 10.1186/1471-2458-13-639] [Medline: 23837771]
73. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behaviour Change Theories: An Essential Resource for Researchers, Policy Makers and Practitioners. Croydon: Silverback Publishing; 2014.
74. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
75. Short C, Rebar AL, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Eur Health Psychol 2015;17(1):32-38.
76. Atkins L, Michie S. Designing interventions to change eating behaviours. Proc Nutr Soc 2015 May;74(2):164-170. [doi: 10.1017/S0029665115000075] [Medline: 25998679]
77. Pearson ES. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. Patient Educ Couns 2012 Apr;87(1):32-42. [doi: 10.1016/j.pec.2011.07.018] [Medline: 21852063]
78. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]

79. Robinson TN. Stealth interventions for obesity prevention and control. In: Dube L, Bechara A, Drewnowski A, LeBel J, James P, Yada RY, editors. Obesity Prevention: The Role of Brain and Society on Individual Behavior. First Edition. Amsterdam: Academic Press; 2011:319-327.
80. Lisetti C, Yasavur U, Leon C, Amini R, Rishe N, Visser U. Building an On-Demand Avatar-Based Health Intervention for Behavior Change. In: Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference. 2012 Presented at: FLAIRS'12; May 23-25, 2012; Marco Island, Florida p. 175-180.

Abbreviations

APEASE: affordable, practicable, effective, acceptable, side-effects, equity
BCT: behavior change technique
BCW: behavior change wheel
C2go: Challenge to go
COM-B: capability-opportunity-motivation-behavior
HAPA: health action process approach
mHealth: mobile health
NCD: noncommunicable disease

Edited by G Eysenbach; submitted 13.07.18; peer-reviewed by P Lunde, L König; comments to author 30.08.18; revised version received 23.01.19; accepted 04.03.19; published 12.08.19.

Please cite as:
Rohde A, Duensing A, Dawczynski C, Godemann J, Lorkowski S, Brombach C
An App to Improve Eating Habits of Adolescents and Young Adults (Challenge to Go): Systematic Development of a Theory-Based and Target Group–Adapted Mobile App Intervention
JMIR Mhealth Uhealth 2019;7(8):e11575
URL: http://mhealth.jmir.org/2019/8/e11575/
doi:10.2196/11575
PMID:30903746

Original Paper

Childrens' and Parents' Willingness to Join a Smartphone-Based Emergency Response Community for Anaphylaxis: Survey

Michael Khalemsky[1], MBA; David G Schwartz[1], PhD, MBA; Tamar Silberg[2,3], PhD; Anna Khalemsky[1], MBA; Eli Jaffe[4], PhD; Raphael Herbst[4], MEM

[1]Graduate School of Business Administration, Bar-Ilan University, Ramat Gan, Israel

[2]Department of Psychology, Bar-Ilan University, Ramat Gan, Israel

[3]Edmond and Lily Safra Children's Hospital, Tel Hashomer, Ramat Gan, Israel

[4]Israel Magen David Adom, Tel Aviv, Israel

Corresponding Author:
Michael Khalemsky, MBA
Graduate School of Business Administration, Bar-Ilan University
Bar-Ilan University
Ramat Gan, 5290002
Israel
Phone: 972 523499011
Fax: 972 25695334
Email: khalemsky@gmail.com

Abstract

Background: Medical emergencies such as anaphylaxis may require immediate use of emergency medication. Because of the low adherence of chronic patients (ie, carrying anti-anaphylactic medication) and the potentially long response time of emergency medical services (EMSs), alternative approaches to provide immediate first aid are required. A smartphone-based emergency response community (ERC) was established for patients with allergies to enable members to share their automatic adrenaline injector (AAI) with other patients who do not have their AAI at the onset of anaphylactic symptoms. The community is operated by a national EMS. In the first stage of the trial, children with food allergies and their parents were invited to join.

Objective: This study aimed to identify the factors that influence the willingness to join an ERC for a group of patients at risk of anaphylaxis.

Methods: The willingness to join an ERC was studied from different perspectives: the willingness of children with severe allergies to join an ERC, the willingness of their parents to join an ERC, the willingness of parents to enroll their children in an ERC, and the opinions of parents and children about the minimum age to join an ERC. Several types of independent variables were used: demographics, medical data, adherence, parenting style, and children's autonomy. A convenience sample of children and their parents who attended an annual meeting of a nonprofit organization for patients with food allergies was used.

Results: A total of 96 questionnaires, 73 by parents and 23 by children, were collected. Response rates were approximately 95%. Adherence was high: 22 out of 23 children (96%) and 22 out of 52 parents (42%) had their AAI when asked. Willingness to join the community was high among parents (95%) and among children (78%). Willingness of parents to enroll their children was 49% (36/73). The minimum age to join an ERC was 12.27 years (SD 3.02) in the parents' opinion and 13.15 years (SD 3.44) in the children's opinion.

Conclusions: Parents' willingness to join an ERC was negatively correlated with parents' age, child's age, and parents' adherence. This can be explained by the *free-rider effect*: parents who carried an AAI for their young child, but had low adherence, wanted to join the ERC to get an additional layer of emergency response. Children's willingness to join the community was positively correlated with age and negatively correlated with the child's emotional autonomy. Parents' willingness to enroll their children in an ERC was positively correlated with child's age and negatively correlated with parents' adherence: again, this can be explained by the aforementioned *free-rider effect*. Parents' and children's opinions about the minimum age to join an ERC were negatively correlated with protective parenting style and positively correlated with monitoring parenting style.

(JMIR Mhealth Uhealth 2019;7(8):e13892) doi:10.2196/13892

KEYWORDS

mHealth; emergency; volunteer; community; smartphone; anaphylaxis; epinephrine

Introduction

Objectives of the Study

This study examines different factors affecting willingness of patients and parental-caregivers to join a smartphone-based emergency response community (ERC) for patients with allergy at a risk of anaphylaxis.

Emergency Response Communities

A medical emergency is "an acute injury or illness that poses an immediate risk to a person's life or long-term health" [1]. Nontraumatic medical emergencies include conditions such as stroke, severe asthma attack, heart attack, anaphylactic shock, hypoglycemic coma, and drug overdose. The immediate provision of first aid is a crucial factor in lowering mortality and improving long-term prognosis [2,3].

Emergency medical services (EMSs) are the primary provider of first aid to people in medical emergencies that occur outside medical institutions [4,5]. Unfortunately, there is no ambulance on every street corner, and patients in distress have to wait for help. The response times of EMSs vary significantly across countries and geographies, such as rural versus urban areas [6-8].

EMS organizations and health policy makers try to achieve faster response times through various approaches. These include the deployment of automatic electronic defibrillators in public places [9-11], the use of drones to deliver emergency equipment [12], and the establishment of networks of first responders and volunteers to provide first aid in different medical situations [13-18].

An ERC [19,20] is a social network of patients who are prescribed to carry life-saving medication for themselves and can potentially help other patients who are without medication in a medical emergency. Central servers track the location of different community members and locate members who are carrying the required medication and are in the vicinity of the patient in distress. The social network is regulated: the identities and medical records of the patients are verified by their doctors, and the provision of medication in an emergency is approved in real time by EMS personnel. As such, the social network is an *EMS-mediated community of laypersons*.

Joining an ERC requires adoption of a dedicated mobile health (mHealth) smartphone app. mHealth is defined as "healthcare to anyone, anytime and anywhere by removing temporal and locational constraints" [21]. New ERC members have to complete a registration process and agree to location tracking when they are available to respond. The adoption of mobile apps has been widely studied in the past decade—both in general and specifically among patients [22-24].

Willingness to Join a Mutual Aid Community

ERCs are a type of mutual aid community with members being both potential givers and takers—providers and recipients—of an emergency response. Joining a mutual aid community is a type of volunteerism. Altruism is the most frequently expressed motive for volunteerism. Yet people may volunteer for reasons other than pure altruism. For example, parents may volunteer in an organization from which their children directly benefit [25]. Self-identification as religiously observant is also associated with a higher willingness to join a mutual aid community [26].

The phenomenon of *bystander intervention* has been widely studied over the past 5 decades [27-29]. Mutual aid communities exist in different areas such as among drug addicts [26,30], mental health patients [30,31], and diabetics [32,33].

Another important phenomenon that may influence willingness to join an ERC is *shared identity*: people tend to help those with whom they share something in common [34-36]. As ERC members all share the same medical condition, they may be influenced by this phenomenon—sometimes referred to as *patients like me*.

Prescription Medication Sharing

There are 2 types of medication sharing: recreational (ie, abusive medication sharing to experience nonmedical effects) and nonrecreational (medication sharing for medical treatment) [37]. ERC is designed to facilitate nonrecreational medication sharing in emergency settings. Prescription medication sharing has been studied for decades [38-40]. A recent meta-analysis, drawing on several studies suggesting that gender, age, income, and use of the internet to access health-related information influence patients' willingness to share medication [37], reports a wide range of prevalence of the borrowing and lending of prescription medication (5%-52%). Additional studies report the willingness of parents to share or borrow medication related to their child's medical condition [41]. A recent survey on automatic adrenaline injector (AAI) sharing shows that 76.6% of AAI carriers expressed willingness to share their AAI right away. Respondents who would not share their AAI were concerned about the potential harm to the patient (eg, misdiagnosis or overdose) and to their child (eg, having no medication left or delay in obtaining a refill) [42]. Joining an ERC indicates readiness to share a personal prescription medication with a stranger.

Parenting Style

To the best of our knowledge, no study has examined the correlation between parents' willingness to join an ERC and parenting styles associated with medical decision making [43]. Parenting styles are characteristics that represent how parents relate to and place demands on their children [44]. For example, an overprotective parenting style in a family with food allergies is a common coping method that can inhibit a child's autonomy and lead to the child's emotional distress over his or her medical condition [45]. The Adult Responses to Children's Symptoms (ARCS) questionnaire [46] is intended to measure specific health-related parenting practices. The ARCS identifies 3 distinct parenting styles: (1) the protective style, where the parents engage in caretaking behaviors that place the child in a

passive patient role, (2) the dismissive style, where the parents criticize the child's health complaints, and (3) the monitoring style, where the parents encourage the child's autonomy while monitoring the child's symptoms [47].

Child's Autonomy

Autonomy refers to a person's ability to act on his or her own values and interest. Taken from ancient Greek, the word autonomy means *self-governance*. From a psychological view, autonomy is made up of a set of functional skills, emotional responses, and attitudes [48]. To act, feel, and reason, the autonomous person must have a sense of self-worth and self-respect.

Some experiences that children and adolescents with food allergies undergo may put them at risk of problems related to the development of their autonomy. Studies show that limiting young children's opportunities for independent exploration of their environment can interfere with the development of their autonomy [49]. Children with severe food allergies, especially those who have experienced a severe anaphylactic reaction in the past, are likely to be restricted in their activities [50]. On the other hand, adolescents at risk of anaphylaxis are likely to take an active role in managing their allergies [51]. Thus, the findings are mixed: although having restrictions placed on one's activities is negatively related to the development of autonomy, self-management is positively related to its development.

In this study, we expand the understanding of the sense of autonomy among children with severe allergies. Subsequently, we will refer to autonomy as an agency consisting of 3 components [48]: (1) *attitudinal autonomy*, which refers to the cognitive process of listing one's possibilities and making a choice between different options; (2) *emotional autonomy*, which refers to confidence and trust in defining goals independent of the wishes of parents and peers; and (3) *functional autonomy*, which describes the process of developing a strategy to achieve one's goals by means of self-regulation and self-control.

Description of the field study

EPIMADA is an ERC launched in January 2018 by the Israel National EMS, Magen David Adom (MDA), in cooperation with Bar-Ilan University. EPIMADA comprises patients with allergies who are required to carry an AAI as the first-line treatment against anaphylaxis [52]. Members are equipped with a mobile app that tracks their location and notifies them about relevant emergencies in their vicinity. Members can set preferences such as days of the week and hours of the day during which they are available for dispatch. In an emergency, members are dispatched by the MDA-EMS command center and receive additional real-time instructions from a trained dispatcher-paramedic by phone.

Methods

We studied the willingness to join an ERC from different points of view:

- Willingness of parents to join an ERC
- Willingness of parents to enroll their children in an ERC
- Willingness of children to join an ERC
- Opinions of parents and children about the minimum age to join an ERC

Independent Variables

We used several types of independent variables:

- Demographic data (parents' age, child's age, parents' gender, child's gender, parents' level of religious observance, and parent's years of education).
- Medical data about the child (time since diagnosis, time since last anaphylactic attack, and the number of anaphylactic attacks in the past) and medical data about the parents (whether they themselves are allergy patients).
- Adherence (carrying the AAI) both by parents and by children.
- Parenting styles (protective, dismissive, and monitoring).
- Children's autonomy (attitudinal, emotional, and functional).

Collection of Data

We used a convenience sample of children diagnosed with severe food allergies and their parents who attended an annual conference of a nonprofit organization of patients with food allergies. All parents were asked to fill out a written questionnaire for parents (Multimedia Appendix 1), and all children older than 8 years were asked to fill out a written questionnaire for children (Multimedia Appendix 2). The questionnaires included a brief description of the EPIMADA project followed by questions about basic parent and child demographics and condition-related factors such as whether the parents are patients themselves, time since diagnosis, adherence level, and emergency events in the past. Parents and children were asked about their willingness to join an ERC; in addition, parents were asked about their willingness to enroll their children. Parents and children were asked for their opinion about the minimum age to join an ERC. Parents completed the adults' version of ARCS questionnaire, and children completed the children's version of ARCS questionnaire [46], the Adolescent Autonomy Questionnaire (AAQ) [48], and the Food Allergy Independent Measure (FAIM)[53].

Given that a convenience sample was used, the research is descriptive and does not presume to predict the behavior of the general population.

A total of 23 parent-child pairs attended the conference and answered the parents' and the children's questionnaires (the children had to be at least 8 years old and attend the conference with their parents). All children's questionnaires were paired with same-family parents' questionnaires through a coding system that maintained anonymity. A total of 50 parents attended without their children and answered the parents' questionnaires. The response rates were about 95% (2-3 parents arrived too late and were not able to answer the questionnaire because the conference had started. No one refused to answer the questionnaires.). These activities were performed at the start of the conference before participants were informed of the EPIMADA initiative.

Analytical Techniques

In addition to descriptive statistics, we used several analytical tools to explore our data:

- *t* tests were used, with the assumption of normal distribution, to check if there are significant differences between 2 groups, for example, parents and their children (paired samples) and clusters of parents (independent samples).
- Mann-Whitney nonparametric *U* tests were used as an alternative to *t* tests without the assumption of normal distribution.
- Chi-square independence test was used to check if there is significant association between 2 nominal variables.
- Pearson correlation analysis was used to discover correlations between different variables to plan regression models and to avoid multicollinearity.
- One-way analysis of variances (ANOVAs) were used to check whether there are significant differences between multiple samples, for example, 3 clusters.
- Intraclass correlation (ICC) tests were used to check the consistency of measures between parents and children in the 3 parenting styles.
- Linear regressions were used for scale-dependent variables.
- Ordinal regressions were used for ordinal-dependent variables.
- Binary logistic regressions were used for binary-dependent variables.
- Bootstrapping is a resampling technique that improves the property estimation in small samples. This technique was applied to logistic regressions that initially did not provide significant results.
- Principal component analysis (PCA) was used to find the mix of possibly correlated variables for dimension reduction for cluster analysis.
- Cluster analysis was used in unsupervised learning to enable identification of homogeneous groups without a target attribute by identifying the similarities between objects for a given number of subgroups [54]. This method allows interpretation of the results without relying on an existing target attribute.
- Classification tree (J48) analysis was used in supervised learning with known target variable to enable identification of the most influential variables.

Institutional Review Board Approval

The research was approved by the Institutional Review Board of Bar-Ilan University and by the Research Committee of MDA.

Software Tools

The data were converted to digital form by the researchers and analyzed using IBM SPSS 24 software and WEKA 3.7.11 software developed by the University of Waikato in New Zealand.

Results

Demographic Parameters of the Sample

A total of 57 (78%) parents were female and 16 (22%) were male. A total of 15 (20.5%) parents reported that they are religiously observant. Tables 1-3 present several demographic parameters of the sample.

Table 1. Demographic statistics of parents (n=73).

Parameter	Average	Median	Mode	SD	Min	Max	IQR[a]
Age (N=73)	40.51	40	39	7.15	22	55	9.5 (35.5-45)
Years of education (N=72[b])	15.74	15.5	15	2.40	12	25	2 (15-17)
Number of children (N=73)	2.49	3	3	0.97	1	6	1 (2-3)

[a]IQR: interquartile range.

[b]These data were missing in 1 questionnaire.

Table 2. Age of children.

Parameter	Average	Median	Mode	SD	Min	Max	IQR[a]
Children[b] (all; N=73)	9.01	8.5	14, 17	5.52	1	21	10 (4-14)
Children[c] (attended; N=23)	13.69	14	14	3.72	8	21	6 (11-17)

[a]IQR: interquartile range.

[b]Children's data were provided by all parents about their children.

[c]A total of 23 children attended the conference and filled out children's questionnaires. The statistics of these children (part of the total sample of 73 children) are based on data reported by their parents. Participation was limited to children aged at least 8 years.

Table 3. Gender of children.

Population	Female, n (%)	Male, n (%)
Children (all, N=73)	49 (68)	23 (32)
Children (attended; N=23)	7 (30)	16 (70)

Medical Statistics

Medical statistical data of children are provided in Table 4.

Adherence

Parents reported who carries their child's AAI: in 14 (19%) cases, only the parents carried an AAI; in 19 (26%) cases, only the child carried an AAI; and in 38 (52%) cases, both the parents and the child carried an AAI (in 2 cases, no data were provided).

A total of 52 parents who carried an AAI for their children were asked 3 questions about their own adherence, 57 parents whose children carried an AAI were asked 3 questions about their children's adherence, and 23 children who attended the conference were asked 2 questions about their adherence. Tables 5-7 present the reports.

We compared the parents' answers to their children's answers. A total of 2 children answered "Yes" to the question "Are you carrying an AAI now?" whereas their parents answered "No" to the question "Is your child carrying an AAI now?" When parents and children answered the question "How many days last week did your child have immediate access to an AAI throughout the day?" in 3 cases, parents reported higher adherence (6 vs 4, 7 vs 6, and 7 vs 5) than their children, and in 3 cases parents reported lower adherence than their children (6 vs 7, 5 vs 7, and 1 vs 7).

Table 4. Medical statistics of children.

Parameter and N (valid[a,b])	Average	Median	Mode	SD	Min	Max	IQR[c]
Time since anaphylaxis diagnosis (years)							
62	8.13	7	2	5.59	1	22	9 (3-12)
20	12.45	12	—[d]	5.36	1	22	7.25 (9.25-16.50)
Time since last anaphylactic attack (years)							
52	4.85	4	1	3.92	1	14	6 (1-7)
21	6.62	7	1	4.61	1	14	9.5 (1-10.5)
Number of anaphylactic attacks							
70	1.74	1	1	1.77	0	10	2 (1-3)
22	2.41	2	1	2.15	0	10	2.25 (1-3.25)

[a]Data for all children are reported in the upper row for each variable and data for the children that attended the conference are reported in the lower row.

[b]These data were missing in 1 questionnaire.

[c]IQR: interquartile range.

[d]Multiple modes exist.

Table 5. First question about adherence.

Question	Never, n (%)	Seldom, n (%)	Often, n (%)	Always, n (%)	No answer, n (%)
Reports by parents					
Do you make sure to carry the AAI[a]? (N=52)	1 (2)	1 (2)	3 (6)	45 (86)	2 (4)
Does your child make sure to carry the AAI? (N=57)	2 (4)	0 (0)	10 (18)	43 (75)	2 (3)

[a]AAI: automatic adrenaline injector.

Table 6. Second question about adherence.

Question	Yes, n (%)	No, n (%)	No answer, n (%)
Reports by parents			
Are you (parents) carrying an AAI[a] now? (N=52)	22 (42)	30 (58)	0 (0)
Is your child carrying an AAI now? (N=57)	53 (93)	3 (5)	1 (2%)
Reports by chilren			
Are you (child) carrying an AAI now? (N=23)	22 (96)	1 (4)	0 (0)

[a]AAI: automatic adrenaline injector.

Table 7. Third question about adherence.

Question	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	6, n (%)	7, n (%)
Reports by parents							
How many days last week did you have immediate access to an AAI[a] throughout the day?[b] (N=52)	9 (17)	2 (4)	1 (2)	2 (4)	0 (0)	1 (2)	37 (71)
How many days last week did your child have immediate access to an AAI throughout the day?[b] (N=57)	1 (2)	0 (0)	0 (0)	1 (2)	1 (2)	3 (5)	51 (89)
Reports by children							
How many days last week did you have immediate access to an AAI throughout the day?[c] (N=23)	0 (0)	0 (0)	0 (0)	1 (4.3)	1 (4.3)	1 (4.3)	20 (87)

[a]AAI: automatic adrenaline injector.

[b]Reports by parents.

[c]Reports by children.

Parenting Styles

Parenting styles as assessed by the ARCS questionnaires answered by parents and their children are presented in Table 8.

We performed a paired samples *t* test to compare parenting style assessments based on answers of parents whose children attended the conference with parenting style assessments based on answers of their children who attended the conference. No significant differences were found. A medium positive correlation (R=0.451, P=.03) was observed between the reports of the parents and their children about the dismissive parenting style. The correlations for other parenting styles were not significant at the 5% significance level. We also performed an ICC test for the 3 parenting styles to check the consistency of measures between parents and children. For the protective parenting style, the ICC value was poor (.319) and not significant (P=.06); for the dismissive parenting style, the ICC value was fair (.45) and significant (P=.01); and for the monitoring parenting style, the ICC value was poor (.103) and not significant (P=.31).

We compared our findings with the data reported by Van Slyke and Walker [47], using a summary independent samples *t* test and found significant differences at the 5% significance level (P<.001) in the protective parenting style (Van Slyke and Walker's results: mean 1.37, SD 0.63) and no significant differences in the dismissive (P=.23) and monitoring (P=.25) parenting styles.

Child's Autonomy

Results related to attitudinal, emotional, and functional autonomy among children who attended the conference are given in Table 9.

Table 8. Parenting styles as assessed by the Adult Responses to Children's Symptoms questionnaires answered by parents and their children.

Parenting style and respondents	Average	Median	Mode	SD	Min	Max	IQR[a]
Protective							
All parents (N=73)	2.39	2.4	2.4	0.69	0.87	3.67	1.06 (1.87-2.93)
Parents whose children attended (N=23)	2.34	2.33	1.67	0.78	0.87	3.6	1.4 (1.67-3.07)
Children who attended (N=23)	2.27	2.27	2.93	0.71	1	3.47	1.2 (1.73-2.93)
Dismissive							
All parents (N=73)	0.88	0.83	0.33	0.58	0	2.67	1 (0.33-1.33)
Parents whose children attended (N=23)	0.74	0.67	0.67	0.62	0	2.17	0.67 (0.33-1)
Children who attended (N=23)	0.93	0.67	0.5	0.58	0.33	2.17	1 (0.5-1.5)
Monitoring							
All parents (N=73)	2.91	3	__[b]	0.61	0.88	3.88	0.88 (2.5-3.38)
Parents whose children attended (N=23)	2.94	3.13	3.13	0.61	1.88	3.88	1.12 (2.38-3.5)
Children who attended (N=23)	2.63	2.63	2.63	0.71	1	3.63	0.87 (2.38-3.25)

[a]IQR: interquartile range.

[b]Multiple modes exist.

Table 9. Attitudinal, emotional, and functional autonomy among children who attended the conference (n=23).

Autonomy	Average	Median	Mode	SD	Min	Max	IQR[a]
Attitudinal (N=23)	3.53	3.4	3.4	0.496	2.8	4.4	0.6 (3.2-3.8)
Emotional (N=23)	3.628	3.6	3.8	0.482	2.8	4.4	0.8 (3.2-4)
Functional (N=23)	3.496	3.4	__[b]	0.692	2.4	4.8	1 (3-4)

[a]IQR: interquartile range.

[b]Multiple modes exist.

Cluster Analysis of Parents

We performed dimension reduction using the PCA technique [55]. We selected the component with the highest percentage of variance explained and selected 10 variables that have correlation of >.4 with the chosen component. We used the *k*-means algorithm with 2 to 4 possible clusters to identify differentiated groups of parents in our sample. We indicate the most prominent attributes that differentiate these groups.

An analysis with 2 clusters of parents identified the 2 groups, which are presented in Table 10. We performed an independent samples *t* test and Mann-Whitney nonparametric *U* tests to check the differences between the parents' characteristics in these 2 clusters.

The analysis with 3 clusters of parents identified the 3 groups presented in Table 11.

We performed a one-way ANOVA to check the differences between the parents' characteristics in the 3 clusters. The following differences were significant at the 5% significance level: parent's age (P<.001), child's age (P<.001), adherence of parents (P<.001), adherence of child (P<.001), and willingness to enroll child (P<.001), time since the last attack (P=.002), and time since diagnosis (P<.001).

An analysis with 4 clusters of parents did not reveal any unique cluster with special characteristics. Specifically, the largest group from the 3-cluster analysis was divided into 2 subgroups with slight differences.

Correlations

Table 12 presents Pearson correlations for scale variables in the parents' data. A correlation analysis was not performed on the children's data because of the low number of respondents.

Table 10. Parent clusters.

Parameter	Cluster 0	Cluster 1	*t* test (*P* value)	*U* test (*P* value)
Number of cases	41	32	__[a]	—
Parents' age	36.88	45.16	<.001	<.001
Child's age	5.66	13.23	<.001	<.001
Children who carry AAI[b] (valid %)	25 (64)[c]	32 (100)[d]	—	—
Parents who carry AAI for their children (valid %)	38 (97)[c]	14 (44)[d]	—	—
Time since diagnosis (years)	5.53	11.45	<.001	<.001
Time since the last attack (years)	3.82	6.16	.002	.01
Adherence of all parents	2.934	1.79	<.001	—
Adherence of parents who carry AAI for their children	2.97	2.46	<.001	—
Adherence of all children	1.89	2.88	<.001	—
Adherence of children who carry AAI.	2.48	2.88	<.001	—
Adherence: number of days in past week parents had access to AAI[7]	5.89	4.63	.03	.03
Willingness to enroll child in the community	2.04	4.66	<.001	—

[a]Test is irrelevant.
[b]AAI: automatic adrenaline injector.
[c]N=39.
[d]N=32.

Table 11. Parent clusters.

Parameter	Cluster 0	Cluster 1	Cluster 2
Number of cases	42	17	14
Parents' age	39.62	47.23	35.00
Child's age	8.33	14.85	3.94
Children who carry AAI[a] (valid %)	40 (100)[b]	17 (100)[c]	0 (0)[d]
Parents who carry AAI for their children (valid %)	37 (92)[b]	1 (6)[c]	14 (100)[d]
Time since diagnosis (years)	7.48	13.18	3.95
Time since the last attack (years)	4.36	6.81	3.92
Adherence of parents	2.80	1.09	2.96
Adherence of parents who carry AAI for their children	2.83	1.00	3.00
Adherence of child (children who carry AAI)	2.63 (2.66)	2.82 (2.82)	0.81 (0.00)
Adherence: number of days in past week parents had access to AAI[7]	5.79	4.22	5.36
Willingness to enroll child in the community	3.07	4.82	1.57

[a]AAI: automatic adrenaline injector.
[b]N=40.
[c]N=17.
[d]N=14.

Table 12. Correlations between variables.

Variables[a]	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
V1	1	0.143	0.291[b]	0.828[c]	0.793[c]	0.098	−0.115	0.029	0.502[c]	0.277[b]	−0.267[b]	0.086	−0.292[b]
V2	0.143	1	0.017	0.039	0.087	−0.248[b]	0.099	0.38	0.204	−0.299[b]	−0.054	−0.280[b]	0.099
V3	0.291[b]	0.017	1	0.326[c]	0.316[b]	0.022	−0.009	−0.215	0.131	0.237[b]	0.052	0.060	0.011
V4	0.828[c]	0.039	0.326[c]	1	0.91[c]	0.163	−0.048	−0.086	0.481[c]	0.455[c]	−0.187	0.202	−0.249[b]
V5	0.793[c]	0.087	0.316[b]	0.91[c]	1	−0.021	−0.091	−0.095	0.529[c]	−0.039	0.26[b]	−0.102	0.487[c]
V6	0.098	−0.248[b]	0.022	0.163	−0.021	1	0.200	−0.091	0.011	−0.012	−0.114	0.034	0.027
V7	−0.155	0.099	−0.009	−0.048	−0.091	0.200	1	0.199	−0.035	−.030	−0.075	0.008	0.008
V8	0.029	0.038	−0.215	−0.086	−0.095	−0.091	0.199	1	−0.056	−0.003	0.082	0.060	0.135
V9	0.502[c]	0.204	0.131	0.481[c]	0.529[c]	0.011	−0.035	−0.056	1	−0.205	−0.240	−0.054	−0.177
V10	0.277[b]	−0.299[b]	0.237[b]	0.455[c]	−0.039	−0.012	−0.030	−0.003	−0.205	1	0.142	0.511[c]	0.013
V11	−0.267[b]	−0.054	0.052	−0.187	0.26[b]	−0.114	−0.075	0.082	−0.240	0.142	1	0.188	0.634[c]
V12	0.086	−0.280[b]	0.060	0.202	−0.102	0.034	0.008	0.060	−0.054	0.511[c]	0.188	1	−0.056
V13	−0.292[b]	0.099	0.011	−0.249[b]	0.487[c]	0.027	0.008	0.135	−0.177	0.013	0.634[c]	−0.056	1

[a]The list of variables used are as follows: V1, parents' age; V2, parents' years of education; V3, number of children; V4, child's age; V5, time as diagnosis (years); V6, parents' opinion about the minimum age for a child to join an ERC; V7, number of days in past week parent who had access to AAI; V8, number of days in past week child who had access to AAI; V9, time as last anaphylactic attack; V10, number of anaphylactic attacks in the past; V11, protective parenting style; V12, dismissive parenting style; V13, monitoring parenting style.

[b]Significant at the 5% level.

[c]Significant at the 1% level.

Parents' Willingness to Join an Emergency Response Community

Parents were asked 2 questions about their willingness to join an ERC. A total of 69 parents (95%) answered "Yes" to the yes-or-no question "Do you intend to join the community?" Because of the very high percentage of affirmative questions, no further statistical analysis (eg, logistic regression) was possible.

Parents were also asked about the probability (0 [very unlikely] to 6 [very likely]) that they would join an ERC. Their answers are presented in Table 13.

We used an ordinal regression model to analyze the factors that influence the probability of joining an ERC. We used the following independent variables: child's gender, child's age, parents' age, parents' education, parents' adherence (number of days in past week parent had immediate access to AAI), child's adherence (number of days in past week child had immediate access to AAI), time since diagnosis, the number of anaphylactic attacks in the past, time since the last attack, and parenting style. The model fitting was significant (χ^2_{13}=22.9 with *P*=.03). The goodness of fit was significant at the 5% significance level for the Pearson chi-square test (*P*<.001). The pseudo R^2 indicators were Cox&Shell=0.449 (44.9% of the total variability explained by the model), Nagelkerke=0.507, and McFadden=0.276. A threshold check showed significant differences between most of the values of the dependent variable at the 5% significance level. The following independent variable estimates were significant at the 5% significance level: parents' age (an increase in age was associated with a decrease in the probability of joining the community, with an odds ratio [OR] of 0.688 [95% CI 0.49-0.96], Wald χ^2_1=4.7, *P*=.03), parents' adherence (number of days in past week parent had immediate access to AAI); (an increase in parents' adherence was associated with a decrease in the probability of joining the community, with an OR of 0.586 [95% CI, 0.38 to 0.91], Wald χ^2_1=5.7, *P*=.02), child's adherence (number of days in past week child had immediate access to AAI); (an increase in child's adherence was associated with an increase in the probability of joining the community, with an OR of 2.106 [95% CI 1.18-3.75], Wald χ^2_1=6.4, *P*=.01).

A 2-tailed t test for independent samples assuming equal variances (according to Levin's test) and a Mann–Whitney nonparametric U test showed no significant differences in the probability of the parent joining the community by parents' gender, but a 1-tailed t test revealed that females are more likely to join the community than males at the 5% significance level (P=.04).

A 2-tailed t test for independent samples assuming equal variances (according to Levin's test) and a Mann–Whitney nonparametric U test showed no significant differences in the probability of the parent joining the community by child's gender.

Table 13. Parents' answers about the probability of them joining the community (N=73).

Answer	n (%)
0 (very unlikely)	1 (1.4)
1	2 (2.7)
2	1 (1.4)
3	5 (6.8)
4	7 (9.6)
5	11 (15.1)
6 (very likely)	45 (61.6)
No answer	1 (1.4)

Children's Willingness to Join an Emergency Response Community

Children were asked 2 questions about their willingness to join an ERC. A total of 18 out of 23 children (78%) answered "Yes" to the yes-or-no question "If your parents let you, do you intend to join the community?"

Because of the low number of respondents, a logistic regression was not able to analyze the factors that influence children's opinions about joining the community.

Children were also asked what the probability was (0 [very unlikely] to 6 [very likely]) that they would join an ERC, assuming that their parents would allow them to join. Their answers are presented in Table 14.

We used an ordinal regression model to analyze the factors that influence the probability of joining an ERC. We used the following independent variables: child's gender, child's age, child's adherence (number of days in past week child had immediate access to AAI), parenting style, and child's autonomy. The model fitting was not significant. The goodness of fit was significant at the 5% significance level for the Pearson chi-square test (P<.001). The pseudo R^2 indicators were Cox&Shell=0.437 (43.7% of total variability explained by the model), Nagelkerke=0.453, and McFadden=0.173. A threshold check resulted in nonsignificant differences between all values of the dependent variable at the 5% significance level. The following independent variables estimates were significant at the 5% significance level: age (an increase in age was associated with an increase in the probability of joining the community, with an OR of 1.508 [95% CI 1.13-2.02], Wald χ^2_1=7.6, P=.006) and emotional autonomy (an increase in emotional autonomy was associated with a decrease in the probability of joining the community, with an OR of 0.038 [95% CI 0.0018-0.79], Wald χ^2_1=4.4, P=.04). Because of the low number of respondents, the ability of the model to explain the variability is limited. A series of t tests for independent samples assuming equal variances (according to Levin test) and Mann-Whitney nonparametric U tests showed no significant differences by child's gender in the probability of a child joining the community. A chi-square test showed no significant differences between male and female children in the probability of joining the community.

Table 14. Children's answers about the probability of them joining the community (N=23).

Answer	n (%)
0 (very unlikely)	2 (9)
1	1 (4)
2	3 (13)
3	1 (4)
4	9 (39)
5	2 (9)
6 (very likely)	5 (22)
No answer	0 (0)

Parents' Willingness to Enroll Their Children in an Emergency Response Community

Parents were asked 2 questions about their willingness to enroll their children in an ERC. A total of 36 parents out of 73 (49%) answered "yes" to the yes-or-no question "Do you intend to enroll your child in the community?"

We used a logistic regression to analyze the factors that influence the parents' decision to enroll their children in the community (dependent variable). We used the following independent variables: parents' age, parents' education, number of children, child's age, time since diagnosis, parents' adherence (number of days in past week parents had immediate access to AAI), child's adherence (number of days in past week child had

immediate access to AAI), time since the last anaphylactic attack, the number of anaphylactic attacks in the past, and parenting style. Omnibus tests of model coefficients provided significant results with P=.04, Cox&Shell R^2=0.426, and Nagelkerke R^2=0.574. According to the model, all independent variables were included in the equation, but only 1 was significant at the 5% significance level: parents' adherence (days in the past week parent had immediate access to AAI) was negatively associated with the willingness to enroll the child into the community. Bootstrapping with 1000 iterations provided the following significant variables at the 5% significance level: parents' age (negative), parents' education (negative), child's age (positive), time since diagnosis (positive), the number of anaphylactic attacks in the past (negative), parents' adherence (number of days in the past week parent had immediate access to AAI; negative), and child's adherence (number of days in the past week child had immediate access to AAI; positive).

We performed another analysis of this variable by applying the J48 classification tree to evaluate the influence of different independent variables on the parents' decision to enroll their children in the community. The tree correctly classifies 71.21% of the cases. Figure 1 presents the results.

In an attempt to expand the options scale, the parents were also asked the question, "What is the probability (0–6) that you will enroll your children in an ERC?" Their answers are presented in Table 15.

We used an ordinal regression model to analyze the factors that influence the probability that parents will enroll their child in an ERC. We used the following independent variables: parents' age, parents' education, number of children, child's age, parents' adherence (number of days in the past week parent had immediate access to AAI), child's adherence (number of days in the past week child had immediate access to AAI), time since the last anaphylactic attack, the number of anaphylactic attacks in the past, parenting style, and child's gender. The model fitting was not significant (χ^2_{13}=19.1 with P=.09). The goodness of fit was significant at the 5% significance level for the Pearson chi-square test (P=.02). The pseudo R^2 indicators were Cox&Shell=0.359 (35.9% of total variability explained by the model), Nagelkerke=0.372, and McFadden=0.133. A threshold check resulted in insignificant differences among all 7 values of the dependent variable. Bootstrapping did not improve these results.

Figure 1. The influence of different independent variables on the parents' decision to enroll their children in the community.

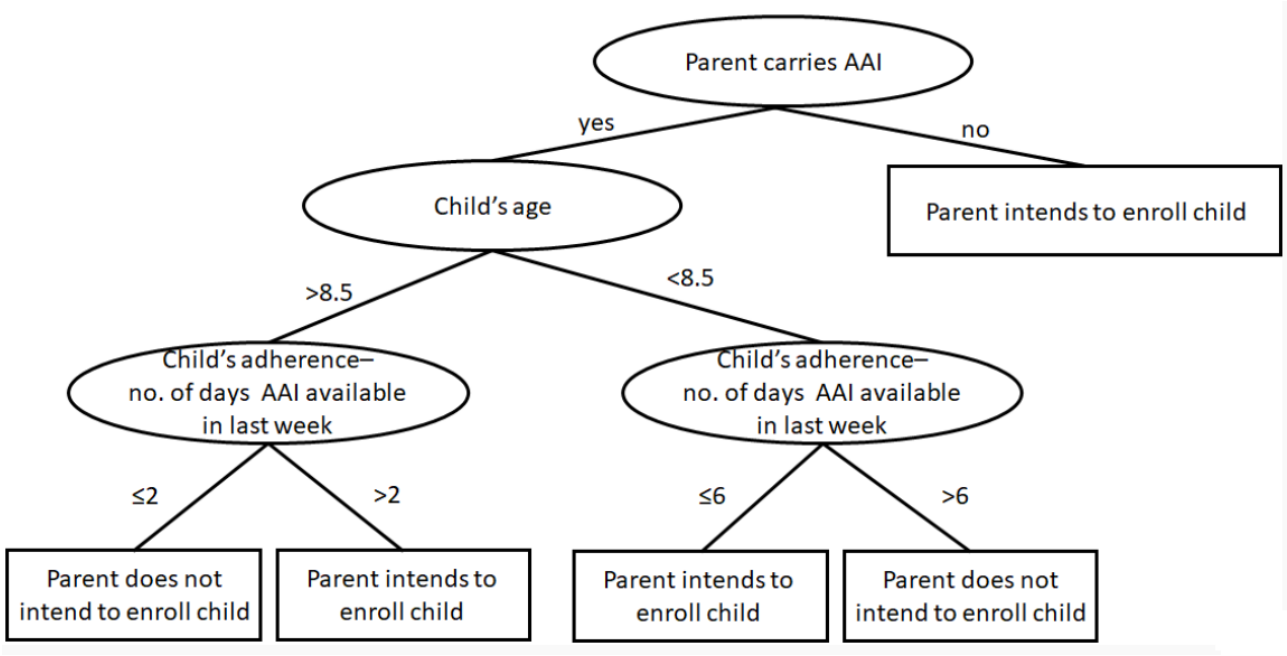


Table 15. Parents' answers about the probability (0 [very unlikely] to 6 [very likely]) of enrolling their children in the community (N=73).

Answer	n (%)
0 (very unlikely)	19 (26.0)
1	3 (4.1)
2	4 (5.5)
3	5 (6.8)
4	6 (8.2)
5	8 (11.0)
6 (very likely)	19 (26.0)
No answer	9 (12.3)

Opinions About Minimum Age to Join an Emergency Response Community

Both parents and children were asked their opinions about the minimum age to join an ERC (see Table 16).

In a paired samples *t* test, no significant differences were found between children's reports and their parents' reports. In an independent samples *t* test, no significant differences were found between males and females. A series of Pearson correlation tests revealed a weak negative (–.246) correlation between parents' opinions and parents' years of education (*P*=.049). No significant correlations of parents' opinions were found either with parents' age or with their children's age.

We used a linear regression to analyze the factors that influence parents' opinions about the minimum age to join an ERC. We used the following independent variables: parents' age, parents' education, number of children, child's age, parents' adherence (number of days in the past week parent had immediate access to AAI), child's adherence (number of days in the past week child had immediate access to AAI), time since the last anaphylactic attack, parenting style, and number of anaphylactic attacks in the past. A multicollinearity analysis did not reveal any evidence of multicollinearity. The model resulted in R^2=0.55, and the model was significant at the 5% significance level (*P*=.004). An analysis of standardized regression coefficients revealed that the most influential significant variables are child's age (beta=0.489, standardized beta=0.801, *P*=.003), protective parenting style (beta=–2.767, standardized beta=–0.561, *P*=.01), parent's age (beta=–0.242, standardized beta=–0.526, *P*=.03), monitoring parenting style (beta=2.584, standardized beta=0.427, *P*=.03), and parents' education (beta=–0.528, standardized beta=–0.418, *P*=.007). All other variables were much less influential and were not significant at the 5% significance level.

Table 16. Opinions about minimum age to join an emergency response community (all values in years).

Population	N (valid[a])	Average	Min	Max	SD	Median	95% CI
Parents	65	12.27	6	18	3.02	12	11.52-13.02
Children (attended)	23	13.15	6.5	20	3.44	12	11.63-14.67

[a]These data were missing in 8 parents' questionnaires.

We used a linear regression to analyze the factors that influence children's opinions about the minimum age to join an ERC. We used the following independent variables: child's age, child's adherence (number of days in the past week child had immediate access to AAI), parenting style, and child's autonomy. A multicollinearity analysis did not reveal any evidence of multicollinearity. The model resulted in R^2=0.602, and the model was very close to significance at the 5% significance level (*P*=.05). An analysis of standardized regression coefficients revealed that the most influential significant variables are monitoring parenting style (beta=4.668, standardized beta=0.952, *P*=.005) and protective parenting style (beta=–4.381, standardized beta=–0.888, *P*=.008). All other variables were much less influential and were not significant at the 5% significance level.

Discussion

Principal Findings

Parents' Willingness to Join an Emergency Response Community

Parents' willingness to join the community was very high, even for a convenience sample. In the following, we describe the main factors that influence parents' willingness to join.

Parents' willingness to join was negatively correlated with parents' age and parents' adherence. Parents' age had a strong positive correlation with child's age. Parents of younger children carried the AAIs for their children. These findings can be explained by the *free-rider effect* [56]: parents who carried an AAI for a young child, but had low adherence, wanted to join the ERC to get an additional layer of response in an emergency. Parents' age was also negatively correlated with a protective parenting style, as this parenting style is associated with a need for additional safety measures [47] and joining an ERC can satisfy this need. Parents' age was strongly positively correlated with child's age, which was positively correlated with time since diagnosis and time since the last anaphylactic attack. Parents of newly diagnosed children had higher levels of parental anxiety [57,58], which led to higher willingness to join to get an additional layer of support.

Parents' willingness to join an ERC was positively correlated with child's adherence. Previous studies have found that parents' psychological characteristics influence their child's adherence [59]. A possible explanation for this finding is that parents' characteristics, such as self-efficacy or parental warmth, which are known to be associated with higher adherence among children [60,61], are also associated with a higher probability of joining an ERC. Further research is required to verify this hypothesis.

We found that females were more likely to join the community than males. This can be explained both from a giving point of view (females participate more in volunteer activities than males [62] and have higher motivation to volunteer [63]) and from a taking point of view (mothers of children with a medical condition experience higher levels of parental anxiety than fathers [58,64] and are thus more likely than males to seek help for their children's condition [65]).

Children's Willingness to Join an Emergency Response Community

Children's willingness to join the community was lower than that of their parents, but still high. Children's willingness to join the community was positively correlated with age. This is a straightforward finding both from a giving point of view (age is positively correlated with volunteerism [66]) and from a

taking point of view (age is positively correlated with help seeking [67]). Children's willingness to join the community was also negatively correlated with children's emotional autonomy. The latter correlation may be more of a methodological issue, as it may be related to the specific questions in the emotional autonomy subscale of the AAQ. For example, "I adapt myself to what other people want" (from question 7 of the AAQ) might be implicitly associated with lack of emotional autonomy in the context of joining an ERC.

Parents' Willingness to Enroll Their Children in an Emergency Response Community

About half of the parents expressed willingness that their children join an ERC. Both the logistic regression and the classification tree identified that being the parent responsible to carry an AAI for his child and parent's adherence as significant factors negatively correlated with parent's willingness to enroll their child in an ERC. It seems that parents who are not able to provide their children with an AAI in the event of an anaphylactic attack want to enroll their children to provide them with an additional layer of support in an emergency.

Child's age was positively correlated with parents' willingness to enroll their children in the community. From a taking point of view, this can be explained by the transition from a protective parenting style, which is more common in parents of younger children, to a monitoring parenting style, which prevails as children grow older [68]. While the protective parenting style is characterized by the parents' attempts to protect their child by themselves, the monitoring parenting style is characterized by the parents' provision of tools to enable the child to cope by himself. Thus, parents of older children, especially those who begin to go out on their own, are more likely to view enrolling their child in an ERC as another tool that he can use in case of an emergency. From a giving point of view, age is positively correlated with volunteerism [66] and is also positively correlated with the ability to help another patient in an emergency, such as to provide him with cardiopulmonary resuscitation [69]. The negative correlation between a history of anaphylactic attacks in the past and parents' willingness to enroll their children can be explained by the parents' concerns that children who use their AAI to help another patient will remain unprotected until they get a replacement [42].

The 2 lower levels of the classification tree reveal another interesting effect: whereas parents of a young child want to enroll him if he has less than perfect adherence, parents of an older child want to enroll him only if he carries an AAI at least 3 days a week. The former finding can be explained by the *free-rider effect* and *taking* behavior among parents of younger children, who understand that their children probably would not be able to provide help to others. The latter finding appears to indicate *giving* behavior among parents of older children, who want to make sure that if their children join, they have a reasonable chance of helping others when called to action.

Cluster analysis with 2 clusters revealed 2 subgroups of 32 and 41 parents. The smaller group is characterized by younger parents, younger children, shorter time since diagnosis, shorter time since last attack, lower percentage of children who carry an AAI, higher percentage of parents who carry an AAI for their children, higher adherence among parents, lower adherence among children, and lower willingness to enroll child in an ERC. Cluster analysis with 3 clusters revealed 3 subgroups:

- Cluster 2: 14 parents to very young children. In this group, all parents carry an AAI for their children, and none of children carries the AAI for himself. Parents' adherence is high when compared with other clusters, and the willingness to enroll child in the community is very low.
- Cluster 1: 17 parents to adolescents. In this group, only 1 parent carries an AAI for his child, and all children carry an AAI for themselves. Adherence among children is the highest when compared with other clusters, and the willingness to enroll child in the community is very high.
- Cluster 0: 42 parents form the third cluster, which is in the middle between the 2 aforementioned clusters. Almost all children already carry an AAI for themselves, but many parents continue to carry an AAI. The adherence is a little bit lower than in cluster 1, but still high. The willingness to enroll a child in the community falls near the midpoint between clusters 1 and 2.

These results are consistent with previously described results obtained by other techniques. Parents' willingness to enroll their child in the community is positively correlated with child's age and child's adherence.

Minimum Age to Join an Emergency Response Community

The minimum age for a child to join an ERC is between the ages of 12 and 13 years in the opinion of both parents and children. This can be explained by the fact that the study was performed in Israel where the ages of 12 years for girls and 13 years for boys are commonly considered the years when a child comes of age. Another possible explanation is that the age of 12 to 13 years is known to be the cutoff for cognitive development [70,71].

The parents' opinion about the minimum age for a child to join an ERC was positively correlated with the age of the parents' own child. This finding may be related to the well-known cognitive bias of *anchoring* [72], suggesting that parents rely on their own child's age as an *anchor* when making decisions about the minimum age to join an ERC.

The parents' opinion about the minimum age for a child to join an ERC was negatively correlated with the protective parenting style and positively correlated with the monitoring parenting style. These findings can be explained by the desire of parents with a higher protective parenting style to provide their children with an additional layer of support as soon as possible, and the desire of parents with a higher monitoring parenting style to provide their children with tools to cope by themselves when they become more independent.

The parents' opinion about the minimum age for a child to join an ERC was negatively correlated with the parents' education, that is, the more the parent is educated, the higher the likelihood she will allow her child to join an ERC at a younger age. Such association may be related to the relationship between the level of education and the exposure to updated technological solutions for treating medical conditions [73].

The children's opinion about the minimum age for a child to join an ERC was negatively correlated with the protective parenting style and positively correlated with the monitoring parenting style. These findings are consistent with the findings about the parents' opinion presented above. Children who have protective parents are prone to anxiety [74] and thus need the additional protection of joining an ERC as soon as possible, whereas children who have monitoring parents view an ERC as a tool that will help them to cope by themselves, more so as they grow older [75].

Limitations

Our research used a convenience sample of highly motivated parents of children with food allergies, namely, those parents who decided to attend the annual meeting of a nonprofit organization for patients with food allergies (these attendees represent about 5% of the total membership of the organization).

Our study was limited to a single emergency condition: anaphylaxis. Most of the participants in our study were parents of patients, but not patients themselves. Our study was conducted in a single country: Israel. Cultural differences could lead to different results in different countries [76].

Comparison With Prior Work

We observed high adherence levels among parents: 45 out of 50 parents (90%) reported that they always carry an AAI, 37 out of 52 parents (71%) reported that they had immediate access to an AAI every day in the past week, and 22 out of 52 parents (42%) had an AAI when filling out the questionnaire. Adherence was also high among children: 43 out of 57 parents (75%) reported that their children always carried an AAI, 51 out of 57 parents (89%) reported that their children had immediate access to AAI every day in the past week, and 53 out of 57 parents (93%) reported that the child had an AAI with him at the conference. These adherence levels are higher than those reported in 2 previous studies that found that 30% [77] and 26% to 45% [78] of patients carried an AAI at all times. However, these differences are not surprising among patients who decided to attend a conference and can thus be considered more motivated than those who stayed at home.

Comparison with the recent study by Shaker et al [42] is very instructive, given the fact that it also targeted parents of allergic children. The research question of Shaker et al was about sharing an AAI, and our research question was about joining an ERC, making our works complementary. Our findings that there are no significant differences in the willingness to join an ERC either by parents' gender or by child's gender or by parents' education are consistent with the results of Shaker et al. They also found that a history of anaphylactic attacks was associated with a slightly higher willingness to share an AAI (79% vs 76%), but this finding was not statistically significant. In our study, the number of anaphylactic attacks in the past did not influence the willingness of parents to join an ERC, but it did negatively affect the willingness of parents to enroll their children in an ERC.

Shaker et al [52] reported that parents expressed concern about "leaving their own child without an AAI" and about the replacement cost of an AAI. It is therefore noteworthy that according to the EPIMADA protocol, an ambulance arriving on scene is required to provide a replacement AAI to the community member who responded to the emergency by giving his own AAI to the patient in distress. In other words, the EPIMADA protocol should alleviate the concerns identified by Shaker et al [42].

EPIMADA is a regulated community, and in every event, the volunteers are provided with guidance by phone from experienced EMS dispatchers. This approach can thus alleviate concerns about "harming the patient if it is not a real allergic reaction" that were also identified by Shaker et al [42].

Conclusions

Our results provide strong support for the creation of ERCs for allergy patients. The willingness to join the community was found to be very high.

We studied a wide range of variables that can affect the willingness to join an ERC and described those variables that we found to be significant in our field study. Our findings can be of interest both to researchers of smartphone-based emergency response communities and to EMS administrators and policy makers who are considering establishing an ERC.

We note that mediation of an ERC by EMSs has the potential to solve several problems:

- An ambulance that arrives to the scene can replace the AAI of the community member who responded to the event by administering his own AAI to the patient in distress. This approach can alleviate the concerns about "leaving oneself or one's child without an AAI" and those about replacement costs and delays.
- EMS dispatchers can provide guidance by phone. This approach can alleviate the concerns about "harming the patient" and can improve responders' chances of providing emergency assistance.

Future research needs to address the following issues:

- Rare use of the app may cause users to forget to install it on new phones. Retention strategies should be developed.
- It is unclear whether ERC membership incentivizes adherence. On the one hand, a community member may feel responsibility to be ready to respond at any time, which could raise his level of adherence. On the other hand, a free-rider effect may lower his level of adherence.
- In this research, we focus on the willingness to join an ERC. Willingness to respond to an event is a different decision and needs to be studied.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire for parents.
[PDF File (Adobe PDF File), 375KB - mhealth_v7i8e13892_app1.pdf]

Multimedia Appendix 2
Questionnaire for children.
[PDF File (Adobe PDF File), 411KB - mhealth_v7i8e13892_app2.pdf]

References

1. Ramanayake RP, Ranasingha S, Lakmini S. Management of emergencies in general practice: role of general practitioners. J Family Med Prim Care 2014;3(4):305-308 [FREE Full text] [doi: 10.4103/2249-4863.148089] [Medline: 25657933]
2. Hansen CM, Kragholm K, Pearson DA, Tyson C, Monk L, Myers B, et al. Association of bystander and first-responder intervention with survival after out-of-hospital cardiac arrest in North Carolina, 2010-2013. J Am Med Assoc 2015 Jul 21;314(3):255-264 [FREE Full text] [doi: 10.1001/jama.2015.7938]
3. Hasselqvist-Ax I, Riva G, Herlitz J, Rosenqvist MA, Hollenberg J, Nordberg P, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med 2015 Jun 11;372(24):2307-2315 [FREE Full text] [doi: 10.1056/NEJMoa1405796] [Medline: 26061835]
4. Hanfling D, Altevogt B, Viswanathan K, Gostin L. Crisis Standards of Care: A Systems Framework for Catastrophic Disaster Response. Washington D.C: The National Academies Press; 2012.
5. Razzak J, Kellermann A. Emergency medical care in developing countries: is it worthwhile? Bull World Health Organ 2002;80(11):900-905 [FREE Full text] [Medline: 12481213]
6. Chanta S, Mayorga M, McLay L. Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. Ann Oper Res 2011 Sep 27;221(1):133-159. [doi: 10.1007/s10479-011-0972-6]
7. Roudsari BS, Nathens AB, Arreola-Risa C, Cameron P, Civil I, Grigoriou G, et al. Emergency Medical Service (EMS) systems in developed and developing countries. Injury 2007 Sep;38(9):1001-1013. [doi: 10.1016/j.injury.2007.04.008] [Medline: 17583709]
8. Waseem H, Naseer R, Razzak JA. Establishing a successful pre-hospital emergency service in a developing country: experience from Rescue 1122 service in Pakistan. Emerg Med J 2011 Jun;28(6):513-515. [doi: 10.1136/emj.2010.096271] [Medline: 20844094]
9. Kitamura T, Kiyohara K, Sakai T, Matsuyama T, Hatakeyama T, Shimamoto T, et al. Public-access defibrillation and out-of-hospital cardiac arrest in Japan. N Engl J Med 2016 Dec 27;375(17):1649-1659. [doi: 10.1056/NEJMsa1600011] [Medline: 27783922]
10. Murakami Y, Iwami T, Kitamura T, Nishiyama C, Nishiuchi T, Hayashi Y, Utstein Osaka Project. Outcomes of out-of-hospital cardiac arrest by public location in the public-access defibrillation era. J Am Heart Assoc 2014 Apr 22;3(2):e000533 [FREE Full text] [doi: 10.1161/JAHA.113.000533] [Medline: 24755149]
11. Nakahara S, Tomio J, Ichikawa M, Nakamura F, Nishida M, Takahashi H, et al. Association of bystander interventions with neurologically intact survival among patients with bystander-witnessed out-of-hospital cardiac arrest in Japan. J Am Med Assoc 2015 Jul 21;314(3):247-254 [FREE Full text] [doi: 10.1001/jama.2015.8068] [Medline: 26197185]
12. Claesson A, Bäckman A, Ringh M, Svensson L, Nordberg P, Djärv T, et al. Time to delivery of an automated external defibrillator using a drone for simulated out-of-hospital cardiac arrests vs emergency medical services. J Am Med Assoc 2017 Dec 13;317(22):2332-2334 [FREE Full text] [doi: 10.1001/jama.2017.3957] [Medline: 28609525]
13. Elsner J, Meisen P, Thelen S, Schilberg D, Jeschke S. EMuRgency–a basic concept for an AI driven volunteer notification system for integrating laypersons into emergency medical services. Int J Adv Life Sci 2013;5(3):223-236 [FREE Full text]
14. Folke F, Lippert F, Nielsen S, Gislason G, Hansen M, Schramm T, et al. Location of cardiac arrest in a city center: strategic placement of automated external defibrillators in public locations. Circulation 2009 Aug 11;120(6):510-517. [doi: 10.1161/CIRCULATIONAHA.108.843755] [Medline: 19635969]
15. Kiyohara K, Kitamura T, Sakai T, Nishiyama C, Nishiuchi T, Hayashi Y, et al. Public-access AED pad application and outcomes for out-of-hospital cardiac arrests in Osaka, Japan. Resuscitation 2016 Dec;106:70-75. [doi: 10.1016/j.resuscitation.2016.06.025] [Medline: 27373223]
16. Marshall A, Cairns K, Kee F, Moore M, Hamilton A, Adgey A. A Monte Carlo Simulation Model to Assess Volunteer Response Times in a Public Access Defibrillation Scheme in Northern Ireland. 2006 Jun Presented at: 19th IEEE International Symposium on Computer-Based Medical Systems, CBMS'06; June 22-23, 2006; Salt Lake City, UT, USA p. 783. [doi: 10.1109/CBMS.2006.19]
17. Rea T, Blackwood J, Damon S, Phelps R, Eisenberg M. A link between emergency dispatch and public access AEDs: potential implications for early defibrillation. Resuscitation 2011 Aug;82(8):995-998. [doi: 10.1016/j.resuscitation.2011.04.011] [Medline: 21570169]
18. Takei Y, Kamikura T, Nishi T, Maeda T, Sakagami S, Kubo M, et al. Recruitments of trained citizen volunteering for conventional cardiopulmonary resuscitation are necessary to improve the outcome after out-of-hospital cardiac arrests in

remote time-distance area: A nationwide population-based study. Resuscitation 2016 Dec;105:100-108. [doi: 10.1016/j.resuscitation.2016.05.021] [Medline: 27263486]
19. Schwartz DG, Bellou A, Garcia-Castrillo L, Muraro A, Papadopoulos NG. Towards chronic emergency response communities for anaphylaxis. In: Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014).: IEEE; 2014 Presented at: The 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014); 13-15 Aug, 2014; Redwood City, CA, USA. [doi: 10.1109/IRI.2014.7051878]
20. Schwartz D, Bellou A, Garcia-Castrillo L, Muraro A, Papadopoulos N. Exploring mHealth participation for emergency response communities. AJIS 2017 Mar 08;21. [doi: 10.3127/ajis.v21i0.1378]
21. Park H. Pervasive healthcare computing: EMR/EHR, wireless and health monitoring. Healthc Inform Res 2011;17(1):89 [FREE Full text] [doi: 10.4258/hir.2011.17.1.89]
22. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 01;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
23. Huang H, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
24. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
25. Lancy D, Nattiv A. Parents as volunteers: storybook readers/listeners. Child Educ 1992 Jun;68(4):208-212. [doi: 10.1080/00094056.1992.10520876]
26. Atkins RG, Hawdon JE. Religiosity and participation in mutual-aid support groups for addiction. J Subst Abuse Treat 2007 Oct;33(3):321-331 [FREE Full text] [doi: 10.1016/j.jsat.2007.07.001] [Medline: 17889302]
27. Bennett S, Banyard V, Garnhart L. To act or not to act, that is the question? Barriers and facilitators of bystander intervention. J Interpers Violence 2014 Feb;29(3):476-496. [doi: 10.1177/0886260513505210] [Medline: 24097909]
28. Latané B, Darley J. Group inhibition of bystander intervention in emergencies. J Pers Soc Psychol 1968 Nov;10(3):215-221. [Medline: 5704479]
29. Latane B, Darley J. The Unresponsive Bystander: Why Doesn't He Help?. New Jersey, United States: Prentice-Hall Incorporated; 1970.
30. Magura S. Effectiveness of dual focus mutual aid for co-occurring substance use and mental health disorders: a review and synthesis of the. Subst Use Misuse 2008 Oct;43(12-13):1904-1926 [FREE Full text] [doi: 10.1080/10826080802297005] [Medline: 19016171]
31. Hyde B. Mutual aid group work: social work leading the way to recovery-focused mental health practice. Soc Work Groups 2013 Jan;36(1):43-58. [doi: 10.1080/01609513.2012.699872]
32. Simmons D. Diabetes self help facilitated by local diabetes research: the Coventry Asian Diabetes Support Group. Diabet Med 1992 Nov;9(9):866-869. [doi: 10.1111/j.1464-5491.1992.tb01907.x]
33. Gilden J, Hendryx M, Clar S, Casia C, Singh S. Diabetes support groups improve health care of older diabetic patients. J Am Soc Geriatr Dent 1992 Feb;40(2):147-150. [doi: 10.1111/j.1532-5415.1992.tb01935.x]
34. Shostak S, Fox NS. Forgetting and remembering epilepsy: collective memory and the experience of illness. Sociol Health Illn 2012 Mar;34(3):362-378 [FREE Full text] [doi: 10.1111/j.1467-9566.2011.01373.x] [Medline: 21707662]
35. Piliavin I, Rodin J, Piliavin J. Good Samaritanism: an underground phenomenon? J Pers Soc Psychol 1969;13(4):289-299. [doi: 10.1037/h0028433]
36. Levine M, Cassidy C, Brazier G, Reicher S. Self-categorization and bystander non-intervention: two experimental studies1. J Appl Social Pyschol 2002 Jul;32(7):1452-1463. [doi: 10.1111/j.1559-1816.2002.tb01446.x]
37. Beyene K, Sheridan J, Aspden T. Prescription medication sharing: a systematic review of the literature. Am J Public Health 2014 Apr;104(4):e15-e26. [doi: 10.2105/AJPH.2013.301823] [Medline: 24524496]
38. Hogan D, Moreland A, Lane P, Lal S. Exchange of prescription medications by dermatology outpatients. J Am Acad Dermatol 1990 Nov;23(5 Pt 1):953. [doi: 10.1016/S0190-9622(08)80715-3)] [Medline: 2254491]
39. Garnier LM, Arria AM, Caldeira KM, Vincent KB, O'Grady KE, Wish ED. Sharing and selling of prescription medications in a college student sample. J Clin Psychiatry 2010 Mar;71(3):262-269 [FREE Full text] [doi: 10.4088/JCP.09m05189ecr] [Medline: 20331930]
40. Goldsworthy R. Recreational versus nonrecreational prescription borrowing: time for an expanded conceptualization? J Adolesc Health 2010 Apr;46(4):402. [doi: 10.1016/j.jadohealth.2009.12.016] [Medline: 20307833]
41. Valet R, Gebretsadik T, Minton P, Woodward K, Wu A, Hartert T, et al. Prevalence and characteristics of medication sharing behavior in a pediatric Medicaid population with asthma. Ann Allergy Asthma Immunol 2015 Feb;114(2):151-153 [FREE Full text] [doi: 10.1016/j.anai.2014.11.007] [Medline: 25492098]
42. Shaker M, Kanaoka T, Murray R, Toy D, Shaker S, Drew A. A survey of caregiver perspectives on emergency epinephrine autoinjector sharing. J Allergy Clin Immunol Pract 2018;6(5):1792-1795 [FREE Full text] [doi: 10.1016/j.jaip.2018.02.017] [Medline: 29499333]
43. National Institutes of Health. 2006. PA-06-097: Parenting Capacities and Health Outcomes in Youths and Adolescents (R01) URL: https://grants.nih.gov/grants/guide/pa-files/pa-06-097.html [accessed 2018-08-04]

44. DuPen M, van Tilburg MA, Langer S, Murphy T, Romano J, Levy R. Parental protectiveness mediates the association between parent-perceived child self-efficacy and health outcomes in pediatric functional abdominal pain disorder. Children (Basel) 2016 Sep 19;3(3):15 [FREE Full text] [doi: 10.3390/children3030015] [Medline: 27657151]
45. Valentine A, Knibb R. Exploring quality of life in families of children living with and without a severe food allergy. Appetite 2011 Oct;57(2):467-474. [doi: 10.1016/j.appet.2011.06.007] [Medline: 21708202]
46. Walker L, Levy R, Whitehead W. Validation of a measure of protective parent responses to children's pain. Clin J Pain 2006 Oct;22(8):712-716 [FREE Full text] [doi: 10.1097/01.ajp.0000210916.18536.44] [Medline: 16988567]
47. Van Slyke DA, Walker L. Mothers' responses to children's pain. Clin J Pain 2006 May;22(4):387-391. [doi: 10.1097/01.ajp.0000205257.80044.01] [Medline: 16691093]
48. Noom M, Deković M, Meeus W. Conceptual analysis and measurement of adolescent autonomy. J Youth Adolesc 2001 Oct;30(5):577-595. [doi: 10.1023/A:1010400721676]
49. Allen J, Hauser S, Bell K, O'Connor TG. Longitudinal assessment of autonomy and relatedness in adolescent-family interactions as predictors of adolescent ego development and self-esteem. Child Dev 1994 Feb;65(1):179-194. [doi: 10.1111/j.1467-8624.1994.tb00743.x] [Medline: 8131646]
50. Herbert L, Dahlquist L. Perceived history of anaphylaxis and parental overprotection, autonomy, anxiety, and depression in food allergic young adults. J Clin Psychol Med Settings 2008 Dec;15(4):261-269. [doi: 10.1007/s10880-008-9130-y] [Medline: 19104982]
51. Gallagher M, Worth A, Cunningham-Burley S, Sheikh A. Strategies for living with the risk of anaphylaxis in adolescence: qualitative study of young people and their parents. Prim Care Respir J 2012 Dec;21(4):392-397 [FREE Full text] [doi: 10.4104/pcrj.2012.00072] [Medline: 22875142]
52. Kemp SF, Lockey RF, Simons FE, World Allergy Organization ad hoc Committee on Epinephrine in Anaphylaxis. Epinephrine: the drug of choice for anaphylaxis. A statement of the World Allergy Organization. Allergy 2008 Aug;63(8):1061-1070. [doi: 10.1111/j.1398-9995.2008.01733.x] [Medline: 18691308]
53. van der Velde JL, Flokstra-de Blok BM, Vlieg-Boerstra B, Oude Elberink JN, DunnGalvin A, Hourihane J, et al. Development, validity and reliability of the food allergy independent measure (FAIM). Allergy 2010 May;65(5):630-635. [doi: 10.1111/j.1398-9995.2009.02216.x] [Medline: 19845570]
54. Jain A. Data clustering: 50 years beyond K-means. Pattern Recognit Lett 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]
55. Ding C, He X. K-means clustering via principal component analysis. In: Proceedings of the twenty-first international conference on Machine learning ACM. 2004 Presented at: The twenty-first international conference on Machine learning ACM; July 4-8, 2004; Banff, Alberta, Canada p. 29. [doi: 10.1145/1015330.1015408]
56. Kerr NL, Bruun SE. Dispensability of member effort and group motivation losses: free-rider effects. J Pers Soc Psychol 1983;44(1):78-94. [doi: 10.1037/0022-3514.44.1.78]
57. Northam E, Anderson P, Adler R, Werther G, Warne G. Psychosocial and family functioning in children with insulin-dependent diabetes at diagnosis and one year later. J Pediatr Psychol 1996 Oct;21(5):699-717. [doi: 10.1093/jpepsy/21.5.699] [Medline: 8936898]
58. Streisand R, Mackey ER, Elliot BM, Mednick L, Slaughter IM, Turek J, et al. Parental anxiety and depression associated with caring for a child newly diagnosed with type 1 diabetes: opportunities for education and counseling. Patient Educ Couns 2008 Nov;73(2):333-338. [doi: 10.1016/j.pec.2008.06.014] [Medline: 18692342]
59. Nock M, Ferriter C. Parent management of attendance and adherence in child and adolescent therapy: a conceptual and empirical review. Clin Child Fam Psychol Rev 2005 Jun;8(2):149-166. [doi: 10.1007/s10567-005-4753-0] [Medline: 15984084]
60. Davis C, Delamater A, Shaw K, La Greca AM, Eidson M, Perez-Rodriguez J, et al. Parenting styles, regimen adherence, and glycemic control in 4- to 10-year-old children with diabetes. J Pediatr Psychol 2001 Mar;26(2):123-129. [doi: 10.1093/jpepsy/26.2.123] [Medline: 11181888]
61. Ott J, Greening L, Palardy N, Holderby A, DeBell W. Self-efficacy as a mediator variable for adolescents' adherence to treatment for insulin-dependent diabetes mellitus. Children's Health Care 2000 Mar;29(1):47-63. [doi: 10.1207/S15326888CHC2901_4]
62. Einolf CJ. Gender differences in the correlates of volunteering and charitable giving. Nonprofit Volunt Sect Q 2010 Sep 28;40(6):1092-1112. [doi: 10.1177/0899764010385949]
63. Fletcher TD, Major DA. Medical students' motivations to volunteer: an examination of the nature of gender differences. Sex Roles 2004 Jul;51(1-2):109-114. [doi: 10.1023/B:SERS.0000032319.78926.54]
64. Yeh C. Gender differences of parental distress in children with cancer. J Adv Nurs 2002 Jun;38(6):598-606. [doi: 10.1046/j.1365-2648.2000.02227.x] [Medline: 12067399]
65. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003 Jan;58(1):5-14. [Medline: 12674814]
66. van Goethem AA, van Hoof A, van Aken MA, Raaijmakers QA, Boom J, de Castro BO. The role of adolescents' morality and identity in volunteering. Age and gender differences in a process model. J Adolesc 2012 Jun;35(3):509-520. [doi: 10.1016/j.adolescence.2011.08.012] [Medline: 21944561]

67. Schonert-Reichl K, Muller J. Correlates of help-seeking in adolescence. J Youth Adolescence 1996 Dec;25(6):705-731. [doi: 10.1007/BF01537450]
68. Thomasgard M, Metz WP. Parental overprotection and its relation to perceived child vulnerability. Am J Orthopsychiatry 1997 Apr;67(2):330-335. [doi: 10.1037/h0080237] [Medline: 9142367]
69. Plant N, Taylor K. How best to teach CPR to schoolchildren: a systematic review. Resuscitation 2013 Apr;84(4):415-421. [doi: 10.1016/j.resuscitation.2012.12.008] [Medline: 23246989]
70. Vakil E, Blachstein H, Sheinman M, Greenstein Y. Developmental changes in attention tests norms: implications for the structure of attention. Child Neuropsychol 2009 Jan;15(1):21-39. [doi: 10.1080/09297040801947069] [Medline: 18608225]
71. Vakil E, Blachstein H, Sheinman M. Rey AVLT: developmental norms for children and the sensitivity of different memory measures to age. Child Neuropsychol 2010 Aug 09;4(3):161-177. [doi: 10.1076/chin.4.3.161.3173]
72. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974 Sep 27;185(4157):1124-1231. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]
73. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
74. Hudson J, Rapee R. Parent-child interactions and anxiety disorders: an observational study. Behav Res Ther 2001 Dec;39(12):1411-1427. [doi: 10.1016/S0005-7967(00)00107-8] [Medline: 11758699]
75. Lee SM, Daniels MH, Kissinger DB. Parental influences on adolescent adjustment: parenting styles versus parenting practices. The Family Journal 2006 Jun 05;14(3):253-259. [doi: 10.1177/1066480706287654]
76. Anheier HK, Salamon LM. Volunteering in cross-national perspective: Initial comparisons. Law and Contemp Probs 1999;62(4):43 [FREE Full text]
77. Song T, Worm M, Lieberman P. Anaphylaxis treatment: current barriers to adrenaline auto-injector use. Allergy 2014 Aug;69(8):983-991. [doi: 10.1111/all.12387] [Medline: 24835773]
78. Goldberg A, Confino-Cohen R. Insect sting-inflicted systemic reactions: attitudes of patients with insect venom allergy regarding after-sting behavior and proper administration of epinephrine. J Allergy Clin Immunol 2000 Dec;106(6):1184-1189. [doi: 10.1067/mai.2000.110927] [Medline: 11112904]

Abbreviations

AAI: automatic adrenaline injector
AAQ: Adolescent Autonomy Questionnaire
ANOVA: analysis of variance
ARCS: Adult Responses to Children's Symptoms
EMS: emergency medical service
ERC: emergency response community
ICC: intraclass correlation
MDA: Magen David Adom—the Israeli national EMS
OR: odds ratio
PCA: principal component analysis

Edited by G Eysenbach; submitted 03.03.19; peer-reviewed by J del Hoyo Francisco, D Carvalho, H Inaba; comments to author 11.04.19; revised version received 12.05.19; accepted 21.05.19; published 27.08.19.

Please cite as:
Khalemsky M, Schwartz DG, Silberg T, Khalemsky A, Jaffe E, Herbst R
Childrens' and Parents' Willingness to Join a Smartphone-Based Emergency Response Community for Anaphylaxis: Survey
JMIR Mhealth Uhealth 2019;7(8):e13892
URL: https://mhealth.jmir.org/2019/8/e13892/
doi:10.2196/13892
PMID:31456582

Viewpoint

Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations

Igbe Tobore[1,2*], BTech, MTech; Jingzhen Li[1*], BSc (Hons), MEng; Liu Yuhang[1], BSc (Hons), MEng; Yousef Al-Handarish[1], BSc, MSc; Abhishek Kandwal[1], BSc, MSc, PhD; Zedong Nie[1], BSc, MSc, PhD; Lei Wang[1], BSc, MSc, PhD

[1]Center for Medical Robotics and Minimally Invasive Surgical Devices, Shenzhen Institutes of Advance Technology, Chinese Academy of Sciences, Shenzhen, China

[2]Graduate University, Chinese Academy of Sciences, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zedong Nie, BSc, MSc, PhD
Center for Medical Robotics and Minimally Invasive Surgical Devices
Shenzhen Institutes of Advance Technology
Chinese Academy of Sciences
1068 Xueyuan Avenue, Shenzhen University, Xili Town, Nanshan District
Shenzhen,
China
Phone: 86 755 86585213
Email: zd.nie@siat.ac.cn

Abstract

The use of deep learning (DL) for the analysis and diagnosis of biomedical and health care problems has received unprecedented attention in the last decade. The technique has recorded a number of achievements for unearthing meaningful features and accomplishing tasks that were hitherto difficult to solve by other methods and human experts. Currently, biological and medical devices, treatment, and applications are capable of generating large volumes of data in the form of images, sounds, text, graphs, and signals creating the concept of big data. The innovation of DL is a developing trend in the wake of big data for data representation and analysis. DL is a type of machine learning algorithm that has deeper (or more) hidden layers of similar function cascaded into the network and has the capability to make meaning from medical big data. Current transformation drivers to achieve personalized health care delivery will be possible with the use of mobile health (mHealth). DL can provide the analysis for the deluge of data generated from mHealth apps. This paper reviews the fundamentals of DL methods and presents a general view of the trends in DL by capturing literature from PubMed and the Institute of Electrical and Electronics Engineers database publications that implement different variants of DL. We highlight the implementation of DL in health care, which we categorize into biological system, electronic health record, medical image, and physiological signals. In addition, we discuss some inherent challenges of DL affecting biomedical and health domain, as well as prospective research directions that focus on improving health management by promoting the application of physiological signals and modern internet technology.

(JMIR Mhealth Uhealth 2019;7(8):e11966) doi:10.2196/11966

KEYWORDS

machine learning; deep learning; big data; mHealth; medical imaging; electronic health record; biologicals; biomedical; ECG; EEG; artificial intelligence

Introduction

The continuous advancement in medicine, genome, pharmaceutical, and health care monitoring is a result of the development and application of technological devices. This has made it possible to easily capture data for analysis and processing. Similarly, improvement in technology also makes it possible to store very large amount of data with useful information. Currently, camera to detect the movements of monitored patients (Panasonic BL-C230A), wireless necklace and badges for acquisition of bioacoustic signals and blood flow, wearable fiber-type smart material, cuffless blood pressure meter, and sensor devices are capable of generating large volumes of data in the form of images, sounds, text, graphs,

and signals creating the concept of big data [1-4]. The term big data can be described as the exponential growth and wide availability of discrete, continuous, categorical, or hybrid data, which are difficult or even impossible to manage and analyze using conventional software tools and technologies [5,6]. Furthermore, estimate shows that 30% of the world storage was occupied by medical images in 2011 and will progressively increase in subsequent years [2,7]. This shows the extremely large and often underestimated amount of data produced in medical institutions. Mobile health (mHealth) is referred to as one of the technological breakthroughs in this decade [8]. The global proliferation of mobile devices and health applications has made mHealth synonymous with big data. This large amount of unutilized generated data calls for attention.

Big data provides the opportunity for health policy experts, physicians, and health care institutions to make data-driven judgments that will enhance patient treatment, disease management, and health care decisions. Many experts have used internet tools for big data services and related applications. This is depicted in the graph in Figure 1, which was obtained from Google Trends for "big data in healthcare" between 2010 and 2018. Google Trends is a free Web service by Google Inc that provides statistical occurrence of activities by people on the internet all over the world. The trend in the graph is calculated as interest over time on a scale from 0 to 100, where 100 refers to the maximum computed score for total search and related activity for the topic.

The first graph in Figure 1 shows the continuous rise in activities regarding big data in health care, and the top 5 countries where it was most popular is given in the second graph, with India, United States, and United Kingdom leading the occurrence chat. The size of medical data is too large for comprehensive analysis with the available analytical tools to maximize the knowledge available in big data. Traditional machine learning (ML) techniques and algorithms have limited capacity to utilize big data and, in most cases, the solution becomes complex and undesirable. Deep learning (DL) is proposed and provides a prospective solution to this challenge. Figure 2 shows the performance between DL and other ML techniques in the situation of increasing data size. The primary advantage of DL is that the performance of large architecture of DL increases with increase size of available data [9].

The main question will be what is DL? Human experts in a specific domain have ample knowledge about the subject in that domain. The limitation with human experts is because of their subjectivity, large variations across interpreters, availability, and fatigue [10,11]. To help the accomplish task performed by humans and overcome these limitations, intelligence demonstrated by humans is built into machines and computers to create the concept of artificial intelligence (AI). ML is a branch of AI that gives computers the ability to learn and perform the role of experts without being explicitly programmed [12,13]. Some examples of ML include support vector machine (SVM), decision tree, logistic regression, Naïve Bayes, K-means clustering, and so on. On the basis of a broad classification scheme, ML can be categorized into 3 groups [13-15]. The first is supervised learning; the computer learns the classification system from the class labels provided. The second is unsupervised learning, where no labels are given; the purpose is to program the computer to do things without telling it how to do it. The third is semisupervised, where the computer learns from a combination of available and unavailable labeled data; usually the size of unavailable labeled data for learning is larger. The recent hunger for data consumption and analysis has opened up new frontier for more ideas and applications [16]. Artificial neural network or neural network (NN) is another example of ML, with an interconnection of nodes called neurons with 3 major layers, input, hidden, and output layers, where the hidden layer is a single layer that connects the input layer to the output layer. The purpose of NN is to gradually approximate a function that maps an input to a corresponding output through an iterative optimization process. NN has transformed from its inception as a simple perceptron to solve simple problem to the advanced concept of deep neural network (DNN), which has many cascaded interconnected hidden layers that are able to process and analyze audios, text, signals, images, and more complex data types. DL is the recurrent learning process performed in DNN that enables it to find an optimal function for representing data. The innovation of DL is a developing trend in data analysis and is ranked as one of the best inventions in technologies [17]. DNN is an active branch of ML and its goal is to make machines think and understand as humans by mimicking the grid of the human brain connection, to focus on learning data representation (DR) rather than task-specific algorithms [14]. Figure 3 reveals the relationship between DL, ML, and AI. Currently, DL is set to take over the ML space because of its increasing attention and performance.

Figure 1. Google Trends for "big data in healthcare" between 2010 and 2018; (a) occurrence timeline graph and (b) prevalence occurrence by country.

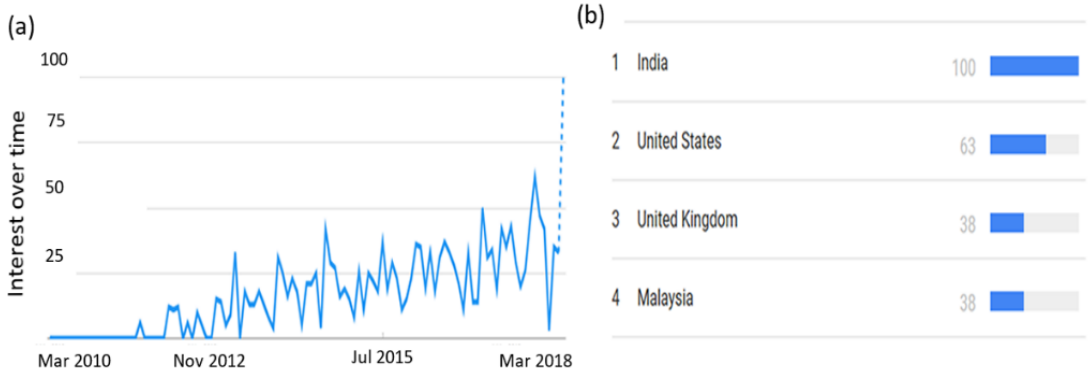


Figure 2. How machine learning techniques scale with amount of data.

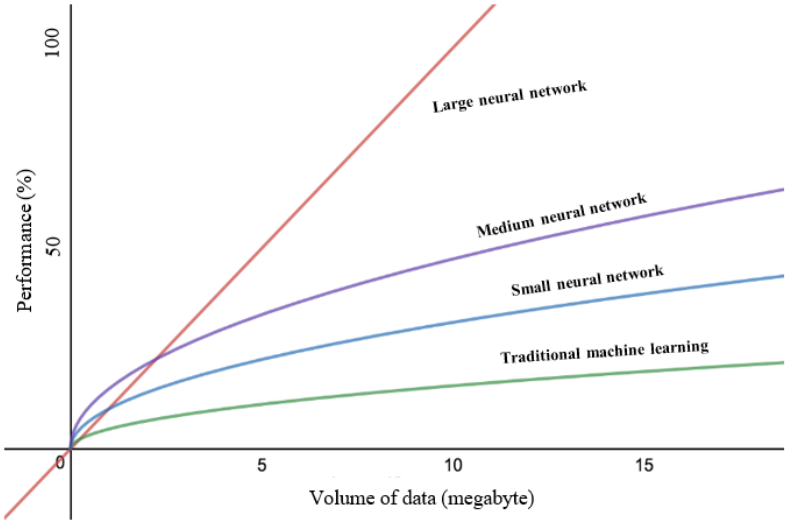
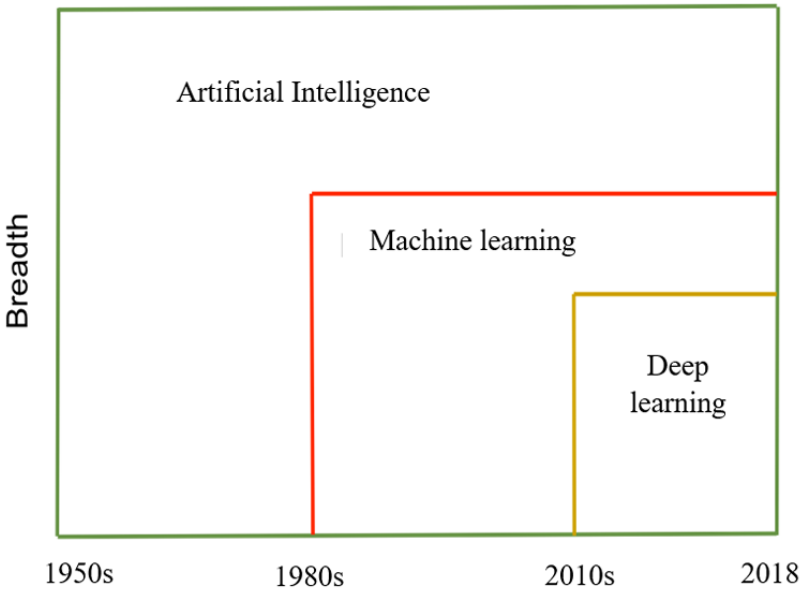


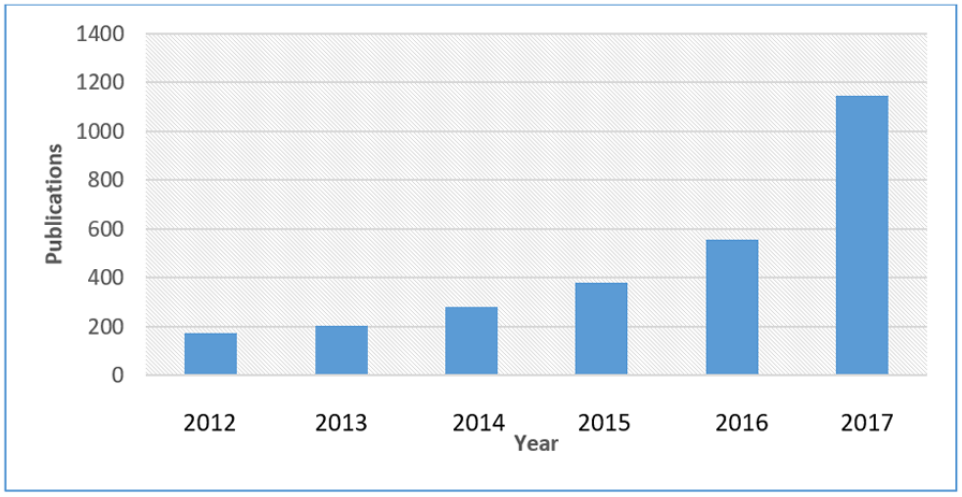
Figure 3. Relationship between artificial intelligence, machine learning and deep learning with emerging timeline.



ML and DL have in recent times attracted a lot of awareness from different sectors such as academia, industry, media, security, and government alike, and its impact on biomedical and health care cannot be over emphasized. DNN has been applied to solve many traditional problems where available large data need to be analyzed and many impressive results have been reported in different areas such as medical image processing [18], speech analysis [19], and electronic health record (EHR) translation [20,21]. Figure 4 shows the trend in application of DL found in research publications. The number in 2017 is almost twice that of 2016. The trend observed in the figure is a result of the performance and results achieved by DL. Therefore, we can say that more areas and domains are moving toward DL to achieve high performance and better results.

Currently, DL has started making huge impact across different areas in health care. The increasing availability of health care data and rapid development of variations in DL techniques have made it possible to have the impressive results recorded in health care [22,23]. DL techniques can reveal clinically relevant information hidden in large amount of health care data, which in turn can be used for decision making, treatment, control, and prevention of health conditions. Some application areas of DL include health behavior reaction [24,25], EHR processing and retrieving scientifically sound treatment from text [26,27], eye related analysis and classification [28-30], gait analysis and robotic-assisted recovery [31,32], hearing disorder treatment [33], cancer treatment [34,35], heart diagnosis [36,37], and brain activity analysis [38-40]. This makes the treatment easier for health care provider and convenient for patients, with faster and productive monitoring. The advancement in DL in medicine has translated the use of simple equipments, such as thermometer and stethoscope, into computed tomography (CT), ultrasound diagnostic devices, radio nuclear imaging, radiation therapy, lithotripsy, dialysis, ventilators, and so on, which have taken conventional patient care to highly adaptive treatment, capable of challenging many dreaded diseases [23]. There is no doubt that in the coming years, health care treatment and equipment will witness greater improvements in many more areas, to make it more effective with qualitative services.

Figure 4. Trends of published papers that implement deep learning techniques. The data are generated by searching for "deep learning" on PubMed database.



Compared with the traditional ML algorithm, the depth of learning and feature extraction in DL has unparalleled superiority. The deep network structure can realize the approximation of complex functions through nonlinear transformation in the hidden layers. From low to high level, the representation of features is more and more abstract, and the original data can be characterized more accurately [41]. A large number of experimental works have applied DL models and techniques and there are variants of DL models. The goal of this paper is not to show all the techniques and models, but to highlight the important principles and the applications of DL in health care and medical field. A simple feedforward DNN architecture is the autoencoder (AE) that comprises encoder and decoder functions for input and output layers, respectively. Convolutional neural networks (CNNs) have had the greatest impact within the field of health informatics [42]. Its architecture can be described as an interleaved set of feedforward layers implementing convolutional filters followed by reduction, rectification, or pooling layers. For each layer, the CNN creates high-level abstract feature. Another variant of DL is the recurrent neural network (RNN), which is a sequential data NN with an inbuilt memory that updates the state of each neuron with previous input. The deep belief network (DBN) model has only several layers of hidden units and there is connection between each unit in a layer with each unit in the next layer. Another architecture is the deep Boltzmann machine (DBM) which has completely undirected connections, unlike DBN, between neurons in all layers. DL is computationally intensive. The success and proliferation recorded in DL can be attributed to the advancement in graphics processing units (GPUs), which play a significant role in accelerating the computation requirement of DL [43,44].

The proceeding sections of this paper are organized as follows. We discuss 5 common DL techniques and their basic principle of operation in the next section that describes DL methods. A review of literature in health care and biomedical domains that have applied DL was examined and presented in the section review of DL implementation in health care. In the section challenges in health care for DL applications, we discuss challenges and setbacks encountered in the application of DL and plausible solutions. In the section future trends for deep learning, we present critical discussion about the future trends for DL algorithm for health and biomedical field, and the conclusion section of the paper closes the discussion.

Deep Learning Methods

Basic Principles of Operation

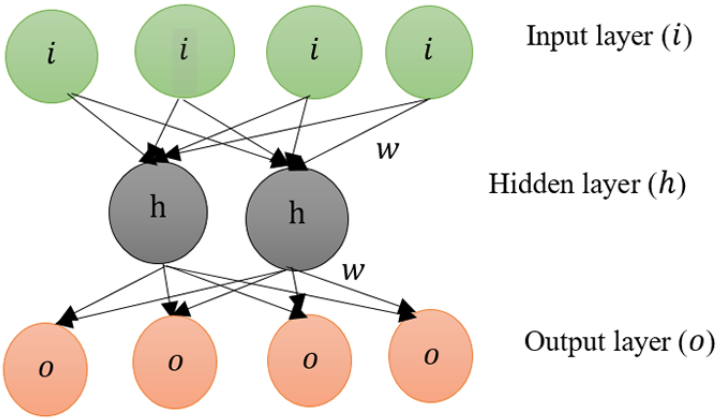
In this section, we describe the principles of operation of 5 DL models. There are several variants for each model. The underlying principle is to approximate a function that produces the expected output for a given input. The different models are more suited to handle different challenges and for different kinds of data type and expected task to be performed. The model more suited for image classification is different from speech or time series classification. Some models can be applied as a preprocessing phase to reduce the dimensionality of the data. Basically, the structure of the model comprises interconnected neurons, connecting the input to the output, known as the hidden layer. Therefore, this produces a sequence of activation through the weighted connection from neurons perceiving the environment (input); this is referred to as feedforward [45]. The differences between DL model and NN include the use of more hidden layers in DL compared with NN, which only has 1 or 2 hidden layers. DL can be trained for both unsupervised and supervised learning tasks but NN can only be trained for supervised learning task. At the end of the feedforward process, the result from the output unit is evaluated with the expected value. This evaluation will produce an error value that will lead to the adjustment of connected weight working backward from the output layer to the hidden layer and to the input layer until the output is close to the expected result. This procedure is referred to as backpropagation [46,47].

Autoencoders

AE is designed for feature extraction using data-driven learning. It is trained in an unsupervised manner as it is trained to recreate the input vector rather than assign class label. The normal design and structure of AE is to have the same number of neurons in the output and input layers, with full connections between neurons in each layer to subsequent layer as shown in Figure

5. The number of neurons in the hidden layer is smaller than the input and output layer. The purpose of this structure is to encode data in low dimensionality space and to achieve extraction of features. However, where the dimensionality of the data is high to achieve the same purpose, many AE can be stacked together to create a deep AE architecture. There are many deviations of AE presented over the last decade to handle different data patterns for performing specific functions. For example, denoising AE was first proposed by Vincent et al [48]. The purpose was to increase the robustness of the regular AE model. The method recreates the input introducing some noise to the patterns, thus forcing the model to capture just the structure of the input. Another variation is the sparse AE that forces the representation to be sparse, which is used to make the data more separable [49]. Another idea proposed shared weights between nodes to preserve spatial locality and process 2-dimensional (2D) patterns called the convolutional AE [50]. Contractive AE is similar to denoising AE, but instead of injecting noise to corrupt the training set, it modifies the error function by adding analytic contractive cost [51,52]. The learning process for AE is described as minimizing a loss function L such that $L(i, g(f(i)))$. $f(i)$ is a function that maps i to h and the function g maps h to the output which is a reconstruction of the input. w is the weight connecting the layers.

Figure 5. Structure of a simple autoencoder showing input, hidden, and output layers. The interconnection between the neurons is shown in the direction of the arrows.



Recurrent Neural Network

This class of DL has connections between neurons in the hidden layer to form a sequence of directed graph. This feature gives it a temporal dynamic state. This is important in applications where the output depends on the previous computations such as the analysis of text, sounds, DNA sequences, and continuous electric signals from the body. The training of RNN is performed with data that have interdependencies to maintain information about what occurred in the previous interval. The performance result at time *t-1* affects the choice at time *t*. It considers the previous output (O_{t-1}) and current input (I_t) and produces a number between 0 and 1 from the cell state M_{t-1}, where 1 represents *save this value* and 0 represents *dispose this value*. This decision is made by a sigmoid layer called the *gate layer*.

Therefore, the principle of RNN is to define recurrent relation over time steps which can be approximated with the formula: $M_k = f(M_{k-1} \times W_r \times I_k \times W_i)$, where M_k is the state at time k, M_{k-1} is the output of the previous state, I_k is the input at time k, and W_r and W_i are the weight parameters in the network. As a result, RNN can be viewed as a state with feedback loop. The final output of the network O at a certain time step k is typically computed from one or more states such as $M_{k-1}\ldots M_{k+j}$ and $j=1,2,\ldots k-1$.

Hence, the result of new data is dependent on 2 sources of input, the present and the recent past. Owing to this principle, RNNs are said to operate with memory [53]. Figure 6 shows a sample of RNN structure and the connection between neurons in each layer. Beside the structural difference, RNN uses the same weight across for all layers, but other DL uses different weights. This significantly cuts down the total number of parameters that the network needs to learn. Despite the successful application of this model, the setback includes vanishing gradient by long input sequence and exploding gradient problems as described in [54]. To handle the limitation, long short-term memory unit (LSTM) was invented by [55]. Specifically, LSTM in Figure 7 is particularly suitable for applications where there are very long time lags of unknown sizes between important events.

To achieve this, LSTMs utilize new sources of information so that data can be stored in, written to, or read from a node at each step. During the training, the network learns what to store and when to allow either reading or writing so as to minimize the classification errors [56]. Another variant of RNN is the gated recurrent unit, which is a simplified model of LSTM with an equal performance as LSTM [57].

Figure 6. Feedforward recurrent neural network implementation. The final output from the output layer is fed back as part the input in the input layer. Where It and Ot are the input and output at time t and Ot-1 is the output for the previous input at time t-1.

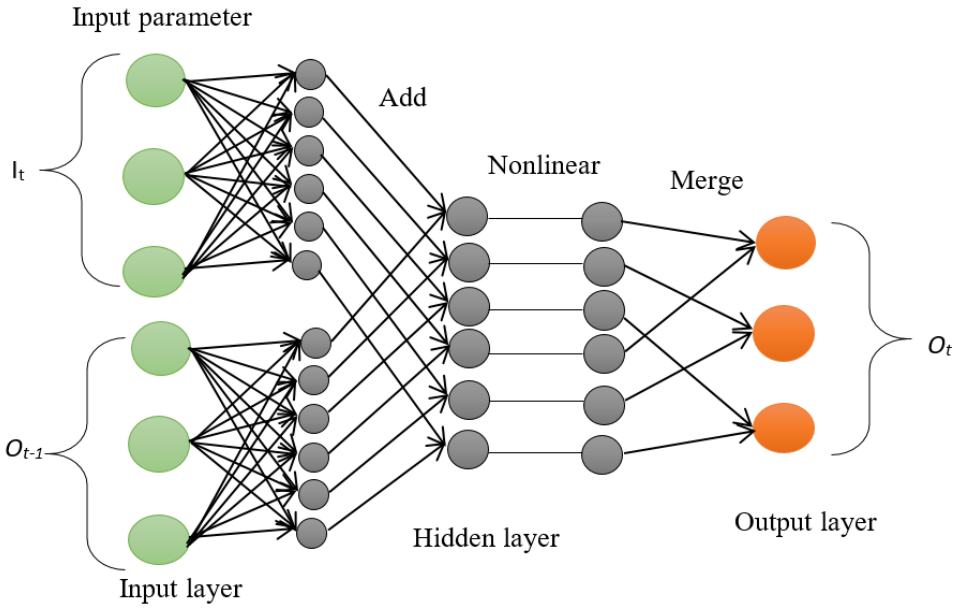
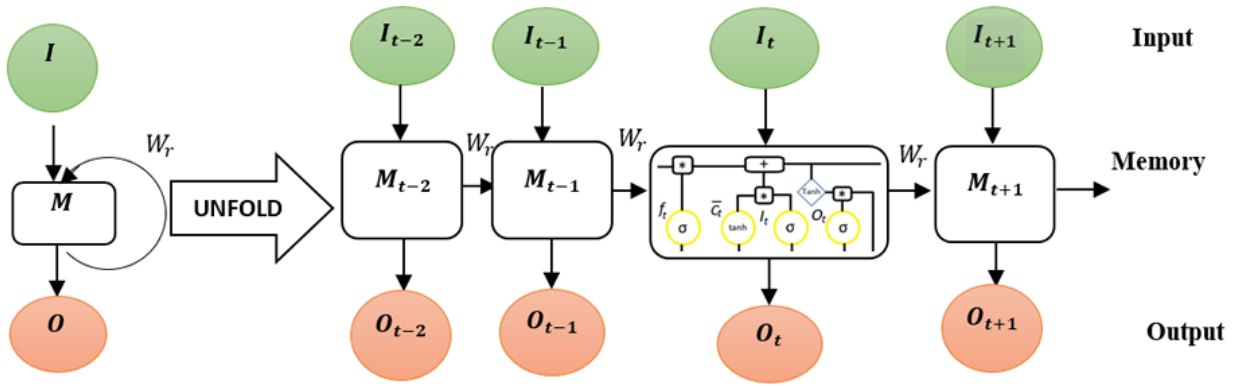


Figure 7. Long short-term memory representation for output sequence influence by input sequence and previous output. Where It-2 and It-1, Ot-2 and Ot-1, Mt-2 and Mt-1 are inputs, outputs, memory respectively for previous time steps. It, Ot, and Mt are the current input, output and memory state of the LSTM cell. Ot+1, and Mt+1 represent subsequent output and memory state respectively for subsequent time step input It+1. I, O, M represent the recurrent input, output and memory state respectively for a simplified LSTM cell operation and Wr is the weight for the computation in the cell.



Convolutional Neural Network

CNN was inspired by biological processes of the human brain, where the connectivity pattern between neurons resembles the concept of the human visual cortex [58,59]. A typical CNN comprises an input, multiple hidden layers, and an output layer. The hidden layers of a CNN usually comprise the following constituents: convolutional, pooling, fully connected (FC), and normalization layers. An example of CNN was proposed to analyze imagery data [60]. Figure 8 shows a simple implementation to identify a character from a 3×3 pixelated matrix image sliding 2 filters of size 2×2 square matrix (kernel=2) with stride of 1. The example is designed to recognize *X*, O, **,** and / characters. The convolution layer applies the filter across the input image. The operation is performed with 2 filters over the input image and having the same weight. This produces a total of 8 parameters. In this example, the bias is omitted for simplicity. Often, a nonlinear (activation) layer is added after the convolution layer, usually rectified linear unit (ReLU). The activation layer applies the function *f(x)=max(0,x)* to all the values from the convolution layer. This process increases the nonlinear properties of the model and the overall network without affecting the receptive fields of the convolution layer. In this way, it resolves the vanishing problem compared with training traditional multilayer NNs with backpropagation. Pooling layer combines the output of neuron clusters in the convolution layer into a single neuron [56]. This is sometimes achieved by using max, sum, or average pooling, which consider the maximum, sum, or average value from each cluster of neurons, respectively [61]. FC layers connect the neurons in the previous layer to the neuron in the final layer by translating input image into a single vector for classification. This layer holds the filter that is used to determine the class of the input image. The output with the highest value is assigned the class label. The main benefit of a CNN is that during backpropagation, the network has to adjust a number of parameters in the filter with techniques such as gradient decent, which drastically reduce the connections of the CNN architecture.

In Figure 9, the general architecture of a simple CNN is presented, which shows the input, convolution+pooling, FC layer, and output layer. The convolution+pooling is responsible for feature extraction. The FC layer acts as a classifier on top of the features and assigns a probability score for the input image to define the output. The input to the convolution layer is an $m \times m \times r$ image, where m is the height and width of the image and r is the number of channels. k is the filters (or kernels) in each convolution layer of size $n \times n \times q$, where n is smaller than the dimension of the image and q can be the same as r. The dimension of k can be $m - n + 1$ which form the size of the filter (a locally connected structure). Each map is then subsampled with mean or max ($f(x)=max(0,x)$) pooling over contiguous region (x); additive bias and sigmoidal nonlinearity is applied to each feature map.

FC layer represents the feature vector of the input, a composite and aggregated information from all the convolution+pooling layers. Each node in the FC layer learns its own set of weights on all of the nodes in the layer below it. The final feature vector is used to predict the input image.

Figure 8. Simple implementation of convolutional neural network to show the sequence of operation to identify "X" with 2 filters.

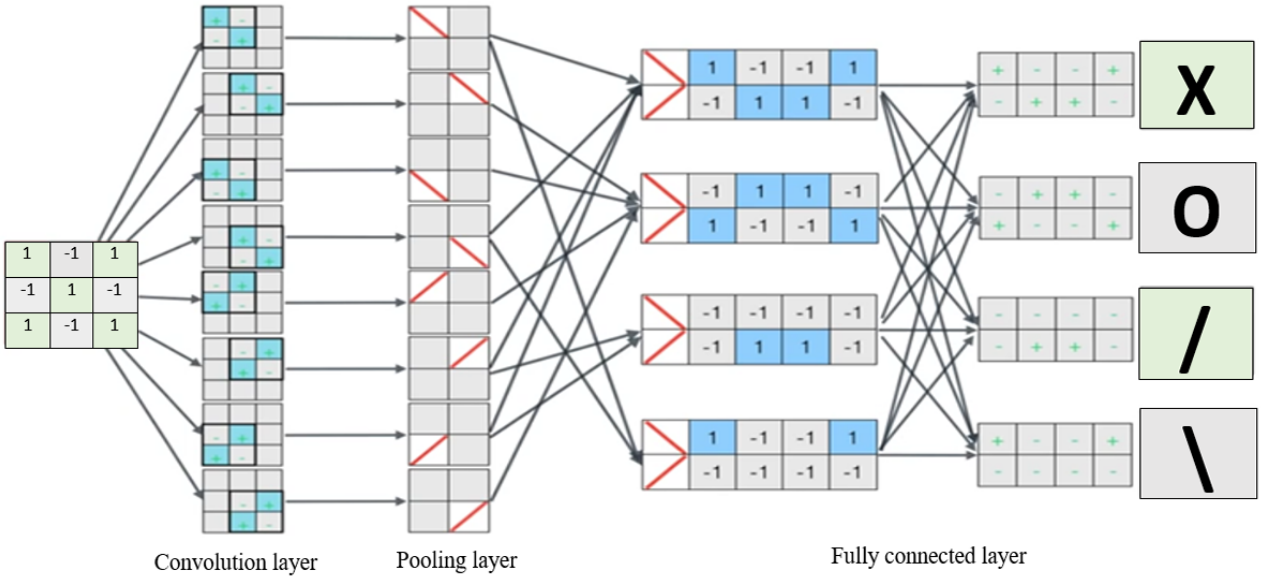
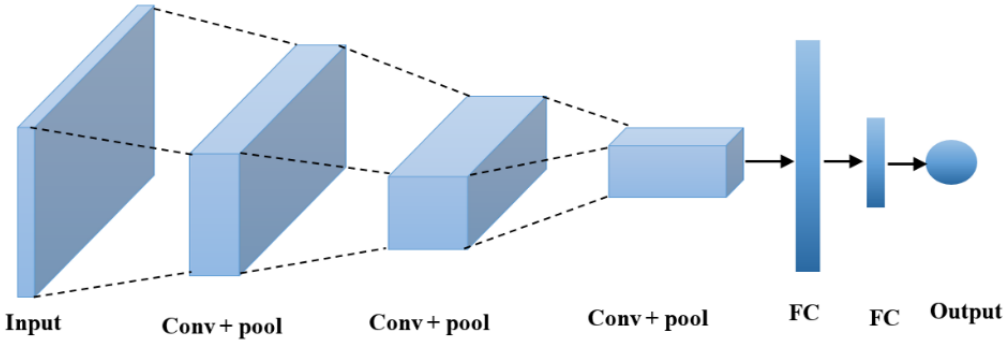


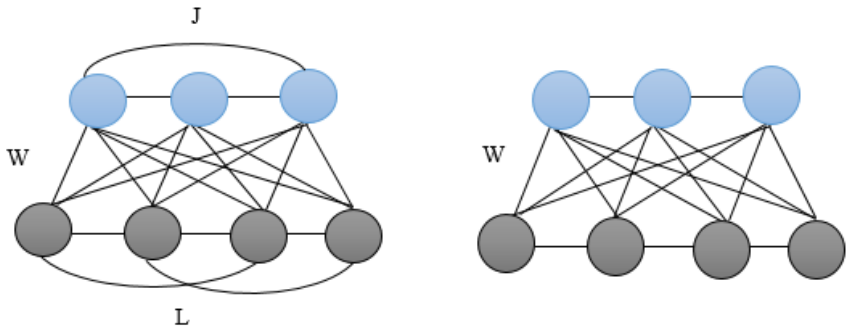
Figure 9. Structure of convolutional neural network with 3 convolution and pooling (conv+pool) layers and 2 fully connected (FC) layers.



Deep Boltzmann Machine

A Boltzmann machine (BM) is a network of symmetrically coupled stochastic visible and hidden units. The first diagram in Figure 10 shows the structure of BM, where the labels W, L, and J represent visible-to-hidden, visible-to-visible, and hidden-to-hidden symmetric interactions, respectively. BM model is suitable for modeling and extracting latent semantic representations from a large unstructured collection of documents [62]. The original algorithm for BM requires randomly initialized Markov chains to achieve equilibrium distributions to evaluate the data-dependent and data-independent expectations in a connected pair of binary variables [63]. Learning procedure is very slow in practice using this system [62]. To achieve an efficient learning, restricted Boltzmann machine (RBM) was created, which has no connections between hidden units [64]. The second diagram in Figure 10 shows a simple architecture of RBM with connections between neurons. A beneficial feature of RBM is that the conditional distribution over the hidden units factorizes, given the visible units. This makes inferences tractable as the RBM feature representation is taken to be a set of marginal posterior distributions obtained by directly maximizing the likelihood. Furthermore, 2 main DL frameworks in this category that have been presented in literatures are DBM and DBN [42].

Figure 10. Left: A general Boltzmann machine. The top layer represents a vector of stochastic binary "hidden" features and the bottom layer represents a vector of stochastic binary "visible" variables. Right: A restricted Boltzmann machine with no hidden-to-hidden and no visible-to-visible connections. Where L, J, and W represent the visible layer, hidden layer, and connection weight between the layers respectively.



The architecture of DBM NN is similar to RBM but with more hidden variables and layers. DBM architecture has entirely undirected connections between neurons within all layers [62]. The right image in Figure 10 shows the architecture of a simple DBM NN for 1 visible layer and 1 hidden layer. It has undirected connections between all layers of the network, but not within the neurons in a layer. For training a DBM, a stochastic maximum probability–based algorithm is usually applied to maximize the lower bound of the probability. This is because calculating the distribution over the posterior hidden neurons, given the visible neurons, cannot be achieved by directly maximizing the likelihood because of the interactions between the hidden neurons.

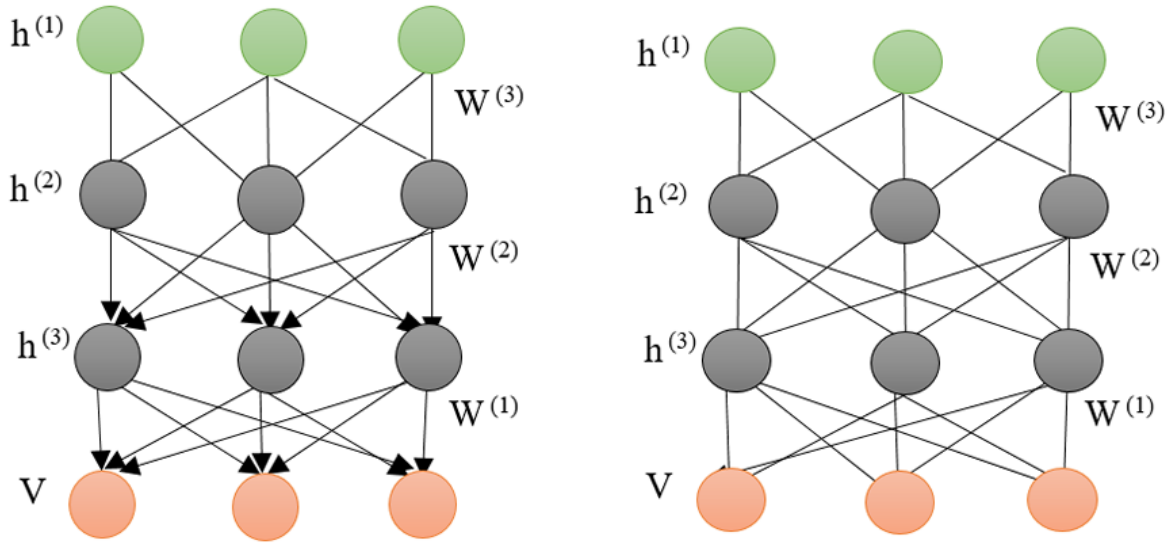
Implementation of DBM is remarkable as DBM has the capability to learn internal representations that become increasingly complex, which is regarded as a promising way of solving recognition problems. Moreover, in cases of semisupervised learning, high-level representations can be built from very limited labeled data and large supply of unlabeled inputs can then be used to fine-tune the model for specific task. In addition, to enable DBM propagate uncertainty and hence deal more robustly with ambiguous inputs, it can incorporate top-down feedback, in addition to an initial bottom-up pass.

Deep Belief Network

DBN is another variant of RBM, where the multiple hidden layers can learn by treating the hidden output of one RBM as the input data for training the next layer of RBM [63,64]. It has undirected connections between its top 2 layers and directed connections between all its subsequent layers. The training strategy is greedy layer wise, which is performed when training the DBN using unsupervised learning and adjusting its parameters based on the expected output. The left diagram in Figure 11 illustrates the architecture of DBN with 3-layer configuration showing visible-to-hidden and hidden-to-hidden symmetric connections. The structure comprises several hidden layers of neurons, which are trained using backpropagation algorithm [65,66].

From Figure 11, the connection units in the DBN architecture is between each neuron in a layer with each neuron in the next layer; however, unlike RBM, there are no intraconnections among neurons within each layer.

Figure 11. Left: A 3-layer deep belief network. Right: A stack of modified restricted Boltzmann machine constructed to create a deep Boltzmann machine. V: visible vector; h: a set of hidden neurons; w: connections.



Trends in Deep Learning Methods

The use of different DL techniques is proliferating into more domains in biomedical engineering applications. This is because of the achievements recorded in previously implemented applications. Figures 12 and 13 describe the trends in the use of different methods of DL over 5 years, from 2012 to 2017. The purpose of constructing the trend is to observe the implementation of DL methods over a period of time, and the choice of publication was based on implemented DL methods without any specific application domain. These statistics are obtained from 2 different sources, PubMed database and the Institute of Electrical and Electronics Engineers archive. Both Figures 12 and 13 show similarity in the pattern of increasing growth in the use of DL.

The observable pattern in both Figures 12 and 13 shows that RNN and CNN have a steady increase in application over the years, with CNN exhibiting tremendous growth rate. This can be attributed to the success recorded in image data and the many available variants of the model. Positron emission tomography and CT scan image processing are at the forefront of many health care applications. CNN has provided the needed processing techniques required to achieve expected performance. The growth rate in the application of this technique is expected to continue as more biomedical image applications will switch to this technique. Nevertheless, the growth rate is expected to slow down after a while as many applications would have migrated to this technique. Another DL method that has also shown promising performance is AE. The steady increase in the number of publications in Figure 12 and 13 indicates the successful implementation results and efficiency. DBN and BM have the least progression. Despite the small positive difference between successive years, the major challenge is in training the NN, which is computationally expensive. Another reason for the low rate of application is because of its combination with other methods and as it is sometimes implemented as a preprocessing phase. The graph in Figure 12 is obtained by searching for the DL techniques in publication title and abstract from PubMed. The result returns graphical statistics of publications grouped by years. An advanced search technique is applied to obtain Figure 13, where the query method is similar to the previous approach, but the query string is applied in title and abstract search fields. Therefore, the final query becomes ((*Publication Title*: Autoencoder) OR (*Abstract*: Autoencoder)). However, to get the total publication for each year, a manual filtering is used to get the distribution. The total publication is made up of both journals and conference articles.

Figure 12. Research publications in different category of deep learning methods. These statistics are obtained from PubMed database by searching for publications containing any of the deep learning method in title or abstract.

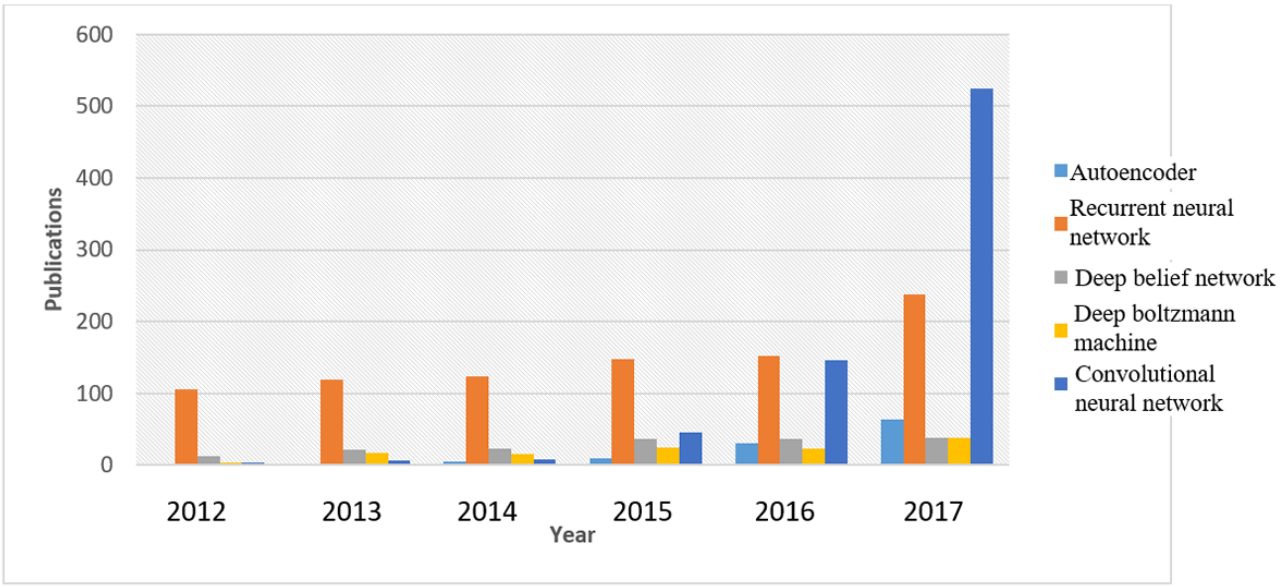
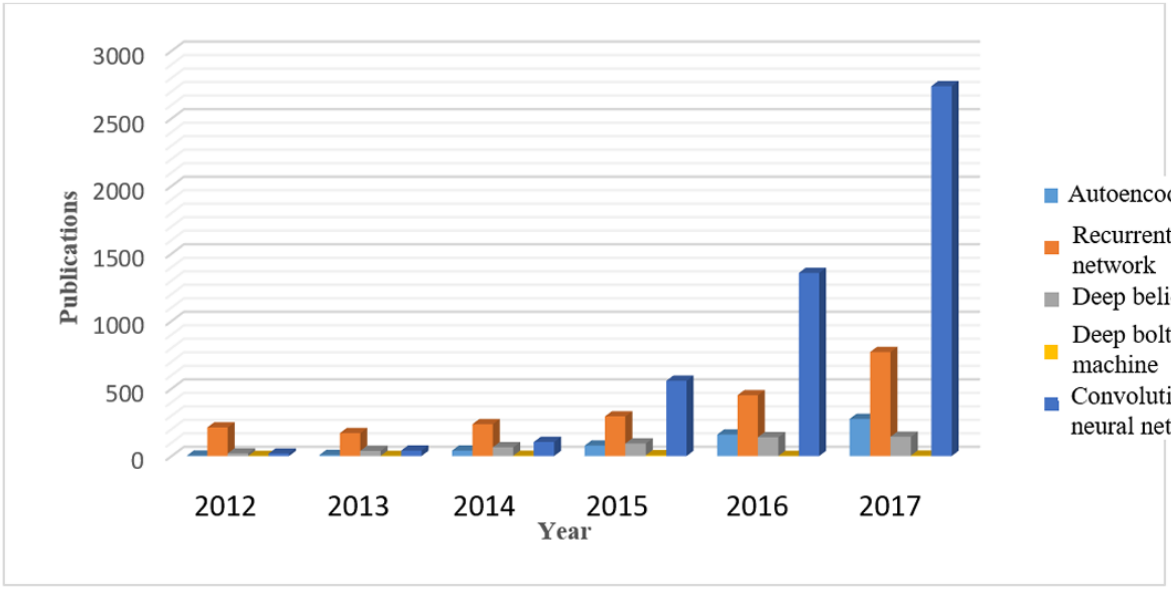


Figure 13. Publications distributions for 5 years in different categories of deep learning methods. These figures are extracted from the Institute of Electrical and Electronics Engineers database of papers from conferences and journals and magazines by using advanced query to search in publication title and abstract containing any of the deep learning methods ((“Publication Title”: Autoencoder) OR (“Abstract”: Autoencoder)).



Comparative Analysis of Deep Learning Methods

The results reported in Figures 12 and 13 for different architectures of DL are based on the conceptual advantage of each method. Despite the fact that each DL model is more suitable for a particular kind of data or situation, there is however some relationship and cross-application of these methods. The fact that RNN and CNN are 2 commonly used DL techniques can be attributed to the need to solve data problems in the form and shape that only these techniques can handle effectively. In addition, most of the common data are either visual or time-dependent. Although CNN is a feedforward NN where information only flows in one (forward) direction, in RNN, the information flows back and forth as it operates on the principle of saving the output of the previous layer and feeding this back to the input to predict the output of the current layer. The principle of operation of CNN is influenced by the consecutive layer organization of the animal visual cortex. Therefore, it is designed to learn to recognize patterns across space. This makes CNN ideal for images (eg, 2D or 3-dimensional [3D] magnetic resonance images [MRI]), videos (eg, gait pattern, moving pattern in organs), and graphics (eg, tumor representation) to recognize features such as lines, edges, curves, and so on. On the contrary, RNN technique is suited to recognize patterns across time such that the information available now will subsequently influence what information will become available later. This makes it suitable for time series analysis such as sound (eg, heartbeat, speech), text (eg, medical records, gene sequence), and signals (eg, physiological signals such as electrocardiogram (ECG)).

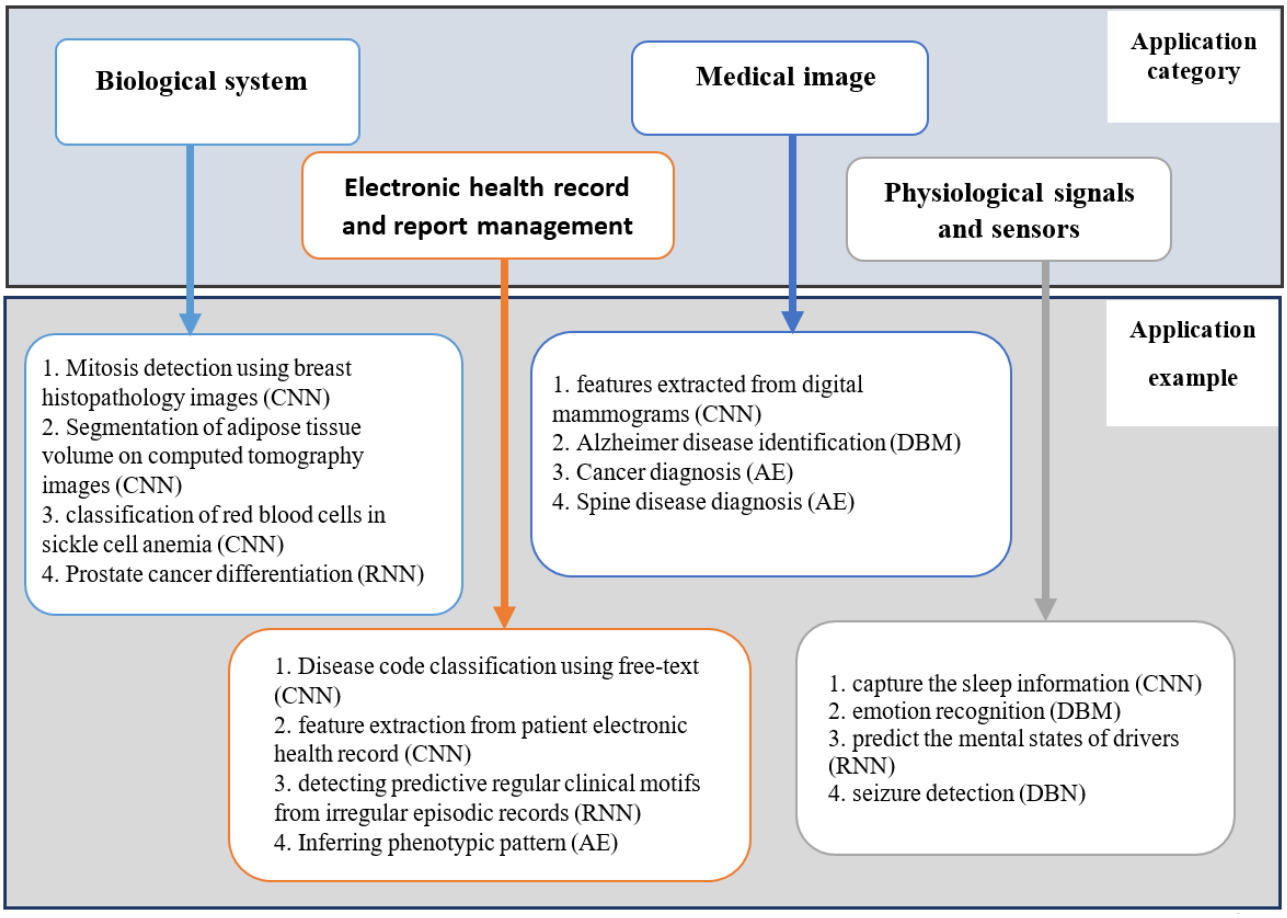
There is a close comparison between DBN and DBM architectures. Although there is a diagrammatic similarity between DBN and DBM, similar to conventional DNN, both methods are deep, which means it is possible to create many hidden layers connecting the input and output unit. In addition, there is the presence of RBM in both architectures. However, they are qualitatively different. The connections between the layers in a DBN are directed, whereas it is undirected in DBM. The first 2 layers in a DBN is undirected connection of RBM; the subsequent layers are directed generative connections. However, in DBM, all the connections between the layers are undirected RBM. Another difference between these 2 techniques can be described in a general picture with the connected layers, where the connected layers in DBN function as sigmoid belief network but in DBM they are Markov random fields. DBN and DBM can be used to extract features from unprocessed physiological signals and image data (MRI) to reduce the size of features required for classification modeling. These models can also be applied as a generative model for human motions completion for fall detection or gait analysis.

The main function of AE architecture is to reconstruct the input data given to it. Therefore, if a vector M is given to AE, it tries to create $M^i=h(g(M))$ so that during training it obtains the parameter for h and g such that M^i is the same as M. In DBN and DBM, h and g exist between the input and hidden layers, although, to compute these functions a probabilistic (or Markov chains) approach is applied. However, unlike AE, there exists a special connection between h and g to make it a valid probabilistic model. In addition, although the model from AE ensures that the input and output are the same, DBN and DBM give a range of outputs for a given input it has been trained on because of its probabilistic principle. The similarity between AE and RBM model is to encode the visible layer with hidden layer in a constructive functional way; encoding the hidden layer with another hidden layer leads to stack AE and RBM (DBN and DBM). Considering the growing size of medical data, an efficient data coding with AE makes it possible to minimize memory requirements and reduce processing costs. Stacked AE can be applied in unsupervised feature learning for detection of tumors, cancers, or inflamed organs.

Review of Deep Learning Implementation in Health Care

This section reviews some health and biomedical areas that have successfully implemented DL techniques to create a model to solve specific task. We considered the DL methods discussed in the previous section and have presented a tabular representation of references for application in 4 areas: biological system, EHR and report management, medical image, and physiological signals and sensors. The tables provided represent the summary of applications of DL methods in each of these categories. The choice of literatures is selected from papers published between 2012 and 2018 that are related to health and medical applications. The purpose is to reveal some of the applications of DL methods that have been designed to solve biomedical-related tasks which initially have poor results with other techniques, such as handcrafted, or which seem unsolvable because of the complication of the task. Figure 14 shows an illustrative summary of application areas and implemented DL methods. It is an overview of the information presented in the tables provided. The figure is divided into 2 distinct blocks, application category and application example. The connection between the 2 blocks is created from the relationship between the content in the application category and application example.

Figure 14. Descriptive summary of biomedical and health applications category and example implemented with deep learning methods: convolutional neural network (CNN), recurrent neural network (RNN), autoencoder (AE), deep boltzmann machine (DBM) and deep belief network (DBN).



Biological System

Biological records such as DNA, RNA, genomes, sickle cell behavior, bacterial and viral multiplication, and mutation have features that make it possible to create a predictive model with DL algorithms to achieve performance exceeding human experts. Prediction could be in the form of discriminative gene identification, structured data in protein-protein interactions, image data from biological cell activities, drug composition–reaction profiling, and binding between DNA and proteins or between protein sequences. The structure of the data and expected goals determine the DL technique to be implemented. RNN, CNN, DBN, and AE methods have largely been applied in many aspects. Table 1 shows DL implementation in biological systems.

Biomedical analysis has benefited from CNN architecture. An example can be found in cell mitosis described in [67]. The CNN architecture primarily comprises 5 convolution layers, 4 max-pooling layers, 4 ReLUs, and 2 FC layers. The activation function used after each layer is ReLU and to avoid overfitting of the model, dropout layer was included after the first FC layer. In [68], a proposed CNN architecture for cell membrane and nuclei classification for breast cancer was developed, which comprised convolution and deconvolution sections. The framework mainly comprised multiple convolution layers, max-pooling layers, spatial pyramid pooling layers, deconvolution layers, upsampling layers, and trapezoidal LSTM. To achieve automatic red blood cell classification, [69] constructed a CNN architecture with alternating convolution and pooling operations to deal with nonlinear and complex

patterns. A combination of 2 CNN techniques was constructed in [70], selection-CNN and segmentation-CNN, to achieve segmentation of adipose tissue volume on CT images.

Moreover, RNN architecture has been used in biomedical process to efficiently model problems with sequence and time feature. Sequence-specific bias correction problem for RNA sequence data was addressed using RNN architecture to model nucleotide sequence without predetermined sequence structures [75]. A variant of combined CNN and LSTM was created for prostate cancer identification with Gleason score of 7 [76]. The model assesses the correlation between histopathology images and genomic data with disease recurrence in prostate tumors to identify prognostic biomarkers within tissue by modeling the spatial relationship from automatically created patches as a sequence within the tissue. In [77], biomedical LSTM and conditional random fields (CRF) were combined to create a relationship between biological entities for trigger detection. The method is based on the sequence annotation that does not require initial complex feature engineering but only requires a simple labeling mechanism to complete the training.

The synergistic effect of drug combination is one of the most desirable properties for treating cancer. A method based on DBN was created to predict drug synergy from gene expression and pathway and ontology fingerprints [78]. A multiclassifier DBN technique was proposed to detect mitotic cells in hematoxylin and eosin-stained images using step by step refinement of segmentation and classification stages [79]. The multiclassifier DBN algorithm segments cell nuclei from background stroma. Critical proteins exhibiting dramatic structural changes in dynamic protein-protein interactions networks were identified in [80] using a DBN framework; the reconstruction errors and the variabilities across time were analyzed in the biological process. In [81], a DBM framework called stacked RBM was proposed to analyze the RNA-seq data of Huntington disease. In addition, the framework was able to screen the key genes during the Huntington disease development. The initial step for the framework was to select disease-associated factors with different time period datasets according to the differentially activated neurons in hidden layers. Then, the disease-associated genes were selected according to the changes of the gene energy at different time periods.

Furthermore, AE technique has been considered as a method to delineate signals from noise in imaging to enhance image quality. In [83], a proposed deep count AE to denoise small cytoplasmic RNA sequence (scRNA-seq) datasets was created. The model takes the count distribution, dispersion, and sparsity of the data into account using a negative binomial noise model with or without zero inflation. The technique was capable of capturing nonlinear gene-gene dependencies. Another design of AE model proposed was called stacked denoising AE for multilabel learning [84]. The purpose was to facilitate gene multifunction discovery and pathway completion. The technique can capture intermediate representations robust to partial corruption of the input pattern for cancer research, pathway analysis, and gaining insight into the underlying biology associated with cancer. A gene superset AE [85], a multilayer model with the incorporation of a priori defined gene sets, retains the crucial biological features in the latent layer. The method introduced the concept of gene superset, an unbiased combination of gene sets with weights trained by AE, where each node in the latent layer is a superset. Furthermore, to analyze the transcriptomic heterogeneities at the single cell level, a deep variational AE technique for scRNA-seq data was proposed [86]. The technique is a deep multilayer generative model for unsupervised dimension reduction and visualization of scRNA-seq data. The AE technique can explicitly model the dropout events and find the nonlinear hierarchical feature representations of the original data.

Table 1. Deep learning implementation in biological systems.

Reference	Task	Method	Remark
Saha et al, 2018 [67]	Mitosis detection	CNN[a]	The prediction model has an improved 92% precision, 88% recall, and 90% F-score over conventional machine learning methods
Saha et al, 2018 [68]	Cell membranes and nuclei classification	CNN	The identification model achieved predictive value of 98.33% accuracy and 6.84% false-positive rate which was comparable with human expert
Xu et al, 2017 [69]	Red blood cells classification	CNN	Framework classified sickle-shaped red blood cells in an automated manner with above 90% accuracy
Wang et al, 2017 [70]	Segmentation of adipose tissue	CNN	The model was tested on 2 datasets; the accuracy produced 95.8% and 96.8% for computed tomography slice selection-CNN and fat pixel segmentation-CNN, respectively
Xu et al, 2017 [71]	Classification, segmentation of tissue	CNN	Outcome generates patterns that reveal biological insights that have been verified by pathologist
Hughes et al, 2016 [72]	Reactivity to biological macromolecules	CNN	The model captured molecules that would have been missed by standard reactivity screening experiments
Song et al, 2018 [73]	Segmentation of cervical cytoplasm	CNN	Experimental results achieved an accuracy of 94.50% for nucleus region detection and a precision of 0.9143(SD 0.0202) and a recall of 0.8726(SD 0.0008) for nucleus cell segmentation
Gurcan et al, 2001 [74]	Detection of microcalcifications	CNN	The results demonstrated optimization of cost surface whose characteristics are not known
Han et al, 2015 [30]	Membrane bioreactor permeability	RNN[b]	Simulation and experimental results demonstrate the reliability and effectiveness of the proposed intelligent detection system
Zhang et al, 2017 [75]	Sequence-specific correction for RNA	RNN	RNN-based bias correction method compares well with the state-of-the-art sequence-specific bias correction method
Ren et al, 2018 [76]	Prostate cancer differentiation	RNN	Their study demonstrates that prostate cancer patients with Gleason score of 4+3 have a higher risk of disease progression and recurrence compared with prostate cancer patients with Gleason score of 3+4
Wang et al, 2018 [77]	Detecting biomedical event trigger for protein and gene	RNN	F-score potentially reached about 80%, which is better than comparative experimental methods
Chen et al, 2018 [78]	Effective drug combination	DBN[c]	Predict effective drug combination from gene expression and pathway and ontology fingerprints properties for treating cancer
Beevi et al, 2017 [79]	Cell mitosis detection	DBN	The algorithm provides improved performance compared with other state-of-the-art techniques with average F-score of 84.29% for the MITOS[d] dataset and 75% for the clinical dataset from Regional Cancer Centre
Zhang et al, 2014 [80]	Identification of critical proteins	DBN	The results of comparison showed that DBN had higher reconstruction rate compared with baseline methods and more proteins of critical value to yeast cell cycle process were identified
Jiang et al, 2016 [81]	Huntington disease identification	DBM[e]	Results demonstrate that the model can detect important information for differential analysis of time series gene expression datasets
Ghasemi et al, 2017[82]	Biological activity prediction	DBN	The output of the model demonstrated significant superiority to traditional neural network with random parameters
Eraslan et al, 2019 [83]	Single cell RNA-seq denoising	AE[f]	Outperforms existing methods for enhancing biological discovery and data imputation in terms of quality and speed
Guan et al, 2018 [84]	Gene function annotation	AE	The model can capture intermediate representations to partial corruption of input pattern and generate low-dimensional codes superior to conditional dimension reduction tools
Chen et al, 2018 [85]	Genomics functional characterization	AE	Retains sufficient biological information with regard to tumor subtypes and clinical prognostic significance. Provides high reproducibility on survival analysis and accurate prediction for cancer subtypes
Wang et al, 2018 [86]	Visualization of single cell RNA- seq	AE	Reconstructs the cell dynamics in preimplantation embryos and identifies several candidate marker genes associated with early embryo development

Reference	Task	Method	Remark
Maggio et al, 2018 [87]	Prognostic profiling for survival prediction	AE	The embedding technique can be used to better stratify patients' survival
Hu et al, 2018 [88]	Prediction of drug-likeness	AE	The classification accuracy of drug-like/nondrug–like models are 91.04% on WDI-ACD[g] databases and 91.20% on MDDR-ZINC[h] database

[a]CNN: convolutional neural network.

[b]RNN: recurrent neural network.

[c]DBN: deep belief network.

[d]MITOS: mitosis detection in breast cancer histological images.

[e]DBM: deep Boltzmann machine.

[f]AE: autoencoder.

[g]WDI-ACD: world drug index-available chemicals directory.

[h]MDDR-ZINC: MDL drug data report-zinc compound.

Health Record and Report Management

EHR stores patient's data for improving health care and providing personalized treatment and historical account. This information can be in the form of radiological images, diagnosis, temporal event extraction, doctor health review, disease classification, demographic details, prescription, laboratory tests, and results. Over the years, these records have increased proportionally making it difficult and challenging for health workers and medical professionals to handle. Until the last decade, most approaches were based on statistical techniques and few attempts to use ML. Currently, those methods have become infeasible because of the large and increasing amount of these data. DL becomes imperative to be able to make meaning from these data. Table 2 presents some literature of DL implementation in EHR.

CNN techniques have been adapted to solve some task related to EHR to achieve better health management system. Feature engineering remains a major bottleneck when creating predictive systems for EHRs [89]. Word embedding from discharge notes combined with CNN was applied to disease code classification [90]. The model was based on a 1-layer CNN with a filter region size of 1 to 5 to increase comparability with traditional ML techniques. One study [91] proposed a CNN method for phenotyping from patients' EHRs. For the initial setup, every patient's record was represented as a temporal matrix with time on one dimension and event on the other dimension. A 4-layer CNN model was created for extracting phenotypes and performing prediction. The first layer comprised the temporal matrix. The second layer was a one-side convolution layer that could extract phenotypes from the first layer. The third layer was a max-pooling layer introducing sparsity on the detected phenotypes so that only those significant phenotypes would remain. The fourth layer was an FC softmax prediction layer.

Similarly, RNN methods have also been applied to solve problems relating to EHR to understand symptoms and achieve improved health care quality and personalized medication. A combination of bidirectional LSTM and CRF network has been implemented to recognize entities and extract relationship between entities in EHR [93]. To improve the model, multitask learning was included to handle hard parameter sharing, parameter regularization, and task relation learning. In addition, another LSTM and CRF to recognize clinical entities from Chinese EHR data was proposed [94]. Character embedding and segmentation information were used as features to be able to semantically understand diagnoses, tests, body parts, symptoms, and treatments. In [95], LSTM method was used for structured prediction in clinical text and in [96], RNN frameworks were explored and proved to be significantly better than CRF models. In [97], the LSTM implementation was for a single data structure that could be used for many predictions, rather than requiring custom, hand-created datasets for every new prediction. This approach represented the entire EHR in temporal order, which represents the event in a patient's timeline.

In addition, RBM and AE DL techniques have been constructed to handle challenging tasks found in EHRs to provide solutions or serve as a preprocessing step for another technique. A DBN framework was applied to predict the risk of osteoporosis from heterogeneous EHR for monitoring bone disease progression [104]. The framework is capable of pinpointing the underlying causes of the disease to assess the risk of a patient in developing a target disease and discriminating between patients suffering from the disease for the purpose of selecting risk factor of the disease. In [105], 2 novel modifications to DBN training was proposed to address the challenges and exploit properties that are peculiar, if not exclusive, to medical data. First, a general framework was examined for prior knowledge to regularize parameters in the topmost layers. Second, a scalable procedure was described for training a collection of NNs of different sizes but with partially shared architectures. AE model was developed for handling computational task and analysis of EHR [107-109]. In [109], the challenge of traditional supervised learning approach for inferring precise phenotypic patterns was addressed. Conventionally, an expert designates which pattern to look for (by specifying the learning task and the class labels) and where to look for them (by specifying the input variables). Although this is appropriate for individual tasks, this approach scales poorly and misses the patterns. Unsupervised feature learning with AE was able to handle these limitations by identifying patterns (or *features*) that collectively form a compact and expressive representation of the source data, with no need for expert input or labeled examples.

Table 2. Deep learning implementation in electronic health records and medical report management.

Reference	Task	Method	Remark
Wickramasinghe et al, 2017 [89]	Extract features from medical records	CNN[a]	It achieves superior accuracy compared with traditional techniques to detect meaningful clinical motifs and uncovers the underlying structure of the disease
Lin et al, 2017 [90]	Disease code classification	CNN	The method had a higher testing accuracy (mean AUC[b]=0.9696; mean F-score=0.9086) than traditional NLP[c]-based approaches (mean AUC range 0.8183-0.9571; mean F-score range 0.5050-0.8739)
Cheng et al, 2016 [91]	Risk prediction of chronic congestive heart failure	CNN	The model performance increases the prediction accuracy by 1.5% when 60% training data were used and 5.2% when it is 90% training data
Zeng et al, 2017 [92]	MobileDeepPill: Recognition of unconstrained pill image	CNN	DL[d]-based pill image recognition algorithm won the first price of the NIH[e] NLM[f] Pill Image Recognition Challenge
Li et al, 2018 [93]	Extraction of adverse drug events	RNN[g]	The DL model achieved a result of F-score=65.9%, which is higher than F-score=61.7% from the best system in the MADE[h]1.0 challenge
Zhang et al, 2018 [94]	Identify clinical named entity	RNN	CRF[i] and bidirectional LSTM[j]-CRF achieved a precision of 0.9203 and 0.9112, recall of 0.8709 and 0.8974, and F-score score of 0.8949 and 0.9043, respectively
Jagannatha et al, 2016 [95]	Prediction based on sequence labeling	RNN	Prediction model improved detection of the exact phrase for various medical entities
Jagannatha et al, 2016 [96]	Extraction of medical events	RNN	Cross-validated microaverage of precision, recall, and F-score for all medical tags for gated recurrent unit–documents are 0.812, 0.7938, and 0.8031, respectively, which are higher than other methods
Rajkomar et al, 2018 [97]	Representation of patients' record	RNN	Achieved high accuracy for tasks such as predicting in-hospital mortality, prolonged length of stay, and all of a patient's final discharge diagnoses
Hou et al, 2018 [98]	Extraction of drug-drug interaction	RNN	DL can efficiently aid in information extraction (drug-drug interaction) from text. The F-score ranged from 49% to 81%
Choi et al, 2015 [99]	Predicting clinical events	RNN	On the basis of separate blind test set evaluation, the model can perform differential diagnosis with up to 79% recall, which is significantly higher than several baselines
Choi et al, 2016 [100]	Detection of heart failure onset	RNN	When using an 18-month observation window, the AUC for the RNN model increased to 0.883 and was significantly higher than the 0.834 AUC for the best of the baseline methods
Volkova et al, 2017 [101]	Forecasting influenza-like illness	RNN	LSTM model outperformed previously used models in all metrics, for example, Pearson correlation (0.79), RMSE[k] (0.01), RMSPE[l] (29.52), and MAPE[m] (69.54)
Yadav et al, 2016 [102]	Patient data deidentification	RNN	The proposed approach achieved best performance, with 89.63, 90.73, 90.18 for recall, precision, and F-score, respectively
Hassanien et al, 2013[103]	Classification of diagnoses	RNN	Models outperformed several strong baselines, including a multilayer perceptron trained on hand-engineered features
Li et al, 2014 [104]	Identifying informative risk factors and predicting bone disease	DBN[n]	Proposed framework predicted the progression of osteoporosis from risk factors and provided information to improve the understanding of the disease
Che et al, 2015 [105]	Detection of characteristic patterns of physiology	DBN	The empirical efficacy of the technique was demonstrated on 2 real-world hospital datasets and the model was able to learn interpretable and clinically relevant features
Tran et al, 2015 [106]	Harness electronic health record with minimal human supervision	DBM[o]	The model achieved F-scores of 0.21 for moderate-risk and 0.36 for high-risk, which are significantly higher than those obtained by clinicians and competitive with the results obtained by support vector machine

Reference	Task	Method	Remark
Miotto et al, 2016 [107]	Predict future of patients	AE[p]	Results significantly outperformed those achieved using representations based on raw electronic health record data and alternative feature learning strategies
Lv et al, 2016 [108]	Clinical relation extraction	AE	The proposed model is validated on the dataset of i2b2 2010. The DL method for feature optimization showed great potential
Lasko et al, 2013 [109]	Inferring phenotypic patterns	AE	The model distinguished the uric acid signatures of gout and acute leukemia despite not being optimized for the task

[a]CNN: convolutional neural network.
[b]AUC: area under the curve.
[c]NLP: natural language processing.
[d]DL: deep learning.
[e]NIH: national institutes of health.
[f]NLM: national library of medicine.
[g]RNN: recurrent neural network.
[h]MADE: medication and adverse drug events.
[i]CRF: conditional random fields.
[j]LSTM: long short-term memory.
[k]RMSE: root mean square error.
[l]RMSPE: root mean square percentage error.
[m]MAPE: mean absolute percentage error.
[n]DBN: deep belief network.
[o]DBM: deep Boltzmann machine.
[p]AE: autoencoder.

Medical Image

The success of DL algorithms for image segmentation, localization, classification, and recognition task in recent years is timely with remarkable increase in medical image data. Analysis of these data has become an active field, partly as image data are easier for clinicians to interpret and they are relatively structured and labeled. There have been reported accuracies in some publications for detecting a range of anomalies such as malignant tumor, breast mass localization, recognition of pathology (organ parts), infectious diseases, and coronary artery stenosis classification. CNN and AE have been commonly implemented to solve challenging medical image problems. This is because the structure of these DL methods makes it possible to learn salient features from the data to create different levels of abstraction to achieve the required result. Table 3 lists some areas in medical image that have implemented DL solution. The performance exceeds ML techniques such as SVM and random forest classifiers.

The use of statistical pooling strategy was crafted into building CNN model in [110], with feature extraction at different convolutional layers and a multivariate classifier trained to predict which tumor contained occult invasive disease. Another pooling technique is stochastic pooling for alcoholism detection [111]. In [112], a CNN method that uses multiple patch sizes and multiple convolution kernel sizes was proposed to acquire multiscale information about each voxel to achieve accurate segmentation of tissue in brain images. Chest diseases are very serious health problems in the life of people. One study [113] presented CNN model for diagnosis of chest diseases. The designed model was trained and tested using chest x-ray images containing different diseases.

Moreover, [115] presented a modified DBN technique to minimize the computation load from 3D ultrasound data for the first trimester of pregnancy, which is an important parameter in prenatal screening. The technique converts the sagittal plane into a symmetry plane and axis searching problem. Feature extraction requires technical and task-specific approach such as neuroimaging, which contains features for diagnosing diseases. A DBM technique was proposed for a high-level hierarchical latent and shared feature representation from 3D patch neuroimaging modalities [116]. An integrated visual and textual multimodal image retrieval approach was created for cancer clinical practice and research with DBM method [117]. In addition, the DBM method can be used to extract volumetric representations from 3D brain image for classification of sensorimotor activities. The weights in the higher level of the architecture show spatial patterns that can identify specific tasks and the third layer represents distinct patterns or codes. In [118], a deep generative shape model–driven level set method was developed and evaluated to address automatic heart motion tracking to minimize radiation-induced cardiotoxicity. The proposed heart motion tracking method made use of MRI image sequences that characterize the statistical variations in heart shapes. This heart shape model was established by training a 3-layered DBM to characterize both local and global heart shape variations.

Table 3. Application of deep learning techniques in medical images.

Reference	Task	Method	Remark
Shi et al, 2018 [110]	Occult invasive disease prediction	CNN[a]	The performance result exceeded handcrafted computer vision technique that was designed with prior domain knowledge. It achieved operating characteristic curve of 0.70 and 95% CI
Wang et al, 2018 [111]	Alcohol detection	CNN	The method used multiple images in the experiment and achieved 96.88% sensitivity, specificity of 97.18%, and accuracy of 97.04%
Moeskops et al, 2016 [112]	Tissue segmentation	CNN	The result demonstrates accurate segmentation in all datasets and the robustness to different age and acquisition protocol
Abiyev et al, 2017 [113]	Chest disease detection	CNN	Demonstrate accurate classification of chest pathologies such as chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, and lung diseases in chest x-rays
Liu et al, 2016 [114]	Food image recognition for dietary assessment	CNN	These results outperformed all other reported work such as DeepFoodCam using UEC-256 and Food-101 dataset
Nie et al, 2017 [115]	Detection of standard sagittal plane in pregnancy	DBN[b]	The model provides knowledge to avoid unnecessary massive searching and corresponding huge computation load
Zhang et al, 2016 [116]	Benign and malignant breast tumors differentiation	DBM[c]	Results showed that the deep learning method achieved better classification performance with an accuracy of 93.4%, a sensitivity of 88.6%, a specificity of 97.1%, and an area under the receiver operating characteristic curve of 0.947
Cao et al, 2014 [117]	Cancer clinical practice and research	DBM	Experimental results with large volume of real-world medical images showed that multimodal approach is a promising solution for the next generation medical image indexing and retrieval system
Wu et al, 2018 [118]	Tracking motion of the heart	DBM	Heart shape model that characterizes the statistical variations in heart shapes present in a training dataset for tracking motion of the heart
Jang et al, 2017 [119]	Four sensorimotor classification	DBM	Identified task-specific (left hand clenching, right hand clenching, auditory attention, and visual stimulus) features and classification of functional/structural magnetic resonance imaging volumes
Suk et al, 2014 [120]	Alzheimer disease identification	DBM	Achieved maximum accuracy of 95.35%, outperforming other computing methods
Khatami et al, 2016 [121]	Extract high-level features from medical images	DBN	Experimental results show that the proposed model improves about 0.07% performance compared with other models
Zhang et al, 2019 [122]	Discovering hierarchical common brain networks	DBN	Three hierarchical layers with hundreds of common and consistent brain networks across individual brains were successfully constructed
Hu et al, 2019 [35]	Cancer diagnosis	AE[d]	Diagnosed malignant mesothelioma, a rare but aggressive cancer because of its composite epithelial/mesenchymal pattern
Uzunova et al, 2018 [123]	Pathology detection	AE	Experiments on 2-dimensional and 3-dimensional datasets show that the approach is suitable for detection of pathologies and deliver reasonable dice coefficient result
Lee et al, 2018 [124]	Benign and malignant tumor classification	AE	The results show that when deep learning algorithm is applied on sonograms after intensity inhomogeneity correction, there is a significant increase of the tumor classification accuracy
Seebock et al, 2018 [125]	Age-related macular degeneration classification	AE	Used markers to classify early and late age-related macular degeneration cases. The model yields an accuracy of 81.40%
Wang et al, 2019 [126]	Spine disease diagnosis	AE	Achieved higher localization accuracy, low model complexity, and without the need for any assumptions about visual field in computed tomography scans
Malek et al, 2017 [127]	Image description for visually impaired	AE	Fusing a set of AE-learned features gave higher classification rates with regard to using the features individually
Zhang et al, 2016 [128]	Histopathological images analysis	AE	The method effectively combined the strength of multiple features adaptively as inputs and achieves 91.67% classification accuracy

Reference	Task	Method	Remark
Xia et al, 2016 [129]	Human attention process	AE	Experimental results on several benchmark datasets show that in accordance with different inputs, the network can learn distinct basic features for saliency modeling in its encoding layer
Mano et al, 2018 [130]	Chronic back pain detection	AE	Experimental results from patients in the United Kingdom and Japan (41 patients, 56 controls), achieved accuracy of 63%, with 68% in cross-validation of all data

[a]CNN: convolutional neural network.

[b]DBN: deep belief network.

[c]DBM: deep Boltzmann machine.

[d]AE: autoencoder.

Furthermore, preprocessing with AE can handle noise in images for tumor detection and intensity inhomogeneity correction before performing classification. A combination of multiple techniques can also be constructed, for example, [119] considered point-wise gated BM and RBM to identify tumor from image data. In [123], it was demonstrated that conditional variational AE can learn the reconstruction and encoding distribution of different variabilities of 2D and 3D images to learn the appearance of pathological structures. Consequently, preprocessing of image can be done to highlight features that are fed into DL architecture to increase distinguishing tumor accuracy [124]. Most applications of DL methods for diagnosis and classification of diseases require that the images are marked for training; however, because of the limitation of marked entities and training examples, supervised training does not scale well. Furthermore, a multiscale deep denoising AE was constructed without the constraint of prior definition for the identification of anomalies occurring frequently in retinal optical coherence tomography image data [125]. Qualitative analysis of these markers shows predictive value in the task of detecting healthy, early, and late age-related muscular degeneration. A combination of contextual features from deep stacked sparse AE (SSAE) and structured regression forest for vertebrae localization and identification was created to overcome handcrafted and low-level features of spine structure [126]. The method employs SSAE to learn deep contextual features by building larger-range input samples to improve the contextual discrimination ability.

Physiological Signals and Sensors

Advancement in sensing technology has made it possible to acquire and analyze signals from patients for monitoring mental state, heart condition, and disease diagnosis. An important process for achieving accurate and high-performance model depends on feature extraction and feature selection. DL algorithms have recorded successful achievement in modeling physiological signals and sensor data with better accuracy compared with traditional ML techniques. Table 4 presents some implementations of DL methods described in literatures using sensors and physiological signals for health care management. One significant characteristic of this kind of medical data is that they are sequential and time-dependent. Therefore, a technique for creating a model should not only take into consideration the shape of the data (spatial features), but also the time factor (temporal features). Skin conductance sensor, microphone for speech sequence, blood volume pulsation, electroencephalogram (EEG), photoplethysmogram (*PPG*), ECG, and so on are some examples of data in this category. RNN and its variants have been predominantly applied in this domain as it is able to model the data in the context of time and sequence to reveal hidden features. DBN and AE have also been implemented to solve problems with this kind of data.

Many variations of CNN techniques have been created to achieve enhanced performance and there are some optimization techniques that have been applied to achieve better accuracy. In [131], a technique based on CNN called fast discriminative complex-valued CNN for automatic sleep stage classification was developed. The method can capture the sleep information hidden inside EEG signals and automatically extract features from the signal. The constructed CNN method eliminated the need for deep digital signal processing skills, which usually serve as preprocessing phase for most signal operations. The time series plots generated directly from accelerometer and gyroscope signals were employed for classification of human activity and exercise detection [132]. The classification task was achieved with CNN, where the generated signals were formatted into image dimension. Estimation of brain activation with CNN was addressed in [133], where the temporal region was modified with unscented Kalman filter and a corresponding unscented smoother to observe inference relation of task-specific brain network. The CNN model parameters were estimated using expectation-maximization algorithm to exploit the partial linearity of the model. Moreover, the initial challenge encountered in building a model is data preprocessing and setup. Often data are irregular, inconsistent, and sometimes contain irrelevant information. Therefore, the preprocessing bottleneck for the creation of 2 mental state classification models for drivers from EEG signals was handled with CNN and deep residual learning [38]. The model contains 8 layers: the input layer, 3 convolutional layers, a pooling layer, a local response normalization layer, an FC layer, and the output layer.

Table 4. Deep learning technique for sensors and physiological signal task.

Reference	Task	Method	Remark
Zeng et al, 2018 [38]	Predict mental states of drivers	CNN[a]	Predicted the mental states of drivers from electroencephalography (EEG) signals using 2 mental state classification models called EEG-Conv and EEG-Conv-R
Zhang et al, 2017 [131]	Sleep stage classification	CNN	The total accuracy and kappa coefficient of the proposed method are 92% and 0.84, respectively
Veiga et al, 2017 [132]	Human activity classification	CNN	The exercises could be recognized with 95.89% accuracy
Lenz et al, 2011 [133]	Interactions in human brain	CNN	Result showed regions of human brain affected by interactions and activities
Murad et al, 2017 [32]	Human activity recognition	RNN[b]	Experimental results showed that the proposed method outperforms methods employing conventional machine learning algorithms, such as support vector machine and k-nearest neighbors
Liu et al, 2016 [134]	Predicting driving fatigue	RNN	Identified brain dynamics in predicting car driving fatigue. The model was evaluated using the generalized cross-subject approach
Yu et al, 2015 [135]	Human action classification	RNN	Real-time human action classification for recording and regenerating both action sequences and action classification tasks from continuous signal
Vakulenko et al, 2017[136]	Human body motion analysis	RNN	The method generated missing information from human body motions from sparse control marker settings
Mo L et al, 2016; Ordóñez et al, 2016 [137,138]	Human physical activity recognition	RNN	The model can recognize 12 types of activities and the accuracy rate was 81.8%. CNN was applied for feature extraction and long short-term memory model for the human physical activity recognition
Mathews et al, 2018 [36]	Cardiac arrhythmias diagnosis	DBM[c]	Single-lead electrocardiogram model detected cardiac abnormalities which was comparable with human expert
Chu et al, 2018 [40]	Recovery motor imagery	DBN[d]	Recognize and restructure the incomplete motor imagery in electroencephalogram signals for recovery motor imagery-based treatment
Turner et al, 2014 [139]	Seizure detection	DBN	High resolution biosensing multichannel model achieved personalized health monitoring
Chao et al, 2018 [140]	Emotion recognition	DBM	The results showed that the proposed framework outperforms other machine learning classifiers
Jindal, 2016 [141]	Heart rate monitoring	DBM	Technique was able to predict heart rate with a 10-fold cross-validation error margin of 4.88%
Hassan et al, 2018 [142]	Human activity recognition	DBN	The proposed approach outperformed traditional expression recognition approaches such as typical multiclass support vector machine and artificial neural network
Ruiz-Rodríguez et al, 2014 [141]	Blood pressure monitoring	DBM	Continuous noninvasive blood pressure monitoring system with performance higher than benchmark methods
Yuan et al, 2019 [144]	Epileptic seizures detection	AE[e]	Experimental results showed that the proposed model was able to achieve higher average accuracy and F-score of 94.37% and 85.34%, respectively
Jirayucharoensak et al, 2014 [145]	Detection of emotion	AE	The model provided better performance compared with support vector machine and Naive Bayes classifiers
Jokanović 2017 [146]	Human fall detection	AE	Experimental data were used to demonstrate the superiority of the model over principal component analysis method
Xia et al, 2018 [147]	Cardiac arrhythmia classification	AE	The results from the model (99.8% accuracy) showed that the classification performance of the proposed approach outperforms most of the state-of-the-art methods

[a]CNN: convolutional neural network.

[b]RNN: recurrent neural network.

[c]DBM: deep Boltzmann machine.

[d]DBN: deep belief network.

[e]AE: autoencoder.

Similarly, RNN methods have been effective in mining discriminative features from raw input sequences acquired from body-worn sensors. The use of deep RNN for building recognition models for human activities was proposed in [32]. It was capable of capturing long-range dependencies in variable-length input sequences. Moreover, the use of fuzzy logic with RNN to create a recurrent fuzzy NN increases adaptability and the bottleneck of regression problem to handle driving fatigue for preventing road accidents [134]. In addition, classification of multiple types of motion when observing a human action can be difficult because of the complex nature of the timescale signal. Therefore, [135] proposed supervised multiple timescale RNN architecture for handling the issue of action classification. To overcome the difficulty of setting the initial states, a group of slow context nodes, known as *classification nodes*, was created. The supervised model provides both prediction and classification outputs simultaneously. In [136], continuous-time RNN networks were considered as dynamic models for the simulation of human body motion. These networks comprise a few centers and many satellites connected to them. The centers evolve in time as periodical oscillators with different frequencies.

Furthermore, [40] developed a decoding scheme from a combination of Lomb-Scargle periodogram and DBN to recognize incomplete EEG signal data to solve the problem of motor imagery recovery for performing classification tasks such as heart rate variability. The use of a variety of representations and DBM algorithms was explored for seizure detection in high resolution, multichannel EEG data [138]. In addition, a DL framework based on improved DBN with glia chains (DBN-GCs) for handling emotion recognition task was constructed [140]. In the framework, DBN-GCs are employed for extracting intermediate representations of EEG raw features from multiple domains separately, as well as for mining interchannel correlation information by glia chains. The higher-level features describe time domain characteristics and frequency domain characteristics. The time-frequency characteristics are fused by a discriminative RBM to implement emotion recognition task. Owing to the increasing monitoring of heart rate through mobile phones and wearable devices, [141] presented a novel technique for accurately determining heart rate during intensive motion by classifying PPG signals obtained from mobile phones or wearable devices integrated with motion data obtained from accelerometer sensors.

Moreover, AE models have also been constructed to solve health and biomedical challenges using signals and sensors. For example, AE-based multiview learning was implemented to monitor and analyze multichannel EEG signals of epileptic patients to prevent complications caused by epileptic seizures [142]. The implemented approach was an end-to-end model that was able to jointly learn multiview features from both unsupervised multichannel EEG reconstruction and supervised seizure detection via spectrogram representation. In [145], the utilization of stacked hierarchical AE learning approach was proposed for automatic emotion recognition with nonstationary EEG signals. To alleviate overfitting problem, principal component analysis was applied to extract the most important components of the initial input. In addition, covariate shift adaptation of the principal components was implemented to minimize the nonstationary effect of EEG signal. In another similar implementation, a stacked AE was proposed to detect human fall [146]. The proposed approach automatically captures the intricate properties of signal from radar. To minimize false alarms in human fall detection, information from both the time-frequency and range domains was fused together.

Challenges in Health Care for Deep Learning Applications

In spite of the all the impressive achievements and capabilities of DL discussed in the previous section, the technique is still in its infancy in biomedical and bioengineering applications. There are significant challenges that need to be resolved for DL to be able to handle the inherent medical and health care challenges. This section highlights some of the challenges confronting the implementation of DL methods.

Medical Data Representation and Transformation

DL algorithms can make the most effective observations and predictions with the appropriate type and quantity of data. Currently, real-world medical data are in unstructured formats such as sequences (time series, audio and video signals, DNA, and so on), trees (XML documents, parse trees, RNA, etc), text data (symptoms description, tumor description, medical records), or combinations of any of these formats [148]. Unfortunately, the core of DL technique can only process numeric input data as eventually it is broken down to strings of zeros and ones for computing system. Some qualitative data are not easily converted into a usable format and processing can sometimes become complicated. Humans can easily process and make meaning of these data and when there is a simultaneous change, for example, in intensity and quantity, it can easily be understood and adjustment can be made with regard to the changes; an example is temperature and light. The representation of similar processes and conditions in DL requires a lot of encoding and thoughtful mathematical expressions in DR and transformation. Cross-domain feature learning algorithm based on stacked denoising AE has been considered for effective feature representation to describe data with multimodal property (eg, signals, image, video, and audio) [149]. A DL architecture that is capable of integrating multiple types of data concurrently is needed to handle some real-world situations.

Figures 15 and 16 show the differences between currently available implementation of DL methods and the expected real-world implementation, respectively. In Figure 15, different medical data have corresponding DR formats which fit a particular DL architecture. The design is expected to respond to one type of input or DR and produce a response for the expected input. However, Figure 16 is a smart DL that mimics the simultaneous process of the human body where it is capable of handling multiple inputs simultaneously to decide a response.

Figure 15. Current working technique for application of deep learning with biomedical data.

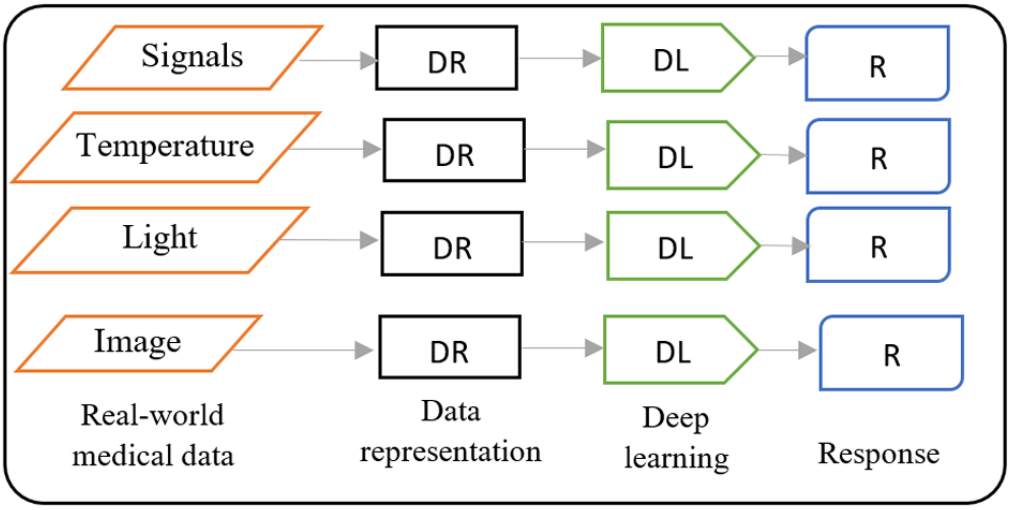
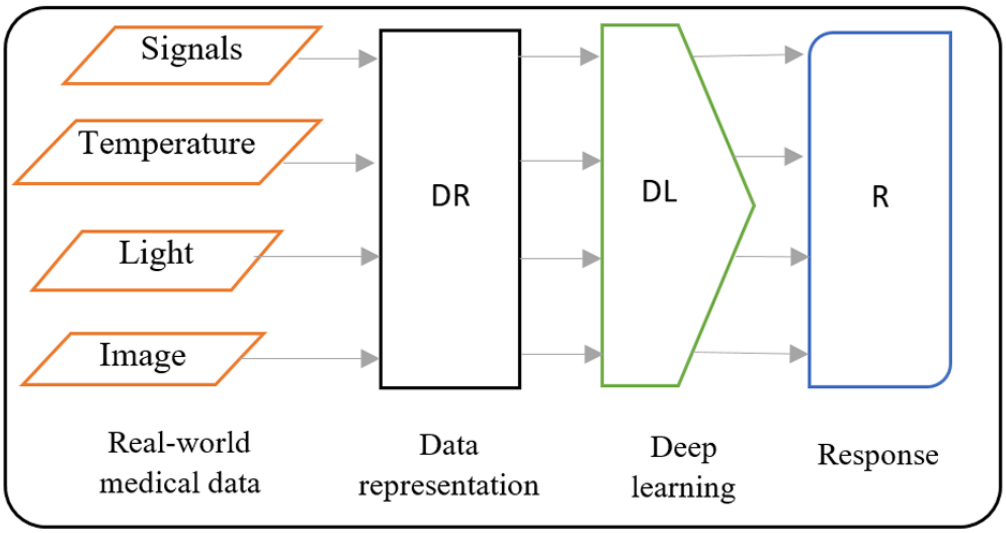


Figure 16. Expected required technique in biomedical application of deep learning.



Consider, for example, a multiple DR DL model for monitoring a baby's health condition. When the model observes an abnormal behavioral pattern of the baby, the temperature and the physiological signals from the baby are analyzed and considering the input of light intensity and the temperature of the environment and the baby, the model produces a response for the baby's health status.

Handling Biomedical Data Stream

Another challenge with DL is dealing with fast moving and streaming data [150]. There is a rapid change in the health care industry with huge volume of health care data emanating at a rapid rate. The benefit of this is that the medical practitioners can leverage on these with the support of DL model to diagnose and deliver health care services for different pathological conditions. These data can be found in real-time biomedical signals from many sources, including blood glucose monitoring, brain activity, blood pressure and oxygen saturation level, biomedical imaging from ultrasound, electrography, MRI, in thousands of terabytes for insight into medical conditions. Unstructured data format of useful patient records in the form of clinical text contains useful patterns and genomic data describing relationship between various genetic markers, disease conditions, and mutations. Physiological sensing data from ECG and EEG are important signals that are acquired from different parts of the body. It is important for DL to be able to make meaning of large volumes of continuous input data that change with time and also take into consideration when previous data become obsolete. In [151], continuous activity learning framework for streaming videos by intricately tying together deep hybrid feature models and active learning was proposed. In another architecture, a streaming hardware accelerator was proposed for incremental feature learning with denoising AE [152,153]. Although, some DL architecture variants have tried to proffer techniques for working around this situation, there are unresolved challenges regarding effective analyses of fast moving, large-scale streaming data in terms of memory consumption, feature selection, missing data, and computational complexity.

Figure 17. Deep learning biomedical data streaming architecture, challenges, and applications.

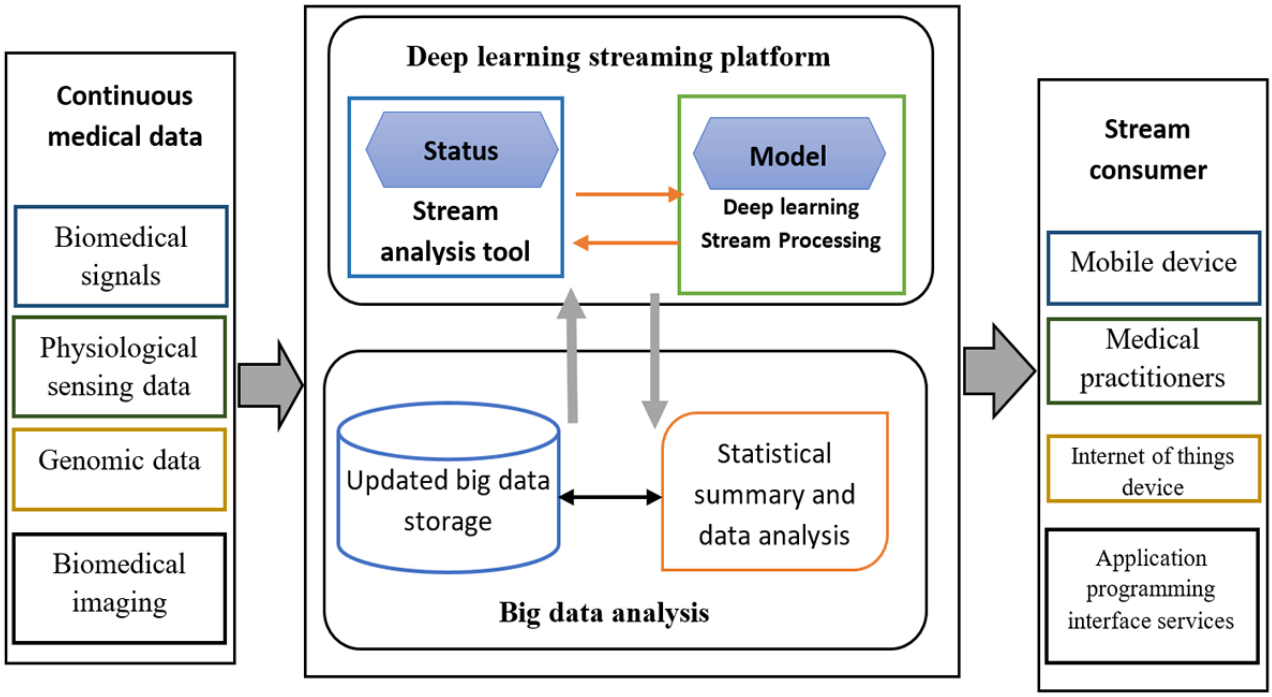


Figure 17 describes the summary between generated continuous medical data (CMD) and the usage of the data by DL. The challenge for DL is to make use of CMD, rather than only performing classification and prediction on the new data. The figure describes the use of updated database storage to keep generated data, which serves as a buffer to hold data until it is used to update the model in the DL stream processing.

The importance of this architecture is that it provides the platform for many consumer streams such as mobile device, medical practitioners, and third-party application programming interface (API) services to utilize the data. However, in the future, there may be variants of DL methods to handle real-time CMD, which may eliminate or reduce the function of big data analysis module.

Analyzing Medical Big Data

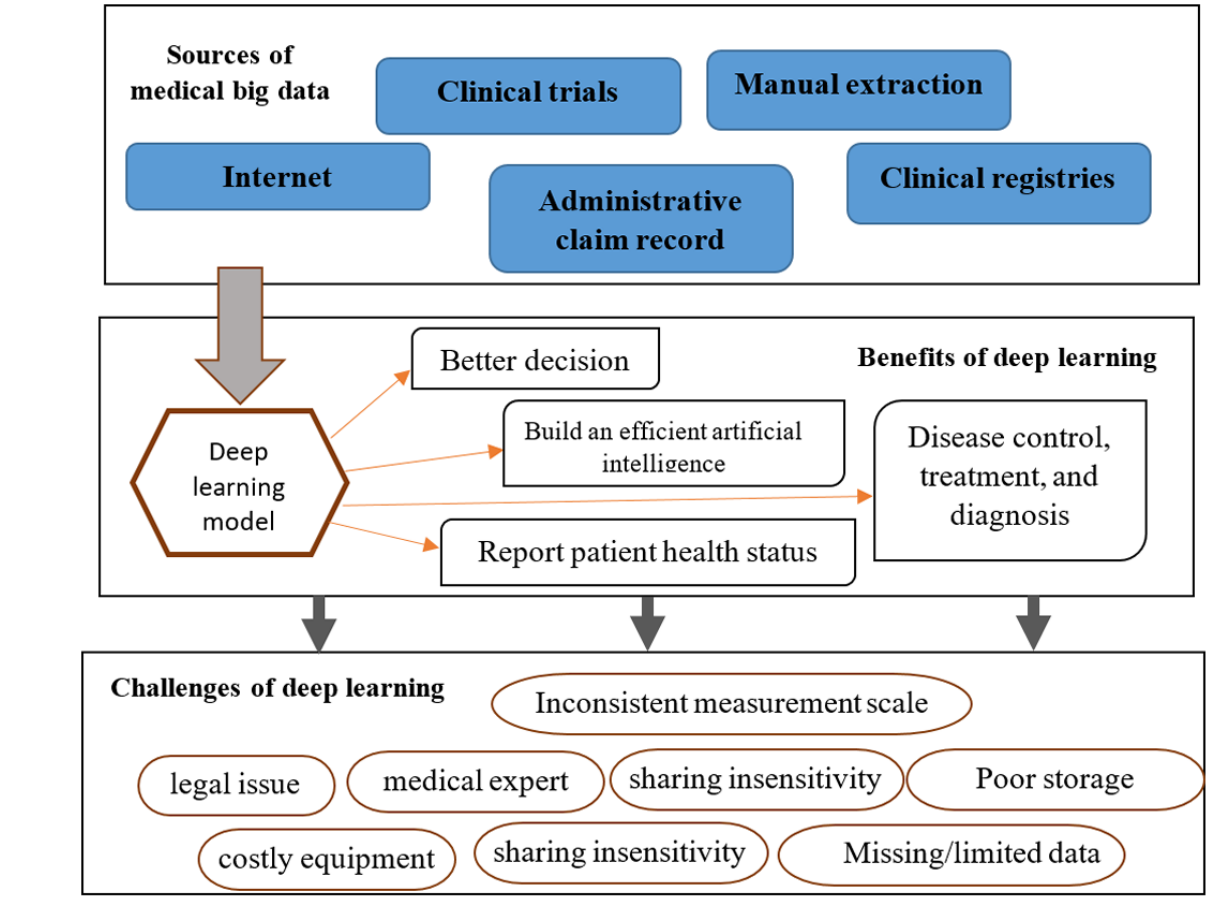
Large quantity of data is responsible for highly accurate results recorded in DL. During the learning phase, features in the data are used to build or create parameters in the neurons to achieve prediction. It is important for the data to be large. In addition, it is also important for the data to contain important and required features for the training. Some medical domains that want to take advantage of DL are restricted because of the difficulty in generating or acquiring data and sometimes labeling data requires domain experts, who are not readily available. In [154], a technique was presented for dealing with fine-grained image classification scarcity. Furthermore, [155,156] presented an approach to deal with useful features in data for DL. The question is how much information is available in large data. Tuning parameters of neurons is achieved largely through validation computation procedures and the choice of DL structure. Medical big data (MBD) analysis has numerous advantages, including disease control, treatment, and diagnosis. Some MBD can be obtained from various sources, for example, medical imaging devices, the internet, biometric data, large clinical trials, biomarker data, clinical registries, and administrative claim record [157]. DL provides the tool to make smart and accurate analysis of data in this field, which will assist experts take better decision, report patient health status, and build an efficient AI. However, there are some challenges confronting DL in this domain. The major challenge is the difficulty in acquiring MBD because of danger of data misuse and lack of data sharing insensitivity which could compromise privacy of patients, legal issue, costly equipment and medical expert involved [158].

Another issue is the method of data collection which is done through application forms and protocols which could be hectic. The data are sometimes relatively small compared with data from other environments (eg, social media) and they are generated from nonreplicable situations or not readily common conditions. There are other inherent challenges encountered in MBD. Apart from missing data, there is the issue of errors in encoding medical record data during storage and difference in measurement equipment and measurement scale. In some areas, data are not available or are not enough because of the lack of knowledge on the importance of big data and data analysis. Synthetic data are sometimes constructed and integrated with acquired real data to achieve a large data size and maintain a balance of variables, nevertheless, how much trust should be given to synthetic data. In the aspect of analysis, there are different types of patient characteristics which can result in differences in physiological signals such as weight and the time of treatment, which may be an additional dimension [159]. These issues need to be resolved to provide an integrated health care system that will improve service delivery and reduce dependence on experts. DL methods require reliable and large data for successful results. Although it is relatively easy to acquire data from other sources, such as Web or user internet pattern, online customer reviews, electronic devices, and atmospheric conditions (eg, temperature and wind speed), it is

often difficult to acquire data from human subjects because of the inherent challenges such as maintaining fixed position sometimes over a long period of time, static charges interference during data acquisition from the subject, difference in metabolic conditions, and negative side effects and reaction by the subject. A summary of the source of MBD, benefits of DL, and challenges of MBD is presented in Figure 18. It is necessary to have either single or multiple repositories to manage MBD; however, there are challenges such as poor network connection, incomplete or inconsistent data, and unavailable computing resources that will mitigate against achieving the desired analysis of the data.

An example is inconsistency in the measurement of date of birth. In one location the order is day, month, and year, but in another location the arrangement can be month, day, and year. In addition, there could be inconsistent units of measurement such as mmol/dl and mg/l for blood glucose measurement and ounce and kilogram for weight measurement. Although the issue of inconsistency in measurement can be managed, security and privacy still make it difficult to acquire data for analysis. Some medical institutions do not support the use of data from patients for studies or the data are not readily available. In some cases, it is expensive because of the high monetary cost needed to get this information from these institutions.

Figure 18. Overview of external challenges in acquiring and analyzing medical big data.



Hardware Requirements for Medical Big Data

DL solution requires large training data to function effectively. Usually real-world medical data are very large and are constantly increasing. To implement tasks and create models, the computing machine needs to be equipped with sufficient processing power. To handle such requirements, data scientists and engineers developed multicore high performing GPUs and similar processing units as the regular central processing unit is impractical to handle large-scale DL tasks [160]. These GPUs are expensive, consume lot of power, and are not readily available for common use or in medical institutes and hospitals where the data are captured and generated. However, some companies, such as Wolfram Mathematica and Nervana Systems, have taken up the project to provide cloud-based services that allow researchers to speed up the training process [42]. The challenge is that industry level DL systems use high-end data centers, which are not available in medical institutions, whereas, deployment after training is done on smart devices, such as laptops, smart wearable devices, embedded computers, and other mobile devices, which have small and inefficient processing units. The larger the DL architecture, the bigger the computing requirement needed to accomplish training. Deploying DL solution to the real world thus becomes a costly and processor consuming situation. mHealth big data are pointless without suitable DL analytic methods to extract meaningful information and hidden patterns from data. One study [161] presents a tutorial on DL in mobile big data analytics and discusses a scalable distributed DL framework over Apache Spark. The framework is executed as an iterative MapReduce computing on many Spark workers. A partial deep model is learned by each Spark worker on a partition of the overall

mobile, and a master deep model is then built by averaging the parameters of all partial models. Moreover, there are current hardware designs intended to implement artificial neurons in a chip, such as Intel Curie, NuPIC, SpiNNaker, and IBM TrueNorth [42]. There are couple of notable software packages that provide implementation of DNN and have API for Python, Java, C++, and MATLAB, for example, TensorFlow by Google, neon by Nervana Systems, and Caffe by Berkeley Center.

Future Trends for Deep Learning

The current achievement in DL will open up more research areas and improvements on existing models. This section describes possible directions for research and development with focus on health care and applications of physiological signals.

Complexity of Computation

There is increasing application of DL in the medical field, especially in the area of physiological signals such as ECG, EEG, electromyography (EMG), and so on. ECG measures the bioelectrical activity of the heart, EEG monitors the bioelectrical activities of the brain, and EMG observes the working condition of the muscles and nerves of the body. The success recorded in this area will bring to the surface more application and implementation variation techniques. The purpose of automated analysis of these signals is for implementation in clinical devices as a practical medical diagnostic tool to improve the efficiency of treatment and continuous health monitoring. To achieve this feat, subsequent studies need to enhance the complexity of classifier algorithm to improve computational efficiency and complexity. For example, one study showed that the memory and complexity of DBN model is higher than other algorithms such as SVM, logistic regression, and K-nearest neighbor (KNN) [139,162]. However, DBN provides high accuracy over the other algorithms. Therefore, improvement is needed to enhance the DL algorithm for practical use.

Multitasking Deep Learning

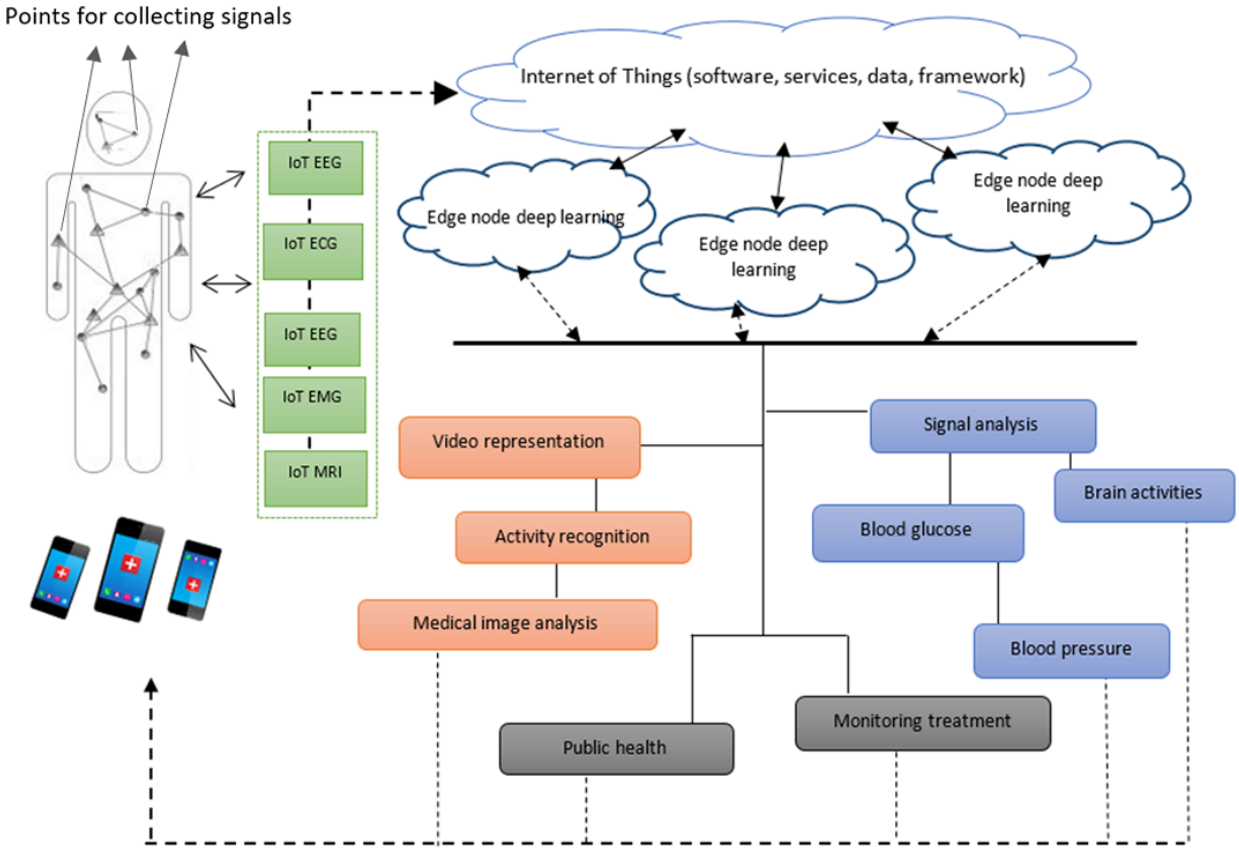
Currently, the rise in the use of wearable devices in recent years means that multiple physiological signals can be captured simultaneously and continuously. The classification and analysis of these signals may require different DL methods for different tasks. Future studies should consider a single generalized DL method that could satisfy multiple classifications. The purpose of this approach will conserve time and effort that would otherwise be needed to create a specific method for each classification. One study [163] considered a complex scenario of learning over multivariate and relational time series with missing data, where relations are modeled by a graph. They not only predicted future values for the time series, but also to fill in missing values. Another future consideration is to design DL methods that combine multiple physiological signals for classification. Multiple signals from wearable devices provide the possibility to integrate these signals to have a unified model. Constructive application of these signals will definitely increase accuracy and will also serve multiple purposes, where when one of the signals is not available, the system can still function with available signal input. Furthermore, a combinational adaptive DL model that is capable of handling multiple physiological signals for different classification tasks (multitasking) will minimize dependence on a single model and open the opportunity for a different approach to the application of DL.

Medical Internet of Things and Application

Internet of Things (IoT) and big data are responsible for the creation of smarter environment. One study [164] describes a smart environment as a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network and smart mobility. IoT devices are on the increase; an estimated 50 billion devices will be connected to the internet by 2020 [165]. This will bring about explosion in the size of data that will be generated. The exponential growth of data from connected devices, such as wireless body sensor, smart meters, and so on, makes DL the desired tool to make meaning from these data. A cloud-based DL becomes a challenge because of connection bottlenecks and overall reduction in the quality of service owing to latency issues [166]. Edge computing is proposed to move computing service from centralized cloud servers to edge nodes closer to the end users [167]. Some research has been conducted in this emerging field, such as seizure prediction in controlling epilepsy in medically refractory patients with EEG and electrocorticography signals via IoT [168]. The rapid proliferation of mobile phone and wearable devices has contributed to the evolution of IoT-enabled technology from usual single center–based system to more personalized health care systems. mHealth uses the wireless connection in IoT and mobile technology from mobile industry to create a connection between patients and health care professionals to make patients become advocates of their own health and promote communication between the professional and patients. mHealth framework in IoT has been used to create a voice pathology detection system using DL [169]. In the system, voices are captured using smart mobile devices. Voice signals are processed before being fed to a CNN. One study [170] presented a DL and mHealth technologies for improving tuberculosis diagnosis in Peru. Figure 19 shows the structure of edge computing technology where all the data captured from IoT devices are stored in the cloud. The edge nodes use specific data from the cloud required by client devices to perform DL analysis and computation. Edge computing adds 2 major enhancements to cloud computing by processing large volumes of data before transferring it to the cloud and enabling computing ability in the edge node, which optimizes the resources in the cloud. An example is DL-based food recognition system for dietary assessment on an edge computing service infrastructure [171]. Further research can be done to improve this area for efficient implementation of DL, for example, a distributed, layer-centered DL architecture that supports edge node operation of cloud resources. DL techniques maximize the number of tasks in computing environment owing to limited service capability, network performance, and scalability.

Performance evaluation and measurement of DL on edge computing is also another area to consider. Moreover, in the future there will be distributed and integrated variants of DL methods for edge computing because of the increase in IoT devices and technology.

Figure 19. Deep learning service for medical internet of things (IoT) with edge computing and mobile apps for continuous health care monitoring using magnetic resonance images (MRI) and signals such as electrocardiogram (ECG), electroencephalogram (EEG), electromyography (EMG).



Semisupervised Learning for Biomedical Data

Another key area of interest would be to explore the situation of both labeled and unlabeled data which occur in many biological domains such as proteins and DNA. The objective of DL in such cases is to integrate semisupervised training techniques toward achieving the criteria for good DR learning. For further studies to enable DL understand the patterns and DRs in such situations (unlabeled/unsupervised data), one approach would be to consider the existing labeled/supervised data to tune learned pattern and representation to achieve the optimal modeling for the data. Another approach is to combine DL and active learning [151]. Variants of semisupervised learning in data mining and active learning methods such as adaptive learning and transfer learning can be exploited toward obtaining improved DRs. Input from crowdsourcing or human experts can be used to obtain labels for some data samples which can then be used to better tune and improve the learned DRs [150]. Furthermore, hybrid DL has been constructed for feature extraction, classification, and verification of faces [172]. Application of DL in physiological signals for health care monitoring and analysis is still in its infancy and often suffers from incomplete or unavailable labeled data. More study needs to be done in this area and the implementation of hybrid techniques and variants of semisupervised learning to overcome the challenges of unlabeled data.

Replacement of Biomedical Research Methods by Deep Learning

There are more areas to implement DL to improve services, operations, devices, and software for health and medical fields. One study [173] proposed a clinical validation technique for improving grading of data collected by crowdsourcing for diabetic retinopathy, a leading cause of vision loss because of diabetes mellitus. A logistic regression method was implemented with 50% dataset for training that have normal and abnormal classification labels. Test and validation was performed on 50% dataset. The result achieved 90% sensitivity. However, the operation requires human decision which is prone to error and bias. CNN can be applied to 50% of labeled images to learn the features through series of convolution and pulling layers to predict 50% test set. The sensitivity is expected to be more than 90% as CNN abstraction of features at different layers will improve the sensitivity results. Furthermore, analysis of fall of individuals with dementia from continuous video monitoring was performed for early detection and prevention [174]. Analysis was carried out using a 4-point Hopkins Falls Grading Scale. A suitable DL method will be a hybrid technique of both RNN and CNN (recurrent CNN), which is able to approximate a function from a series of video frames in the continuous video sequence to systematically determine patient's condition: prefall, fall, and postfall. This can be implemented to trigger alarm for medical attention for the patient. In [175], mainstream wearable device was presented in health monitoring to support consumers in making purchasing decision. The analysis method

implemented may become ineffective as the data grows larger, but a DNN will remain effective despite the size of the data and the performance will not decline. Another area of application is in cardiac auscultation that can provide information about cardiovascular hemodynamics and diseases with simple diagnostic algorithm [176]. LSTM technique will effectively map the sequence of sounds from the device capturing the sound to distinguish between normal and pathologic heart sounds. LSTM is capable of understanding the pattern from the sound data because of the gate and memory circuit which is an integral part of the DL algorithm. One study [177] presented a qualitative and quantitative tablet-based software application for assessing bodily symptoms for both clinical and research purposes. The implementation can be achieved with multilayer stack AE between patients and doctors. The architecture of AE is able to encode and decode the input from patients to the expected output for the treating doctor and vice versa. The multilayer concept will handle the test-retest reliability. Furthermore, a construction of a priori analysis was employed to describe the essential qualities of participant's experience [178]. This included delineation of common and novel themes relating to informed consent, with a self-administered, mobile phone-based electronic consent (eConsent) process over a 6-month period within the Parkinson mPower app. This challenge can effectively be resolved with structured DBN architecture. Data collected for the specified period can be used to train a DBN model that resides in a cloud and the mPower app can communicate with the model to get required result. DBN layer-wise training technique ensures that features in the input data are taken into consideration during the creation of the model. This architecture removes processing responsibility from the mobile app (this makes the device light), provides the possibility for extension of the model, allows multiple users to take advantage of the system, and allows for central update when necessary. In [179], user needs for mobile behavioral-sensing technology was presented for depression management using thematic analysis with an inductive approach. The research was conducted by interviewing 9 clinicians and 12 students with depression, recruited from a counseling center. The interview duration was between 40 and 50 min and there was audio recording and transcription. The success recorded was because of small data size. However, it will become challenging with large size of data and human limitation will affect performance. Therefore, hybrid DL technique will provide a better performance, with the use of LSTM to model the recorded audio and DNN for the structured content. The use of hybrid technique will make better meaning from the data, as the model keeps learning and improving with increasing available data without human bias or the limitation of thematic analytic model. DL can be implemented to understand the use of gyroscope for classification of physical activities using mobile phone motion sensor. In [180], 13 physical activities were considered, and the classification technique required the use of many algorithms: C4.5, Naive Bayes, logistic regression, KNN, and meta-algorithms such as boosting and bagging. DL technique called RBM will be appropriate to replace the implemented algorithms. The conditional distribution over the hidden nodes in RBM makes the feature presentation of each activity from the input signal possible.

Conclusions

ML is gradually influencing the way health care treatment and monitoring is performed. All of this can be attributed to the success recorded by DL. Compared with conventional ML and feature engineering, DL has potentially proven to provide response to data analysis and learning problems found in enormous volumes of data. Different variations of DL techniques have been implemented across many areas such as biomedical image, health record processing, sensors and physiological signal processing, human motion and emotion analysis, and so on. A successful AI system must have an excellent ML component; DL is taking the position as the number one choice for AI. To understand DL, in this review paper, we discussed about the basic architecture of DL methods. The discussion focused on principles of operation and application in health and medical domains. We presented the following models: (1) AE, (2) RNN, (3) CNN, (4) DBN, and (5) DBM. We presented a review of publications that have implemented these models in medical image, physiological signals, biological system, and EHR. We investigated the trend of DL implementation from 2012 to 2017. We observed a steady rise, with CNN having the highest increase occurrence. Computer and network architecture will gradually begin to change to support big data and DL techniques for efficiency and scalability. Moreover, there are some inherent challenges encountered in DL that need to be addressed. Most of these data in the real world are in unstructured format that cannot be processed by DL methods and require extra layer of encoding and representation. Clinical data are expensive to acquire and dataset contains incomplete and inconsistent records.

Statistical results presented in this review paper reveals that future applications and trends in DL will see more application of CNN implemented in medical image processing. There will be more variations of DL techniques across the general DL methods. There will be increase in the application of physiological signals using DL methods for diagnosis. The advancement in IoT and edge computing technology will bring about a different model of DL that will support this technology. Further research and study need to consider targeting this platform and solve issues relating to performance evaluation, scalability, and limited service capability. AI for mHealth will be driven by DL assisted by cloud and edge computing to process big data from wearable and mobile devices. Therefore, we can say that DL offers an excellent algorithm and is the answer to the challenges presented by MBD. However, the use of DL in every application that requires data analysis should not be done at the expense of other ML algorithms with less computation and memory requirement that are capable of producing similar results. Furthermore, attention should be given to other ML algorithms that have good possibility of achieving high performance with big data to deal with the demand for data analysis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NNSFC) under grant number 61871375, U1505251 and U1713219. National Key R&D Program of China under grant number 2018YFC2001002, Shenzhen Basic Research Project under grant number JCYJ20180507182231907, Chinese Academy of Sciences (CAS) key laboratory of health informatics, and the enhancement project for Shenzhen biomedical electronics technology public service platform. The support of Chinse Academy of Sciences and The World Academy of Sciences president's fellowship program.

Authors' Contributions

Authors ZN and LW are both corresponding authors for this paper.

Conflicts of Interest

None declared.

References

1. Hung K, Zhang YT, Tai B. Wearable medical devices for tele-home healthcare. Conf Proc IEEE Eng Med Biol Soc 2004;7:5384-5387. [doi: 10.1109/IEMBS.2004.1404503] [Medline: 17271560]
2. Muller H, Unay D. Retrieval from and understanding of large-scale multi-modal medical datasets: a review. IEEE T Multimedia 2017 Sep;19(9):2093-2104. [doi: 10.1109/TMM.2017.2729400]
3. Tarouco LM, Bertholdo LM, Granville LZ, Arbiza LM, Carbone F, Marotta M, et al. Internet of Things in Healthcare: Interoperatibility and Security Issues. In: Proceedings of the International Conference on Communications. 2012 Presented at: IEEE'12; June 10-15, 2012; Ottawa, Canada p. 6121-6125. [doi: 10.1109/ICC.2012.6364830]
4. Wu X, Zhu X, Wu GQ, Ding W. Data mining with big data. IEEE Trans Knowl Data Eng 2014 Jan;26(1):97-107 [FREE Full text] [doi: 10.1109/TKDE.2013.109]
5. Bergamaschi S. Big Data Analysis: Trends & Challenges. In: Proceedings of the International Conference on High Performance Computing & Simulation. 2014 Presented at: IEEE'14; July 21-25, 2014; Bologna, Italy p. 303-304. [doi: 10.1109/HPCSim.2014.6903699]
6. Chen XW, Lin X. Big data deep learning: challenges and perspectives. IEEE Access 2014;2(3):514-525. [doi: 10.1109/ACCESS.2014.2325029]
7. Wood J. Likythos: DSpace. 2010. Riding the Wave: How Europe Can Gain From the Rising Tide of Scientific Data. A Vision for 2030 URL: https://lekythos.library.ucy.ac.cy/bitstream/handle/10797/14288/4ekt023.pdf?sequence=1 [accessed 2018-04-16] [WebCite Cache ID 71i00en3l]
8. Istepanian RS, Al-Anzi T. m-Health 2.0: new perspectives on mobile health, machine learning and big data analytics. Methods 2018 Dec 1;151:34-40. [doi: 10.1016/j.ymeth.2018.05.015] [Medline: 29890285]
9. Ng AY. Nuts and Bolts of Building Applications Using Deep Learning. In: Proceedings of the Thirtieth Conference on Neural Information Processing Systems. 2016 Presented at: NIPS'16; December 5-10, 2016; Barcelona, Spain.
10. Greenspan H, van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 2016 May;35(5):1153-1159. [doi: 10.1109/TMI.2016.2553401]
11. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017 Dec 21;19:221-248 [FREE Full text] [doi: 10.1146/annurev-bioeng-071516-044442] [Medline: 28301734]
12. Keep M. DZone. 2018. Deep Learning and the Artificial Intelligence Revolution: Part 2 URL: https://dzone.com/articles/deep-learning-and-the-artificial-intelligence-revo-5 [accessed 2018-01-18] [WebCite Cache ID 71i2BVYWu]
13. Das S, Nene MJ. A Survey on Types of Machine Learning Techniques in Intrusion Prevention Systems. In: Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking. 2017 Presented at: IEEE'17; March 22-24, 2017; Chennai, India p. 2296-2299. [doi: 10.1109/WiSPNET.2017.8300169]
14. Guruvayur SR, Suchithra R. A Detailed Study on Machine Learning Techniques for Data Mining. In: Proceedings of the International Conference on Trends in Electronics and Informatics. 2017 Presented at: ICOEI'17; May 11-12, 2017; Tirunelveli, India p. 1187-1192. [doi: 10.1109/ICOEI.2017.8300900]
15. Almalaq A, Edwards G. A Review of Deep Learning Methods Applied on Load Forecasting. In: Proceedings of the International Conference on Machine Learning and Applications. 2017 Presented at: IEEE'17; December 18-21, 2017; Cancun, Mexico p. 511-516. [doi: 10.1109/ICMLA.2017.0-110]
16. Pandey M. Machine Learning and Systems for the Next Frontier in Formal Verification. In: Proceedings of the Formal Methods in Computer-Aided Design. 2016 Presented at: IEEE'16; October 3-6 , 2016; Mountain View, CA, USA. [doi: 10.1109/FMCAD.2016.7886650]
17. Wang G. A perspective on deep imaging. IEEE Access 2016;4:8914-8924. [doi: 10.1109/ACCESS.2016.2624938]
18. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2017 Presented at: IEEE'17; July 21-26, 2017; Honolulu, HI, USA p. 4700-4708. [doi: 10.1109/CVPR.2017.243]

19. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 2012 Nov;29(6):82-97. [doi: 10.1109/MSP.2012.2205597]
20. Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS'14; December 8-13, 2014; Montreal, Canada p. 3104-3112.
21. Perez CE. Advanced Micro Devices. 2018. The Potential Disruptiveness of AMD's Open Source Deep Learning Strategy URL: https://instinct.radeon.com/en/the-potential-disruptiveness-of-amds-open-source-deep-learning-strategy/ [accessed 2018-02-21] [WebCite Cache ID 71i7k6OEu]
22. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, MIT Critical Data Conference 2014 Organizing Committee. Making big data useful for health care: a summary of the inaugural mit critical data conference. JMIR Med Inform 2014 Aug 22;2(2):e22 [FREE Full text] [doi: 10.2196/medinform.3447] [Medline: 25600172]
23. Aggarwal LM. Advances in medical technology and its impact on health care in developing countries. Int J Radiol Radiat Tther 2017 Feb 27;2(2):569-576. [doi: 10.15406/ijrrt.2017.02.00022]
24. Wang CS, Lin PJ, Cheng CL, Tai SH, Yang YH, Chiang JH. Detecting potential adverse drug reactions using a deep neural network model. J Med Internet Res 2019 Feb 6;21(2):e11016 [FREE Full text] [doi: 10.2196/11016] [Medline: 30724742]
25. Zhang Y, Allem JP, Unger JB, Cruz TB. Automated identification of hookahs (waterpipes) on Instagram: an application in feature extraction using convolutional neural network and support vector machine classification. J Med Internet Res 2018 Dec 21;20(11):e10513 [FREE Full text] [doi: 10.2196/10513] [Medline: 30452385]
26. Rivas R, Montazeri N, Le NX, Hristidis V. Automatic classification of online doctor reviews: evaluation of text classifier algorithms. J Med Internet Res 2018 Dec 12;20(11):e11141 [FREE Full text] [doi: 10.2196/11141] [Medline: 30425030]
27. del Fiol G, Michelson M, Iorio A, Cotoi C, Haynes RB. A deep learning method to automatically identify reports of scientifically rigorous clinical research from the biomedical literature: comparative analytic study. J Med Internet Res 2018 Dec 25;20(6):e10281 [FREE Full text] [doi: 10.2196/10281] [Medline: 29941415]
28. Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, et al. An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study. J Med Internet Res 2018 Dec 14;20(11):e11144 [FREE Full text] [doi: 10.2196/11144] [Medline: 30429111]
29. Shi J, Wen H, Zhang Y, Han K, Liu Z. Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision. Hum Brain Mapp 2018 Dec;39(5):2269-2282 [FREE Full text] [doi: 10.1002/hbm.24006] [Medline: 29436055]
30. Han J, Zhang D, Wen S, Guo L, Liu T, Li X. Two-stage learning to predict human eye fixations via SDAEs. IEEE Trans Cybern 2016 Feb;46(2):487-498. [doi: 10.1109/TCYB.2015.2404432] [Medline: 25730840]
31. Yeganegi H, Fathi Y, Erfanian A. Decoding hind limb kinematics from neuronal activity of the dorsal horn neurons using multiple level learning algorithm. Sci Rep 2018 Dec 12;8(1):577 [FREE Full text] [doi: 10.1038/s41598-017-18971-x] [Medline: 29330489]
32. Murad A, Pyun JY. Deep recurrent neural networks for human activity recognition. Sensors (Basel) 2017 Nov 6;17(11):pii: E2556 [FREE Full text] [doi: 10.3390/s17112556] [Medline: 29113103]
33. Bing D, Ying J, Miao J, Lan L, Wang D, Zhao L, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. Clin Otolaryngol 2018 Dec;43(3):868-874. [doi: 10.1111/coa.13068] [Medline: 29356346]
34. Liang M, Li Z, Chen T, Zeng J. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Trans Comput Biol Bioinform 2015;12(4):928-937. [doi: 10.1109/TCBB.2014.2377729] [Medline: 26357333]
35. Hu X, Yu Z. Diagnosis of mesothelioma with deep learning. Oncol Lett 2019 Feb;17(2):1483-1490 [FREE Full text] [doi: 10.3892/ol.2018.9761] [Medline: 30675203]
36. Mathews SM, Kambhamettu C, Barner KE. A novel application of deep learning for single-lead ECG classification. Comput Biol Med 2018 Dec 1;99:53-62. [doi: 10.1016/j.compbiomed.2018.05.013] [Medline: 29886261]
37. Xue W, Islam A, Bhaduri M, Li S. Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Trans Med Imaging 2017 Dec;36(10):2057-2067. [doi: 10.1109/TMI.2017.2709251] [Medline: 28574348]
38. Zeng H, Yang C, Dai G, Qin F, Zhang J, Kong W. EEG classification of driver mental states by deep learning. Cogn Neurodyn 2018 Dec;12(6):597-606. [doi: 10.1007/s11571-018-9496-y] [Medline: 30483367]
39. Lu N, Li T, Ren X, Miao H. A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. IEEE Trans Neural Syst Rehabil Eng 2017 Dec;25(6):566-576. [doi: 10.1109/TNSRE.2016.2601240] [Medline: 27542114]
40. Chu Y, Zhao X, Zou Y, Xu W, Han J, Zhao Y. A decoding scheme for incomplete motor imagery EEG with deep belief network. Front Neurosci 2018;12:680 [FREE Full text] [doi: 10.3389/fnins.2018.00680] [Medline: 30323737]
41. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. Burlington, Massachusetts: Morgan Kaufmann; 2016.
42. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017 Dec;21(1):4-21. [doi: 10.1109/JBHI.2016.2636665] [Medline: 28055930]

43. Alam M, Vidyaratne LS, Iftekharuddin KM. Sparse simultaneous recurrent deep learning for robust facial expression recognition. IEEE Trans Neural Netw Learn Syst 2018 Dec;29(10):4905-4916. [doi: 10.1109/TNNLS.2017.2776248] [Medline: 29993957]
44. Takahashi N, Gygli M, van Gool L. AENet: Learning deep audio features for video analysis. IEEE T Multimedia 2018 Mar;20(3):513-524. [doi: 10.1109/TMM.2017.2751969]
45. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
46. Yousoff SN, Baharin A, Abdullah A. A Review on Optimization Algorithm for Deep Learning Method in Bioinformatics Field. In: Proceedings of the Conference on Biomedical Engineering and Sciences. 2016 Presented at: IEEE'16; December 4-8, 2016; Kuala Lumpur, Malaysia p. 707-711. [doi: 10.1109/IECBES.2016.7843542]
47. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
48. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and Composing Robust Features With Denoising Autoencoders. In: Proceedings of the 25th International Conference on Machine learning. 2008 Presented at: ICML'08; July 5-9, 2008; Helsinki, Finland p. 1096-1103. [doi: 10.1145/1390156.1390294]
49. Ranzato MA, Poultney C, Chopra S, LeCun Y. Efficient Learning of Sparse Representations With an Energy-Based Model. In: Proceedings of the 19th International Conference on Neural Information Processing Systems. 2006 Presented at: NIPS'06; December 4-7, 2006; Vancouver, British Columbia, Canada p. 1137-1144.
50. Masci J, Meier U, Cirean D, Schmidhuber J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. In: Proceedings of the Artificial Neural Networks and Machine Learning. 2011 Presented at: ICANN'11; June 14-17, 2011; Espoo, Finland p. 52-59. [doi: 10.1007/978-3-642-21735-7_7]
51. Ororbia LA, Kifer D, Giles CL. Unifying adversarial training algorithms with data gradient regularization. Neural Comput 2017 Dec;29(4):867-887. [doi: 10.1162/NECO_a_00928] [Medline: 28095194]
52. Rifai S, Vincent P, Muller X, Glorot X, Bengio Y. Contractive Auto-Encoders: Explicit Invariance During Feature Extraction. In: Proceedings of the 28th International Conference on Machine Learning. 2011 Presented at: ICML'11; June 28-July 2, 2011; Bellevue, Washington, USA p. 833-840 URL: https://dl.acm.org/citation.cfm?id=3104587
53. Schmidhuber J. WebCite. 1993. [Demonstrates Credit Assignment Across the Equivalent of 1,200 Layers in an Unfolded RNN] URL: http://www.webcitation.org/71i6G4Jaw [WebCite Cache ID 71i6G4Jaw]
54. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5(2):157-166. [doi: 10.1109/72.279181] [Medline: 18267787]
55. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
56. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]
57. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar p. 1724-1734. [doi: 10.3115/v1/D14-1179]
58. Matsugu M, Mori K, Mitari Y, Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw 2003;16(5-6):555-559. [doi: 10.1016/S0893-6080(03)00115-1] [Medline: 12850007]
59. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 1962;160(1):106-154. [doi: 10.1113/jphysiol.1962.sp006837] [Medline: 14449617]
60. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278-2324. [doi: 10.1109/5.726791]
61. Cirean D, Meier U, Schmidhuber J. Multi-Column Deep Neural Networks for Image Classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2012 Presented at: IEEE'12; June 16-21, 2012; Providence, RI, USA p. 3642-3649. [doi: 10.1109/CVPR.2012.6248110]
62. Salakhutdinov R, Larochelle H. Efficient Learning of Deep Boltzmann Machines. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010 Presented at: AISTATS'10; May 13-15, 2010; Sardinia, Italy p. 693-700 URL: http://proceedings.mlr.press/v9/salakhutdinov10a.html
63. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2006 Jul;18(7):1527-1554. [doi: 10.1162/neco.2006.18.7.1527] [Medline: 16764513]
64. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006 Jul 28;313(5786):504-507 [FREE Full text] [doi: 10.1126/science.1127647] [Medline: 16873662]
65. Ryu S, Noh J, Kim H. Deep neural network based demand side short term load forecasting. Energies 2016 Dec 22;10(1):3. [doi: 10.3390/en10010003]
66. Goodfellow I, Bengio Y, Courville A. Deep Learning (Adaptive Computation and Machine Learning Series). Cambridge, Massachusetts: MIT Press; 2016.

67. Saha M, Chakraborty C, Racoceanu D. Efficient deep learning model for mitosis detection using breast histopathology images. Comput Med Imaging Graph 2018 Dec;64:29-40. [doi: 10.1016/j.compmedimag.2017.12.001] [Medline: 29409716]
68. Saha M, Chakraborty C. Her2Net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. IEEE Trans Image Process 2018 May;27(5):2189-2200. [doi: 10.1109/TIP.2018.2795742] [Medline: 29432100]
69. Xu M, Papageorgiou DP, Abidi SZ, Dao M, Zhao H, Karniadakis GE. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. PLoS Comput Biol 2017 Oct;13(10):e1005746 [FREE Full text] [doi: 10.1371/journal.pcbi.1005746] [Medline: 29049291]
70. Wang Y, Qiu Y, Thai T, Moore K, Liu H, Zheng B. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. Comput Methods Programs Biomed 2017 Jun;144:97-104 [FREE Full text] [doi: 10.1016/j.cmpb.2017.03.017] [Medline: 28495009]
71. Xu Y, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinformatics 2017 May 26;18(1):281 [FREE Full text] [doi: 10.1186/s12859-017-1685-x] [Medline: 28549410]
72. Hughes TB, Dang NL, Miller GP, Swamidass SJ. Modeling reactivity to biological macromolecules with a deep multitask network. ACS Cent Sci 2016 Aug 24;2(8):529-537 [FREE Full text] [doi: 10.1021/acscentsci.6b00162] [Medline: 27610414]
73. Song Y, Zhang L, Chen S, Ni D, Li B, Zhou Y, et al. A Deep Learning Based Framework for Accurate Segmentation of Cervical Cytoplasm and Nuclei. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: IEEE'14; August 26-30, 2014; Chicago, IL, USA p. 2903-2906. [doi: 10.1109/EMBC.2014.6944230]
74. Gurcan MN, Sahiner B, Chan HP, Hadjiiski L, Petrick N. Selection of an optimal neural network architecture for computer-aided detection of microcalcifications--comparison of automated optimization techniques. Med Phys 2001 Sep;28(9):1937-1948. [doi: 10.1118/1.1395036] [Medline: 11585225]
75. Zhang YZ, Yamaguchi R, Imoto S, Miyano S. Sequence-specific bias correction for RNA-seq data using recurrent neural networks. BMC Genomics 2017 Dec 25;18(Suppl 1):1044 [FREE Full text] [doi: 10.1186/s12864-016-3262-5] [Medline: 28198674]
76. Ren J, Karagoz K, Gatza M, Foran DJ, Qi X. Differentiation among prostate cancer patients with Gleason score of 7 using histopathology whole-slide image and genomic data. Proc SPIE Int Soc Opt Eng 2018 Feb;10579:pii: 1057904 [FREE Full text] [doi: 10.1117/12.2293193] [Medline: 30662142]
77. Wang Y, Wang J, Lin H, Tang X, Zhang S, Li L. Bidirectional long short-term memory with CRF for detecting biomedical event trigger in FastText semantic space. BMC Bioinformatics 2018 Dec 21;19(Suppl 20):507 [FREE Full text] [doi: 10.1186/s12859-018-2543-1] [Medline: 30577839]
78. Chen G, Tsoi A, Xu H, Zheng WJ. Predict effective drug combination by deep belief network and ontology fingerprints. J Biomed Inform 2018 Sep;85:149-154. [doi: 10.1016/j.jbi.2018.07.024] [Medline: 30081101]
79. Beevi KS, Nair MS, Bindu GR. A multi-classifier system for automatic mitosis detection in breast histopathology images using deep belief networks. IEEE J Transl Eng Health Med 2017;5:4300211 [FREE Full text] [doi: 10.1109/JTEHM.2017.2694004] [Medline: 29018640]
80. Zhang Y, Du N, Li K, Feng J, Jia K, Zhang A. msiDBN: a method of identifying critical proteins in dynamic PPI networks. Biomed Res Int 2014;2014:138410 [FREE Full text] [doi: 10.1155/2014/138410] [Medline: 24800204]
81. Jiang X, Zhang H, Duan F, Quan X. Identify Huntington's disease associated genes based on restricted Boltzmann machine with RNA-seq data. BMC Bioinformatics 2017 Oct 11;18(1):447 [FREE Full text] [doi: 10.1186/s12859-017-1859-6] [Medline: 29020921]
82. Ghasemi F, Fassihi A, Pérez-Sánchez H, Mehri DA. The role of different sampling methods in improving biological activity prediction using deep belief network. J Comput Chem 2017 Dec 5;38(4):195-203. [doi: 10.1002/jcc.24671] [Medline: 27862046]
83. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun 2019 Dec 23;10(1):390 [FREE Full text] [doi: 10.1038/s41467-018-07931-2] [Medline: 30674886]
84. Guan R, Wang X, Yang MQ, Zhang Y, Zhou F, Yang C, et al. Multi-label deep learning for gene function annotation in cancer pathways. Sci Rep 2018 Dec 10;8(1):267 [FREE Full text] [doi: 10.1038/s41598-017-17842-9] [Medline: 29321535]
85. Chen HI, Chiu YC, Zhang T, Zhang S, Huang Y, Chen Y. GSAE: an autoencoder with embedded gene-set nodes for genomics functional characterization. BMC Syst Biol 2018 Dec 21;12(Suppl 8):142 [FREE Full text] [doi: 10.1186/s12918-018-0642-2] [Medline: 30577835]
86. Wang D, Gu J. VASC: dimension reduction and visualization of single-cell RNA-seq data by deep variational autoencoder. Genomics Proteomics Bioinformatics 2018 Dec;16(5):320-331 [FREE Full text] [doi: 10.1016/j.gpb.2018.08.003] [Medline: 30576740]
87. Maggio V, Chierici M, Jurman G, Furlanello C. Distillation of the clinical algorithm improves prognosis by multi-task deep learning in high-risk neuroblastoma. PLoS One 2018;13(12):e0208924 [FREE Full text] [doi: 10.1371/journal.pone.0208924] [Medline: 30532223]

88. Hu Q, Feng M, Lai L, Pei J. Prediction of drug-likeness using deep autoencoder neural networks. Front Genet 2018;9:585 [FREE Full text] [doi: 10.3389/fgene.2018.00585] [Medline: 30538725]
89. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: a convolutional net for medical records. IEEE J Biomed Health Inform 2017 Dec;21(1):22-30. [doi: 10.1109/JBHI.2016.2633963] [Medline: 27913366]
90. Lin C, Hsu CJ, Lou YS, Yeh SJ, Lee CC, Su SL, et al. Artificial intelligence learning semantics via external resources for classifying diagnosis codes in discharge notes. J Med Internet Res 2017 Dec 6;19(11):e380 [FREE Full text] [doi: 10.2196/jmir.8344] [Medline: 29109070]
91. Cheng Y, Wang F, Zhang P, Hu J. Risk Prediction With Electronic Health Records: A Deep Learning Approach. In: Proceedings of the International Conference on Data Mining. 2016 Presented at: SIAM'16; May 5-7, 2016; Miami, Florida p. 432-440. [doi: 10.1137/1.9781611974348.49]
92. Zeng X, Cao K, Zhang M. MobileDeepPill: A Small-Footprint Mobile Deep Learning System for Recognizing Unconstrained Pill Images. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. 2017 Presented at: MobiSys'17; June 19-23, 2017; Niagara Falls, New York, USA p. 56-67. [doi: 10.1145/3081333.3081336]
93. Li F, Liu W, Yu H. Extraction of information related to adverse drug events from electronic health record notes: design of an end-to-end model based on deep learning. JMIR Med Inform 2018 Nov 26;6(4):e12159 [FREE Full text] [doi: 10.2196/12159] [Medline: 30478023]
94. Zhang Y, Wang X, Hou Z, Li J. Clinical named entity recognition from Chinese electronic health records via machine learning methods. JMIR Med Inform 2018 Dec 17;6(4):e50 [FREE Full text] [doi: 10.2196/medinform.9965] [Medline: 30559093]
95. Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. Proc Conf Empir Methods Nat Lang Process 2016 Nov;2016:856-865 [FREE Full text] [doi: 10.18653/v1/D16-1082] [Medline: 28004040]
96. Jagannatha AN, Yu H. Bidirectional RNN for medical event detection in electronic health records. Proc Conf 2016 Jun;2016:473-482 [FREE Full text] [doi: 10.18653/v1/N16-1056] [Medline: 27885364]
97. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18. [doi: 10.1038/s41746-018-0029-1]
98. Hou WJ, Ceesay B. Extraction of drug-drug interaction using neural embedding. J Bioinform Comput Biol 2018 Dec;16(6):1840027. [doi: 10.1142/S0219720018400279] [Medline: 30567477]
99. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
100. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017 Dec 1;24(2):361-370 [FREE Full text] [doi: 10.1093/jamia/ocw112] [Medline: 27521897]
101. Volkova S, Ayton E, Porterfield K, Corley CD. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. PLoS One 2017;12(12):e0188941 [FREE Full text] [doi: 10.1371/journal.pone.0188941] [Medline: 29244814]
102. Yadav S, Ekbal A, Saha S, Bhattacharyya P. Deep Learning Architecture for Patient Data De-Identification in Clinical Records. In: Proceedings of the Clinical Natural Language Processing Workshop.: The COLING 2016 Organizing Committee; 2016 Presented at: ClinicalNLP'16; December 11, 2016; Osaka, Japan p. 32-41 URL: https://aclweb.org/anthology/papers/W/W16/W16-4206/
103. Hassanien AE, Al-Shammari ET, Ghali NI. Computational intelligence techniques in bioinformatics. Comput Biol Chem 2013 Dec;47:37-47. [doi: 10.1016/j.compbiolchem.2013.04.007] [Medline: 23891719]
104. Li H, Li X, Ramanathan M, Zhang A. Identifying informative risk factors and predicting bone disease progression via deep belief networks. Methods 2014 Oct 1;69(3):257-265. [doi: 10.1016/j.ymeth.2014.06.011] [Medline: 24979059]
105. Che Z, Kale D, Li W, Bahadori MT, Liu Y. Deep Computational Phenotyping. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD'15; August 10-13, 2015; Sydney, NSW, Australia p. 507-516. [doi: 10.1145/2783258.2783365]
106. Tran T, Nguyen TD, Phung D, Venkatesh S. Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). J Biomed Inform 2015 Apr;54:96-105 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.012] [Medline: 25661261]
107. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
108. Lv X, Guan Y, Yang J, Wu J. Clinical relation extraction with deep learning. Int J Hybrid Inf Technol 2016 Jul 31;9(7):237-248. [doi: 10.14257/ijhit.2016.9.7.22]
109. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PLoS One 2013;8(6):e66341 [FREE Full text] [doi: 10.1371/journal.pone.0066341] [Medline: 23826094]
110. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, et al. Prediction of occult invasive disease in Ductal carcinoma in situ using deep learning features. J Am Coll Radiol 2018 Dec;15(3 Pt B):527-534 [FREE Full text] [doi: 10.1016/j.jacr.2017.11.036] [Medline: 29398498]

111. Wang SH, Lv YD, Sui Y, Liu S, Wang SJ, Zhang YD. Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. J Med Syst 2017 Nov 17;42(1):2. [doi: 10.1007/s10916-017-0845-x] [Medline: 29159706]
112. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Isgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging 2016 Dec;35(5):1252-1261. [doi: 10.1109/TMI.2016.2548501] [Medline: 27046893]
113. Abiyev RH, Ma'aitah MK. Deep convolutional neural networks for chest diseases detection. J Healthc Eng 2018;2018:4168538 [FREE Full text] [doi: 10.1155/2018/4168538] [Medline: 30154989]
114. Liu C, Cao Y, Luo Y, Chen G, Vokkarane V, Ma Y. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. In: Proceedings of International Conference on Inclusive Smart Cities and Digital Health. 2016 Presented at: ICOST'16; May 25-27, 2016; Wuhan, China p. 37-48. [doi: 10.1007/978-3-319-39601-9_4]
115. Nie S, Yu J, Chen P, Wang Y, Zhang JQ. Automatic detection of standard sagittal plane in the first trimester of pregnancy using 3-D ultrasound data. Ultrasound Med Biol 2017 Dec;43(1):286-300. [doi: 10.1016/j.ultrasmedbio.2016.08.034] [Medline: 27810260]
116. Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics 2016 Dec;72:150-157. [doi: 10.1016/j.ultras.2016.08.004] [Medline: 27529139]
117. Cao Y, Steffey S, He J, Xiao D, Tao C, Chen P, et al. Medical image retrieval: a multimodal approach. Cancer Inform 2014;13(Suppl 3):125-136 [FREE Full text] [doi: 10.4137/CIN.S14053] [Medline: 26309389]
118. Wu J, Mazur TR, Ruan S, Lian C, Daniel N, Lashmett H, et al. A deep Boltzmann machine-driven level set method for heart motion tracking using cine MRI images. Med Image Anal 2018 Dec;47:68-80. [doi: 10.1016/j.media.2018.03.015] [Medline: 29679848]
119. Jang H, Plis SM, Calhoun VD, Lee JH. Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: evaluation using sensorimotor tasks. Neuroimage 2017 Dec 15;145(Pt B):314-328 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.04.003] [Medline: 27079534]
120. Suk HI, Lee SW, Shen D, Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 2014 Nov 1;101:569-582 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.06.077] [Medline: 25042445]
121. Khatami A, Khosravi A, Lim CP, Nahavandi S. A Wavelet Deep Belief Network-Based Classifier for Medical Images. In: Proceedings of the International Conference on Neural Information Processing. 2016 Presented at: ICONIP'16; October 16-21, 2016; Kyoto, Japan p. 467-474. [doi: 10.1007/978-3-319-46675-0_51]
122. Zhang S, Dong Q, Zhang W, Huang H, Zhu D, Liu T. Discovering hierarchical common brain networks via multimodal deep belief network. Med Image Anal 2019 May;54:238-252. [doi: 10.1016/j.media.2019.03.011] [Medline: 30954851]
123. Uzunova H, Schultz S, Handels H, Ehrhardt J. Unsupervised pathology detection in medical images using conditional variational autoencoders. Int J Comput Assist Radiol Surg 2019 Mar;14(3):451-461. [doi: 10.1007/s11548-018-1898-0] [Medline: 30542975]
124. Lee CY, Chen GL, Zhang ZX, Chou YH, Hsu CC. Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network? J Healthc Eng 2018;2018:8413403 [FREE Full text] [doi: 10.1155/2018/8413403] [Medline: 30651947]
125. Seebock P, Waldstein SM, Klimscha S, Bogunovic H, Schlegl T, Gerendas BS, et al. Unsupervised identification of disease marker candidates in retinal OCT imaging data. IEEE Trans Med Imaging 2019 Apr;38(4):1037-1047. [doi: 10.1109/TMI.2018.2877080] [Medline: 30346281]
126. Wang X, Zhai S, Niu Y. Automatic vertebrae localization and identification by combining deep SSAE contextual features and structured regression forest. J Digit Imaging 2019 Apr;32(2):336-348. [doi: 10.1007/s10278-018-0140-5] [Medline: 30631979]
127. Malek S, Melgani F, Mekhalfi ML, Bazi Y. Real-time indoor scene description for the visually impaired using autoencoder fusion strategies with visible cameras. Sensors (Basel) 2017 Nov 16;17(11):pii: E2641 [FREE Full text] [doi: 10.3390/s17112641] [Medline: 29144395]
128. Zhang X, Dou H, Ju T, Xu J, Zhang S. Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. IEEE J Biomed Health Inform 2016 Dec;20(5):1377-1383. [doi: 10.1109/JBHI.2015.2461671] [Medline: 26241980]
129. Xia C, Qi F, Shi G. Bottom-up visual saliency estimation with deep autoencoder-based sparse reconstruction. IEEE Trans Neural Netw Learn Syst 2016 Dec;27(6):1227-1240. [doi: 10.1109/TNNLS.2015.2512898] [Medline: 26800552]
130. Mano H, Kotecha G, Leibnitz K, Matsubara T, Sprenger C, Nakae A, et al. Classification and characterisation of brain network changes in chronic back pain: a multicenter study. Wellcome Open Res 2018;3:19 [FREE Full text] [doi: 10.12688/wellcomeopenres.14069.2] [Medline: 29774244]
131. Zhang J, Wu Y. A new method for automatic sleep stage classification. IEEE Trans Biomed Circuits Syst 2017 Dec;11(5):1097-1110. [doi: 10.1109/TBCAS.2017.2719631] [Medline: 28809709]
132. Veiga DJ, O'Reilly M, Whelan D, Caulfield B, Ward TE. Feature-free activity classification of inertial sensor data with machine vision techniques: method, development, and evaluation. JMIR Mhealth Uhealth 2017 Aug 4;5(8):e115 [FREE Full text] [doi: 10.2196/mhealth.7521] [Medline: 28778851]

133. Lenz M, Musso M, Linke Y, Tüscher O, Timmer J, Weiller C, et al. Joint EEG/fMRI state space model for the detection of directed interactions in human brains--a simulation study. Physiol Meas 2011 Nov;32(11):1725-1736. [doi: 10.1088/0967-3334/32/11/S01] [Medline: 22027197]
134. Liu YT, Lin YY, Wu SL, Chuang CH, Lin CT. Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network. IEEE Trans Neural Netw Learn Syst 2016 Feb;27(2):347-360. [doi: 10.1109/TNNLS.2015.2496330] [Medline: 26595929]
134. Yu Z, Lee M. Real-time human action classification using a dynamic neural model. Neural Netw 2015 Sep;69:29-43. [doi: 10.1016/j.neunet.2015.04.013] [Medline: 26042805]
135. Vakulenko S, Radulescu O, Morozov I, Weber A. Centralized networks to generate human body motions. Sensors (Basel) 2017 Dec 14;17(12):pii: E2907 [FREE Full text] [doi: 10.3390/s17122907] [Medline: 29240694]
137. Mo L, Li F, Zhu Y, Huang A. Human Physical Activity Recognition Based on Computer Vision With Deep Learning Model. In: International Instrumentation and Measurement Technology Conference Proceedings. 2016 Presented at: IEEE'16; May 23-26, 2016; Taipei, Taiwan p. 1-6. [doi: 10.1109/I2MTC.2016.7520541]
138. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors (Basel) 2016 Jan 18;16(1):pii: E115 [FREE Full text] [doi: 10.3390/s16010115] [Medline: 26797612]
139. Turner JT, Page A, Mohsenin T, Oates T. Deep Belief Networks Used on High Resolution Multichannel Electroencephalography Data for Seizure Detection. In: Proceedings of the Computer Vision and Pattern Recognition; Artificial Intelligence. 2014 Presented at: AAAI'14; March 24-26, 2014; Palo Alto, California p. 75-81 URL: https://arxiv.org/abs/1708.08430
140. Chao H, Zhi H, Dong L, Liu Y. Recognition of emotions using multichannel EEG data and DBN-GC-based ensemble deep learning framework. Comput Intell Neurosci 2018;2018:9750904 [FREE Full text] [doi: 10.1155/2018/9750904] [Medline: 30647727]
141. Jindal V. Integrating Mobile and Cloud for PPG Signal Selection to Monitor Heart Rate During Intensive Physical Exercise. In: Proceedings of the International Conference on Mobile Software Engineering and Systems. 2016 Presented at: MOBILESoft'16; May 14-22, 2016; Austin, Texas p. 36-37. [doi: 10.1145/2897073.2897132]
142. Hassan MM, Uddin MZ, Mohamed A, Almogren A. A robust human activity recognition system using smartphone sensors and deep learning. Future Gener Comput Syst 2018 Apr;81:307-313. [doi: 10.1016/j.future.2017.11.029]
143. Ruiz-Rodríguez JC, Ruiz-Sanmartín A, Ribas V, Caballero J, García-Roche A, Riera J, et al. Innovative continuous non-invasive cuffless blood pressure monitoring based on photoplethysmography technology. Intensive Care Med 2013 Sep;39(9):1618-1625. [doi: 10.1007/s00134-013-2964-2] [Medline: 23740275]
144. Yuan Y, Xun G, Jia K, Zhang A. A multi-view deep learning framework for EEG seizure detection. IEEE J Biomed Health Inform 2019 Dec;23(1):83-94. [doi: 10.1109/JBHI.2018.2871678] [Medline: 30624207]
145. Jirayucharoensak S, Pan-Ngum S, Israsena P. EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. ScientificWorldJournal 2014;2014:627892 [FREE Full text] [doi: 10.1155/2014/627892] [Medline: 25258728]
146. Jokanovic B, Amin M. Fall detection using deep learning in range-Doppler radars. IEEE Trans Aerosp Electron Syst 2018 Feb;54(1):180-189. [doi: 10.1109/TAES.2017.2740098]
147. Xia Y, Zhang H, Xu L, Gao Z, Zhang H, Liu H, et al. An automatic cardiac arrhythmia classification system with wearable electrocardiogram. IEEE Access 2018;6:16529-16538. [doi: 10.1109/ACCESS.2018.2807700]
148. Angelov P, Sperduti A. Challenges in deep learning. In: ESANN'16 - 24th European Symposium on Artificial Neural Networks. Bruges, Belgium: i6doc; 2016:489-495.
149. Yang X, Zhang T, Xu C. Cross-domain feature learning in multimedia. IEEE T Multimedia 2015 Jan;17(1):64-78. [doi: 10.1109/TMM.2014.2375793]
150. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data 2015 Feb 24;2(1):1-21. [doi: 10.1186/s40537-014-0007-7]
151. Hasan M, Roy-Chowdhury AK. A continuous learning framework for activity recognition using deep hybrid feature models. IEEE T Multimedia 2015 Nov;17(11):1909-1922. [doi: 10.1109/TMM.2015.2477242]
152. Du L, Du Y, Li Y, Su J, Kuan YC, Liu CC, et al. A reconfigurable streaming deep convolutional neural network accelerator for internet of things. IEEE Trans Circuits Syst 2018 Jan;65(1):198-208. [doi: 10.1109/TCSI.2017.2735490]
153. Zhou G, Sohn K, Lee H. Online Incremental Feature Learning With Denoising Autoencoders. In: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. 2012 Presented at: AISTATS'12; April 21-23, 2012; La Palma, Canary Island p. 1453-1461 URL: http://proceedings.mlr.press/v22/zhou12b.html
154. Xie S, Yang T, Wang X, Lin Y. Hyper-Class Augmented and Regularized Deep Learning for Fine-Grained Image Classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2015 Presented at: IEEE'15; June 7-12, 2015; Boston, MA, USA p. 2645-2654. [doi: 10.1109/CVPR.2015.7298880]
155. Yang J, Jiang B, Li B, Tian K, Lv Z. A fast image retrieval method designed for network big data. IEEE Trans Ind Inf 2017 Oct;13(5):2350-2359. [doi: 10.1109/TII.2017.2657545]
156. Dang C, Radha H. RPCA-KFE: key frame extraction for video using robust principal component analysis. IEEE Trans Image Process 2015 Nov;24(11):3742-3753. [doi: 10.1109/TIP.2015.2445572] [Medline: 26087486]

157. Slobogean GP, Giannoudis PV, Frihagen F, Forte ML, Morshed S, Bhandari M. Bigger data, bigger problems. J Orthop Trauma 2015 Dec;29(Suppl 12):S43-S46. [doi: 10.1097/BOT.0000000000000463] [Medline: 26584266]
158. Scruggs SB, Watson K, Su AI, Hermjakob H, Yates 3rd JR, Lindsey ML, et al. Harnessing the heart of big data. Circ Res 2015 Mar 27;116(7):1115-1119 [FREE Full text] [doi: 10.1161/CIRCRESAHA.115.306013] [Medline: 25814682]
159. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract 2017 Mar;36(1):3-11 [FREE Full text] [doi: 10.23876/j.krcp.2017.36.1.3] [Medline: 28392994]
160. Raina R, Madhavan A, Ng AY. Large-Scale Deep Unsupervised Learning Using Graphics Processors. In: Proceedings of the 26th Annual International Conference on Machine Learning. 2009 Presented at: ICML'09; June 14-18, 2009; Montreal, QC, Canada p. 873-880. [doi: 10.1145/1553374.1553486]
161. Alsheikh MA, Niyato D, Lin S, Tan HP, Han Z. Mobile big data analytics using deep learning and apache spark. IEEE Netw 2016 May;30(3):22-29. [doi: 10.1109/MNET.2016.7474340]
162. Page A, Turner JT, Mohsenin T, Oates T. Comparing raw data and feature extraction for seizure detection with deep learning methods. : AAAI Publications; May Presented at: The Twenty-Seventh International Florida Artificial Intelligence Research Society Conference; May 21-23, 2014; Florida USA p. 284-287.
163. Ziat A, Contardo G, Baskiotis N, Denoyer L. Learning embeddings for completion and prediction of relationnal multivariate time-series. In: ESANN 2016 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium: i6doc; 2016.
164. Ahmed E, Yaqoob I, Gani A, Imran M, Guizani M. Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges. IEEE Wireless Commun 2016 Oct;23(5):10-16. [doi: 10.1109/MWC.2016.7721736]
165. Evans D. Cisco. 2011. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything URL: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf [WebCite Cache ID 71iOmzVsC]
166. Kaur K, Garg S, Aujla GS, Kumar N, Rodrigues JJ, Guizani M. Edge computing in the industrial internet of things environment: software-defined-networks-based edge-cloud interplay. IEEE Commun Mag 2018 Feb;56(2):44-51. [doi: 10.1109/MCOM.2018.1700622]
167. Li H, Ota K, Dong M. Learning IoT in edge: deep learning for the internet of things with edge computing. IEEE Netw 2018 Jan;32(1):96-101. [doi: 10.1109/MNET.2018.1700202]
168. Hosseini MP, Pompili D, Elisevich K, Soltanian-Zadeh H. Optimized deep learning for EEG big data and seizure prediction BCI via internet of things. IEEE Trans Big Data 2017 Dec 1;3(4):392-404. [doi: 10.1109/TBDATA.2017.2769670]
169. Alhussein M, Muhammad G. Voice pathology detection using deep learning on mobile healthcare framework. IEEE Access 2018;6:41034-41041. [doi: 10.1109/ACCESS.2018.2856238]
170. Cao Y, Liu C, Liu B, Brunette MJ, Zhang N, Sun T, et al. Improving Tuberculosis Diagnostics Using Deep Learning and Mobile Health Technologies Among Resource-Poor and Marginalized Communities. In: Proceedings of the First International Conference on Connected Health: Applications, Systems and Engineering Technologies. 2016 Presented at: IEEE'16; June 27-29, 2016; Washington, DC, USA p. 274-281. [doi: 10.1109/CHASE.2016.18]
171. Liu C, Cao Y, Luo Y, Chen G, Vokkarane V, Yunsheng M, et al. A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. IEEE Trans Serv Comput 2018 Mar 1;11(2):249-261. [doi: 10.1109/TSC.2017.2662008]
172. Sun Y, Wang X, Tang X. Hybrid Deep Learning for Face Verification. IEEE Trans Pattern Anal Mach Intell 2016 Dec;38(10):1997-2009. [doi: 10.1109/TPAMI.2015.2505293] [Medline: 26660699]
173. Brady CJ, Mudie LI, Wang X, Guallar E, Friedman DS. Improving consensus scoring of crowdsourced data using the rasch model: development and refinement of a diagnostic instrument. J Med Internet Res 2017 Dec 20;19(6):e222 [FREE Full text] [doi: 10.2196/jmir.7984] [Medline: 28634154]
174. Bayen E, Jacquemot J, Netscher G, Agrawal P, Noyce LT, Bayen A. Reduction in fall rate in dementia managed care through video incident review: pilot study. J Med Internet Res 2017 Dec 17;19(10):e339 [FREE Full text] [doi: 10.2196/jmir.8095] [Medline: 29042342]
175. Wen D, Zhang X, Liu X, Lei J. Evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. J Med Internet Res 2017 Dec 7;19(3):e68 [FREE Full text] [doi: 10.2196/jmir.6874] [Medline: 28270382]
176. Kang SH, Joe B, Yoon Y, Cho GY, Shin I, Suh JW. Cardiac auscultation using smartphones: pilot study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e49 [FREE Full text] [doi: 10.2196/mhealth.8946] [Medline: 29490899]
177. Neubert TA, Dusch M, Karst M, Beissner F. Designing a tablet-based software app for mapping bodily symptoms: usability evaluation and reproducibility analysis. JMIR Mhealth Uhealth 2018 May 30;6(5):e127 [FREE Full text] [doi: 10.2196/mhealth.8409] [Medline: 29848470]
178. Doerr M, Truong AM, Bot BM, Wilbanks J, Suver C, Mangravite LM. Formative evaluation of participant experience with mobile e-consent in the app-mediated Parkinson mPower study: a mixed methods study. JMIR Mhealth Uhealth 2017 Feb 16;5(2):e14 [FREE Full text] [doi: 10.2196/mhealth.6521] [Medline: 28209557]
179. Meng J, Hussain SA, Mohr DC, Czerwinski M, Zhang M. Exploring user needs for a mobile behavioral-sensing technology for depression management: qualitative study. J Med Internet Res 2018 Dec 17;20(7):e10139 [FREE Full text] [doi: 10.2196/10139] [Medline: 30021710]

180. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. J Med Internet Res 2012 Oct 5;14(5):e130 [FREE Full text] [doi: 10.2196/jmir.2208] [Medline: 23041431]

Abbreviations

2D: two-dimensional
3D: three-dimensional
AE: autoencoder
AI: artificial intelligence
API: application programming interface
BM: Boltzmann machine
CMD: continuous medical data
CNN: convolutional neural network
CRF: conditional random fields
CT: computed tomography
DBM: deep Boltzmann machine
DBN: deep belief network
DBN-GCs: deep belief network with glia chains
DL: deep learning
DNN: deep neural network
DR: data representation
ECG: electrocardiogram
EEG: electroencephalogram
EHR: electronic health record
EMG: electromyography
FC: fully connected
GPU: graphics processing unit
IoT: Internet of Things
KNN: K-nearest neighbor
LSTM: long short-term memory
MBD: medical big data
mHealth: mobile health
ML: machine learning
PPG: photoplethysmogram
RBM: restricted Boltzmann machine
ReLU: rectified linear unit
RNN: recurrent neural network
scRNA-seq: small cytoplasmic RNA sequence
SSAE: stacked sparse autoencoder
SVM: support vector machine

Edited by G Eysenbach; submitted 17.08.18; peer-reviewed by B Wang, MS Liew; comments to author 29.11.18; revised version received 14.04.19; accepted 12.06.19; published 02.08.19.

Please cite as:
Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L
Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations
JMIR Mhealth Uhealth 2019;7(8):e11966
URL: https://mhealth.jmir.org/2019/8/e11966/
doi:10.2196/11966
PMID:31376272

Viewpoint

The SMART Framework: Integration of Citizen Science, Community-Based Participatory Research, and Systems Science for Population Health Science in the Digital Age

Tarun Reddy Katapally[1], MS, PhD

Johnson Shoyama Graduate School of Public Policy, University of Regina, Regina, SK, Canada

Corresponding Author:
Tarun Reddy Katapally, MS, PhD
Johnson Shoyama Graduate School of Public Policy
University of Regina
2155 College Ave
Regina, SK, S4P4V5
Canada
Phone: 1 3065854544
Email: tarun.katapally@uregina.ca

Abstract

Citizen science enables citizens to actively contribute to all aspects of the research process, from conceptualization and data collection, to knowledge translation and evaluation. Citizen science is gradually emerging as a pertinent approach in population health research. Given that citizen science has intrinsic links with community-based research, where participatory action drives the research agenda, these two approaches could be integrated to address complex population health issues. Community-based participatory research has a strong record of application across multiple disciplines and sectors to address health inequities. Citizen science can use the structure of community-based participatory research to take local approaches of problem solving to a global scale, because citizen science emerged through individual environmental activism that is not limited by geography. This synergy has significant implications for population health research if combined with systems science, which can offer theoretical and methodological strength to citizen science and community-based participatory research. Systems science applies a holistic perspective to understand the complex mechanisms underlying causal relationships within and between systems, as it goes beyond linear relationships by utilizing big data–driven advanced computational models. However, to truly integrate citizen science, community-based participatory research, and systems science, it is time to realize the power of ubiquitous digital tools, such as smartphones, for connecting us all and providing big data. Smartphones have the potential to not only create equity by providing a voice to disenfranchised citizens but smartphone-based apps also have the reach and power to source big data to inform policies. An imminent challenge in legitimizing citizen science is minimizing bias, which can be achieved by standardizing methods and enhancing data quality—a rigorous process that requires researchers to collaborate with citizen scientists utilizing the principles of community-based participatory research action. This study advances SMART, an evidence-based framework that integrates citizen science, community-based participatory research, and systems science through ubiquitous tools by addressing core challenges such as citizen engagement, data management, and internet inequity to legitimize this integration.

(JMIR Mhealth Uhealth 2019;7(8):e14056) doi:10.2196/14056

KEYWORDS

community-based participatory research; smartphones; mobile phones; population health; mHealth; eHealth; digital health; big data; evidence-based framework; citizen science; participatory research; participatory surveillance; systems science; ubiquitous tools

Population Health Science in the Digital Age

Global population health crises in the 21st century are extremely complex, with links to economic disasters [1-3], warfare [3,4], and climate change [3,5]. The digital age offers new opportunities and challenges for population health science to tackle these global crises. For instance, digital tools and technologies are increasingly being used to not only address urgent humanitarian crises [6] but also facilitate citizen

participation, population health interventions, and knowledge transfer [7].

However, digital technologies are rarely evaluated for their impact on health outcomes [6], and perhaps more relevant to population health, citizens, communities, and researchers are seldom involved in designing digital tools [7,8]. Given that most digital tools are developed for profit, where privacy and confidentiality are major concerns [8,6], adoption of digital technology in population health science should take special consideration to remedy ethical pitfalls.

Given that addressing health inequities is one of the primary goals of population health science [9-11], there is a role for digital tools and technologies in contributing to the reduction of health disparities. Nevertheless, a digitally driven big data approach has the risk of widening existing health disparities if care is not taken to engage diverse populations to realize potential benefits [12]. Complex global health crises are exacerbated by health inequities and are difficult to address using traditional research practices, thus citizen science is emerging as a powerful approach that policy makers are increasingly employing in determining population health solutions [13,14].

An example of citizen science interventions is the ability to harness technology to inform active, healthy neighborhoods in developing countries [15]. This bottom-up approach of involving community members in research can also be an effective way to translate knowledge to a broader audience [14,16] and has inherent overlap with community-based participatory research, which is known to facilitate community involvement in informing research, policy, and practice [17].

Community-based participatory research has led the crusade of minimizing health disparities [18,19] by engaging communities equitably in all aspects of the research process [20]. However, global population health problems require holistic systems science solutions that enable the examination of complex interconnected factors influencing health outcomes [21,22]. As systems science has been identified as particularly effective for addressing health inequities [22], it is time now to reimagine the implementation of population health science in the digital age by developing a framework that integrates citizen science, community-based participatory research, and systems science (ie, three-pronged approach).

Citizen Science

Citizen science enables citizens to actively contribute to all aspects of the research process, from conceptualization and data collection, to knowledge translation and evaluation [23,24]. Citizen science has its roots in environmental and ecological activism [25], where volunteers across the globe can collect data without being restricted by the constraints of traditional academic research [26].

The applicability of citizen science has become increasingly interdisciplinary over the past couple of decades [27], which has implications for population health science—a field of science that plays a key role in addressing broad health inequities [28]. Moreover, with increasing power of citizens to effect change [13], citizen science is earning a place in national science policies of countries such as the United States and Australia by complementing the efforts of governments and health professionals [14].

Citizen scientists can support and encourage regional data collection in a manner that will help researchers and policy makers to avoid the *one-size fits all* approach [14]. Involvement of citizens in research procedures is also an effective way to communicate health information to a broad audience [14,16]. Rowbotham et al [14] argue that citizen science offers an opportunity to transform population health science by engaging a larger proportion of the population in data collection to bring citizen scientist perspectives closer to traditional decision-making processes.

As citizen science can range from contributory (ie, data collection) and collaborative approaches (ie, analysis and interpretation of data) to cocreation of knowledge (ie, conceptualizing research and translating knowledge) [27], it has a natural overlap with community-based participatory research.

Community-Based Participatory Research and Citizen Science

Community-based participatory research has a strong record of application across multiple disciplines and sectors to address health inequities [29-31]. A significant emphasis of community-based participatory research is on generating empirical evidence on social determinants of health, which in essence could be used to inform and influence population health policies.

However, as community-based participatory research is entrenched in human rights and social justice, it can be applied to promote local and regional policy change by bringing together community needs, scientific evidence, and political power [31]. In the path toward policy advocacy, participatory action can be the catalyst in engaging evidence with political power [31]. A key component in this participatory action is the involvement of diverse stakeholders in all aspects of the research process through shared responsibility and research ownership [29,32-37].

This foundational strength of community-based participatory research can result in symbiosis with citizen science in the realm of population health science, as there is considerable alignment between these 2 research approaches [38]. To translate citizen science into community voices that could potentially inform and influence policies [24], it is imperative that citizen scientist endeavors are structured using community-based participatory research principles, where citizens co-design studies and cocreate knowledge with researchers by contributing to all aspects of the research process [24,39].

As citizens are ultimately potential voters, citizen science can perhaps stimulate renewed impetus among decision makers and catalyze evidence-based policy formulation. The alignment with citizen science can catapult community-based participatory research action from local endeavors to global initiatives because citizen science is not restricted by geography, jurisdictions, or

populations, thus enabling local solutions to global problems. Data collection on a global scale is the essence of citizen science, and in the 21st century, the potential for participatory modeling with a systems lens enables systems science via citizen science to solve population health issues [40].

Systems Science, Community-Based Participatory Research, and Citizen Science

Systems science takes a holistic perspective in understanding complex mechanisms by focusing on causality and going beyond linear correlations to advance data analytics through computational modeling that could inform public health decisions [41-48]. El-Sayed and Galea [46] conclude that a systems science approach can upend traditional analytical tools with novel methods such as machine learning, microsimulation, and social network analysis that do justice to the complexity and dynamism of population health science.

Systems thinking also includes qualitative mapping and problem structuring of ill-defined issues that enable stakeholders to directly address health disparities [47]. A systems science approach for structuring problems aligns with community-based participatory research by taking into consideration the need for stakeholder involvement to tackle systemic health inequities [48-50].

Frerichs et al [41] identified 5 areas of synergy between systems science and community-based participatory research—paradigmatic, socioecological, capacity building, colearning, and translational. These synergies provide a rationale for integrating systems science and community-based participatory research, with the central concept revolving around qualitative problem-structuring and systems mapping — a process that prioritizes research questions for computational modeling to delineate the complex pathways that influence health disparities.

Systems science can change the paradigm of population health science when combined with community-based participatory research and citizen science. Apart from their interdisciplinarity, the 3 research approaches have tremendous potential for symbiosis if they are used together for reciprocal benefit. For instance, community-based participatory research-informed qualitative mapping could benefit from citizen perceptions at the population level [41], an aspect that is core to citizen science.

Nevertheless, the crux of citizen science is citizen-driven quantitative data collection [51], which has implications for both systems science and population health science. Large groups of citizens could provide individual, environmental, and social big data needed for examining interconnected, nonlinear relationships through agent-based modeling that could inform population health policies [24,52]. A systems perspective can broaden citizen science as an anchor for population health science, which is increasingly adopting socioecological frameworks [53-55]. The realization of research conducted using these frameworks requires complex multilevel subjective and objective data that enable linkages across geographic and jurisdictional barriers.

Although there are examples of citizen science projects combining social and ecological data [56-59], participation of citizens in population health endeavors (ie, large-scale linkage of individual-level health data with administrative and contextual data) has rarely been tested or implemented. Citizen science has predominantly been used for environmental change or ecological activism, whether it is reconstruction of native landscapes in disaster zones in Iraq after the collapse of Saddam Hussein or the determination of social, environmental, and economic associations in high-risk industries such as coal mining [60,61].

The bottom-up approach of citizen science to bring about social and environmental change has a place in population health science if integrated with community-based participatory research [51]. Nevertheless, the need for big data will eventually thrust systems science to the forefront in conducting population health science [62], and this need can be addressed by leveraging citizen-owned ubiquitous digital tools for large-scale data collection by combining citizen science and community-based participatory research.

Integration Through Ubiquitous Tools

The growth of citizen science in recent years can be attributed to the global spread of internet-connected devices [63], and among all internet-connected devices, the ubiquitous presence of smartphones is the most significant development for citizen science-enabled population health research. The phenomenal growth of smartphone ownership in both developed and developing countries [64] creates an opportunity for the proliferation of citizen science across international borders.

Methodologically, the technology that powers these devices offers extraordinary research opportunities for population health science to overcome traditional constraints in terms of participant recruitment and retention, data collection and analysis, interventions, and knowledge translation [24]. Because smartphones have become a key part of day-to-day life and the primary one-stop communication device at home, work, and on the go [24,65], researchers can collaborate with citizen scientists to triangulate critical data: qualitative perceptions via audio, video, and photo-enabled ecological momentary assessments; traditional and novel quantitative data via deployment of surveys in real time; and objective data sensed via in-built smartphone sensors (eg, accelerometers, pedometers, and global positioning system) [24]. This triangulation allows consistent and longitudinal capture of big data in terms of volume, velocity, variety, and veracity [66], which is essential for the development of novel computational models informed by systems science approaches.

Although such comprehensive endeavors are not commonplace at the moment, initiatives such as the SMART Platform are leading the integration of citizen science, community-based participatory research, and systems science in implementing studies for population health surveillance, integrated knowledge translation, and policy interventions [24,67]. The success of

this three-pronged approach will ultimately depend on academic leadership in engaging citizens and communities to leverage data required to address population health issues that are of concern to citizens, communities, policy makers, and researchers.

According to some estimates, there will be 6 billion smartphones in circulation by the end of 2020 [64]. Our lives may be difficult to imagine without these gadgets, and out of all the internet-connected devices that we use, smartphones are the most versatile and pervasive tools. As there is no indication that we will revert to the days without these devices, the real question is, how can we transform smartphones into effective research tools?

The three-pronged approach provides a paradigm-changing purpose to these research tools. However, for this approach to come to fruition, academia needs to play a role in empowering citizens through smartphones, especially because leaders in the technology industry do not have a strong record of safeguarding citizen privacy and confidentiality [68].

The SMART Framework

SMART is an evidence-based framework designed to conduct population surveillance, knowledge translation, and interventions by integrating citizen science, community-based participatory research, and systems science. The framework informs the SMART Platform, which researchers are utilizing to engage with citizen scientists via their smartphones to implement multiple studies with varied study designs (eg, cross-sectional and longitudinal studies to quasi-experimental and community trials) across different jurisdictions within and outside of Canada [69]. SMART Platform provides the flexibility to engage with participants (ie, citizen scientists) in real time to capture rich population health data across jurisdictions.

Studies conducted by the SMART Platform apply mixed-methods approaches to understand not just the incidence and prevalence of health behaviors and outcomes but also where, when, how, and, more importantly *why* health behaviors and outcomes change. Comprehensive population health data collection is achieved by triangulating traditional surveys with ecological momentary assessments deployed via smartphones and mobile sensors. These data can be linked with upstream policy and administrative data and downstream health care utilization data to inform policies across jurisdictions in collaboration with primary stakeholders [24].

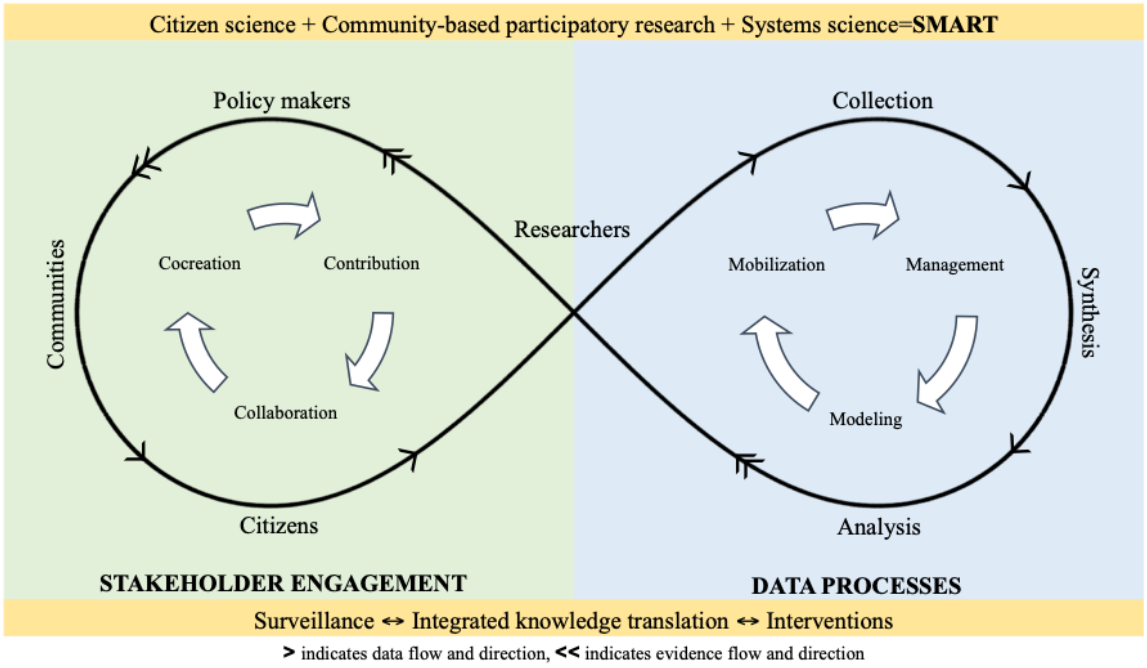
The SMART Framework is rooted in continuous and consistent engagement of citizen scientists by empowering them to cocreate knowledge. The three-pronged approach allows collaborative development of research questions through qualitative systems mapping that enables data collection across the life course to inform dynamic modeling [41]. The evidence generated by this approach is translated back to all relevant stakeholders, including citizen scientists through the same devices that provide the digital data used in all analyses—smartphones (Figure 1).

Figure 1 enumerates the key components of the SMART Framework (ie, stakeholder engagement on the left side and data processes on the right side) as well as the direction of the data (>) and evidence (<<) flow in relationship with the key stakeholders. The engagement of key stakeholders in the SMART Framework is encapsulated by the interaction between citizens, communities, researchers, and policy makers [51].

The contribution, collaboration, and cocreation cycle is central to citizen science and community-based participatory research, and the engagement of stakeholders that materializes within this cycle is essential for qualitative systems mapping to determine research questions.

Researchers have a vital role to play in both empowering citizens to participate in collaboration with communities and to translate evidence back to all the stakeholders, including policy makers.

Figure 1. The SMART Framework: integration of citizen science, community-based participatory research, and systems science via ubiquitous tools.



Thus, researchers occupy a central role that straddles both stakeholder engagement and data processes. The evidence translated will enable evaluation of existing research goals in collaboration with stakeholders to inform future data generation.

Data processes are driven by collection, synthesis, and analysis, with researchers playing the primary role in enabling these processes. Nonetheless, there is a significant potential for citizen involvement in data processes, depending on the level of engagement—contribution, collaboration, or cocreation [51]. These processes are dependent on data management, dynamic modeling, and evidence mobilization to ultimately convert data into evidence and translate it back to all stakeholders through the researchers.

The infinity symbol is representative of the continuous interplay between stakeholders and constant flow of data and evidence within the framework. Digital data leveraged through ubiquitous devices such as smartphones are central to this endless data generation, which is apparent from our dependence on these devices. The stakeholder interplay and data and evidence flow are facilitated by the three-pronged approach, which in turn expedites population health surveillance, integrated knowledge translation, and interventions—processes that are not only interconnected but also interdependent [24]. Overall, the framework enables quick replication of studies in different geographic locations, with an option to centralize or decentralize data collection and storage to follow ethical guidelines for data privacy, anonymity, and security.

An example of this synergy is evident from one of the initiatives informed by the SMART Framework—SMART Indigenous Youth. This initiative is a longitudinal community trial that engages indigenous youth (aged 13-18 years) and educators in rural and remote areas of Canada as citizen scientists [70]. The objective of SMART Indigenous Youth is to improve mental health outcomes among youth by embedding a land-based, culturally appropriate active living intervention into the school curricula.

As part of this initiative, youth and educator citizen scientists are able to cocreate knowledge through a strong partnership between researchers, schools, and rural and remote communities where these citizen scientists reside. The communities are independent jurisdictional units governed by Chief and Council, who also engage with researchers in concert with schools and citizen scientists. This stakeholder engagement accelerates the *big data* collection (surveillance) needed for the dynamic modeling that generates evidence translated to stakeholders (integrated knowledge translation) to evaluate the community trial (intervention).

SMART Indigenous Youth is an ideal example of research informed by the SMART Framework, and this investigation would not be possible without the power of citizen-owned smartphones. These ubiquitous tools are the reason for real-time engagement of citizen scientists in rural and remote areas, thus they serve as tools of equity for a population that is among one of the most vulnerable in Canada [71,72].

Addressing Challenges

Citizen Engagement: Recruitment, Retention, and Compliance

The success of citizen science is dependent on its ability to contribute data, or collaborate and cocreate knowledge with researchers [51]. However, for these actions to transpire, citizen scientists need to be engaged and empowered. Engagement and empowerment are directly connected to recruitment, retention, and compliance, which are key challenges for the success of citizen science. Through the SMART Platform [69], we are

integrating citizen science and community-based participatory research to aid citizen engagement and empowerment. For example, when we recruit youth citizen scientists, we not only engage with schools and school administrators as community partners, but we also form youth citizen scientist councils, where youth interact with researchers and school administrators to drive the research process.

Currently, the SMART Platform is using mobile technology and wearables within and outside of Canada to run simultaneous surveillance and intervention studies with varied designs (eg, cross-sectional and longitudinal studies to quasi-experimental and community trials) among populations across the life course [69,70,73-76]. An important factor in maximizing recruitment, retention, and compliance is to develop strategies that are specific to different studies, cohorts, and demographic groups.

For instance, effective recruitment and retention of older cohorts (aged >65 years) requires continuous engagement and in-person deployment of mobile apps to minimize social isolation and potential technology anxiety. With respect to marginalized populations such as indigenous youth, apart from in-person group deployment through peer-to-peer interaction, incentives such as free data plans play a key role in the success of mobile health interventions.

Finally, to improve engagement from a mobile technology and methodological perspective, it is crucial to understand how mobile interfaces can be made more intuitive to reduce the burden of traditional surveys instruments. Deploying and testing ecological momentary assessments and developing replicable methods for objective data derivation are strategies used by the SMART Platform to reduce burden and increase overall compliance [69,77].

Moreover, the Platform is devised to trigger questions based on time, location, and movement to understand citizen science compliance. The challenges of citizen engagement need to be addressed by a combination of logistical, technological, and methodological solutions—before, during, and after data collection—that are facilitated by the three-pronged approach of the SMART Framework.

Legitimization: Resources, Power Balance, and Traditional Conservatism

The SMART Framework's three-pronged approach inherently faces challenges with legitimization, with leveraging resources being the primary obstacle, as this approach requires significant support in terms of personnel training, funding, and time [41]. Furthermore, as computational modeling processes used by researchers lack transparency, the balance of power between researchers, and communities or citizens could be another valid challenge [78].

In addition, traditional conservatism of academic research could be a larger barrier for this three-pronged approach within population health science, where there is risk of reinforcement and perpetuation of the status quo [79]. With transformative change necessitating more than the involvement of multiple stakeholders [80,81], a need exists for continuous evaluation of the three-pronged approach.

Rowbotham et al [14] state that although citizen science has been neglected by population health science, it has the potential to traverse cross-jurisdictional boundaries to provide new insights in solving population health problems. Moreover, in their study exploring the role of citizen science in transforming population health science, the authors also emphasize the role of community-based participatory research in addressing population health issues [14].

Conducting research using the three-pronged approach requires a long-term vision that cannot be restricted to short-term projects. Existing evidence indicates that citizen science projects, especially those that take a systems perspective, require longevity, and although citizen science projects have been able to obtain private and public funding for longitudinal studies through cross-disciplinary funding initiatives, there is an inherent risk of altering the project goals to meet funding requirements [82].

Utilizing the three-pronged SMART Framework, the SMART Platform has evolved into a digital methodological toolkit that can address broad population health issues ranging from the physical inactivity pandemic and youth mental health, to indigenous health and school health policies [69,70,74-76]. In doing so, the Platform has been able to secure funding from multiple private and public sources by aligning its goals with priorities of both funding agencies, and citizens and communities.

This tactic has triggered the interest of interdisciplinary researchers in citizen science–enabled population health studies, which is important for the success of the integration of citizen science, community-based participatory research, and systems science [52]. The integration also allows the transfer of power to citizens and communities, who play an important role in evaluating the implementation of the SMART Platform. In the SMART Framework, researchers empower citizens to play a larger role and challenge traditional conservatism in academia. This ultimately allows population health science to use citizen science as a tool when integrated with community-based participatory research and systems science.

Data Management: Privacy, Security, and Linkages

Ethical considerations in terms of data privacy, security, and anonymity are at the forefront of the SMART Framework. Smartphone-enabled citizen scientist personal data provide sophisticated granularity in terms of potential identification of participants in real time through a slew of sensors such a global positioning system [83].

Protecting privacy and anonymity of citizens requires advanced encryption processes, which are embedded into the SMART Platform [24]. However, before beginning data collection and storing encrypted data, obtaining informed consent is mandatory [77]. As part of the SMART Platform, citizens not only provide informed consent through their smartphones, but they are also able to drop out of studies and request the deletion of their data. Citizens can also review the informed consent through their smartphones at any point during the study.

Beyond the ability to provide and delete data, citizen scientists in the SMART Platform co-own the data and are able to

participate in data visualization, contribute to analysis, and translate knowledge—all aspects that are enabled by the integration of citizen science, community-based participatory research, and systems science. However, citizen science–enabled data need to be linked with administrative, policy, and even health care access and utilization data to address population health issues [77].

Data linkages require digital and logistical infrastructure as well as policy maker support. For citizen science engagement, technological tools that facilitate secure data collection, synthesis, analyses, and dissemination are essential [52]. These data processes require front-end human-computer interfaces such as smartphone apps and back-end interfaces that encapsulate data management, modeling, and mobilization, as articulated within the SMART Framework.

The SMART Platform utilizes sophisticated front-end interfaces, which comprise constantly evolving smartphone apps specific to various projects. This evolution is informed by the contribution of citizen scientists, communities, and policy makers. Although there is a growing movement for open-source back-end data management [84,85], the SMART Platform functions on a closed back-end database. Nevertheless, the data that are processed as part of this closed back-end are made open to public and interdisciplinary researchers after applying rigorous protocols of data anonymization. These efforts are a part of the strategy to increase interoperability through data linkages with other data sources that maximize the potential benefits of citizen science–enabled research [56,57,86-91].

The ultimate success of such data linkages is dependent on policy maker support, an integral part of SMART Framework to translate data into evidence. The SMART Platform aims to implement this approach by linking upstream behavioral citizen science data with downstream health care utilization data in collaboration with policy makers at multiple levels (local, provincial, and federal).

Internet Inequity: Sociodemographics, Global Gaps, and Policy Initiatives

Of all the challenges that the integration of citizen science, community-based participatory research, and systems science has to overcome, the most systemic barrier is internet inequity. Internet inequity could be defined as differential access to the internet based on wealth of a country (high-, low- or middle-income), geographic region (urban, rural, or remote), socioeconomic status, gender, age, or ethnicity [12,92-94].

If the goal of the three-pronged approach is to address global population health inequities through citizen-owned devices, this integration needs to take into account the fact that there is potential for widening existing health disparities through digital divides that exclude vulnerable groups [12].

Large-scale surveys have shown significant differences in access to the internet across socioeconomic status [92,93]. Buchi et al [92] conducted structural modeling using data from representative surveys from 5 countries (New Zealand, Sweden, Switzerland, United Kingdom, and United States) to show that sociodemographics independently account for 50% of variance in usage, with age as the strongest predictor.

The Pew Research Centre's Internet Project conducted a study to explore how smartphone dependence (ie, when one's only means of accessing the internet is via a smartphone) and smartphone use differ between key demographic groups in the United States. Results showed that minority groups and younger, lower income, and less educated users are more likely to be dependent on smartphones. These findings suggest that smartphones can symbolize both equity and inequity depending on access to the internet [93].

The digital divide is a complex phenomenon as depicted in the study by Hilbert [94], who investigated nationally installed bandwidth potential of 172 countries from 1984 to 2014. The results indicated that internet bandwidth divide between high- and low-income countries first increased during the period of study and then decreased to historic lows between 2012 and 2014. Although there is no clear pattern in terms of the bandwidth divide across high- and low-income countries, in general, there are apparent links between bandwidth divide and income divide [94].

Although more individuals have access to global bandwidth than ever before [94], some countries are more digitally progressive and can provide the road map for better access to the internet [95]. With the United Nations declaring that access to the internet is a human right [96] and with a significant proportion (>75%) of people in 26 countries agreeing with this assessment [97], policy makers have a moral responsibility to address internet inequity.

Countries such as Canada are tackling internet inequity by allocating resources and setting national targets such as providing high speed internet to 100% of homes and businesses by 2030. This strategy is especially important in reducing the urban-rural divide in access to the internet [98].

Although it is beyond the purview of academics to address structural and systemic issues related to internet inequity, as part of the SMART Platform, researchers are engaging with policy makers, communities, and citizens to develop bottom-up approaches for providing internet access to all participating citizen scientists. For example, to provide youth citizen scientists in rural and remote areas with internet access, free data packages are being negotiated as incentives, and schools are increasing internet access to youth during and after school hours.

Conclusions

Citizen science, community-based participatory research, and systems science need to be integrated in addressing global population health problems. However, for this integration to materialize, it is necessary to repurpose citizen-owned ubiquitous communication devices (ie, smartphones) that have revolutionized the ability to sense, share, and link big data. Smartphones have the reach to not only create equity by empowering disenfranchised citizens, but smartphone-based apps also have the capacity to source big data to inform policies through the voice of the citizens. SMART is an evidenced-based framework that integrates citizen science, community-based participatory research, and systems science through ubiquitous

tools by addressing challenges such as citizen engagement, data management, and internet inequity to legitimize this integration.

Acknowledgments

This study was funded by the Canadian Institutes of Health Research's Project Grant (#153226) and the Saskatchewan Health Research Foundation's Establishment Grant (3779). The author acknowledges the support of Ms Jasmin Bhawra, University of Waterloo, to the Digital Epidemiology and Population Health Laboratory and the SMART Platform. Ms Bhawra also contributed to the creation of the SMART Framework diagram.

Conflicts of Interest

None declared.

References

1. Borisch B. Public health in times of austerity. J Public Health Policy 2014 May;35(2):249-257. [doi: 10.1057/jphp.2014.7] [Medline: 24751785]
2. de Vogli R, Gimeno D. The G20 and the three global crises: what prospects for global health? J Epidemiol Community Health 2010 Feb;64(2):99-100. [doi: 10.1136/jech.2009.094789] [Medline: 20056960]
3. Garcia KK, Gostin LO. One health, one world—the intersecting legal regimes of trade, climate change, food security, humanitarian crises, and migration. Laws 2012 Apr 4;1(1):4-38. [doi: 10.3390/laws1010004]
4. Russell MC, Figley CR. Investigating recurrent generational wartime behavioral health crises: part 2 of a preliminary analysis. Psychol Inj Law 2015 May 23;8(2):132-152. [doi: 10.1007/s12207-015-9223-5]
5. Nickerson JW. Global health on the edge--the humanitarian tipping point. Can J Public Health 2013 Mar 6;104(3):e267-e269. [doi: 10.17269/cjph.104.3714] [Medline: 23823894]
6. Mesmar S, Talhouk R, Akik C, Olivier P, Elhajj IH, Elbassuoni S, et al. The impact of digital technology on health of populations affected by humanitarian crises: recent innovations and current gaps. J Public Health Policy 2016 Nov;37(Suppl 2):167-200. [doi: 10.1057/s41271-016-0040-1] [Medline: 27899794]
7. Graham GN, Ostrowski M, Sabina AB. Population health-based approaches to utilizing digital technology: a strategy for equity. J Public Health Policy 2016 Nov;37(Suppl 2):154-166. [doi: 10.1057/s41271-016-0012-5] [Medline: 27899799]
8. Chauvin J, Rispel L. Digital technology, population health, and health equity. J Public Health Policy 2016 Nov;37(Suppl 2):145-153. [doi: 10.1057/s41271-016-0041-0] [Medline: 27899801]
9. Government of Canada. 2019. What is the Population Health Approach? URL: https://www.canada.ca/en/public-health/services/health-promotion/population-health/population-health-approach.html [accessed 2019-07-02]
10. Kindig D, Stoddart G. What is population health? Am J Public Health 2003 Mar;93(3):380-383. [doi: 10.2105/ajph.93.3.380] [Medline: 12604476]
11. Andermann A, CLEAR Collaboration. Taking action on the social determinants of health in clinical practice: a framework for health professionals. Can Med Assoc J 2016 Dec 6;188(17-18):E474-E483 [FREE Full text] [doi: 10.1503/cmaj.160177] [Medline: 27503870]
12. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis 2017;27(2):95-106 [FREE Full text] [doi: 10.18865/ed.27.2.95] [Medline: 28439179]
13. King AC, Winter SJ, Chrisinger BW, Hua J, Banchoff AW. Maximizing the promise of citizen science to advance health and prevent disease. Prev Med 2019 Feb;119:44-47 [FREE Full text] [doi: 10.1016/j.ypmed.2018.12.016] [Medline: 30593793]
14. Rowbotham S, McKinnon M, Leach J, Lamberts R, Hawe P. Does citizen science have the capacity to transform population health science? Crit Public Health 2017 Nov 9;29(1):118-128. [doi: 10.1080/09581596.2017.1395393]
15. Rosas LG, Salvo D, Winter SJ, Cortes D, Rivera J, Rodriguez NM, et al. Harnessing technology and citizen science to support neighborhoods that promote active living in Mexico. J Urban Health 2016 Dec;93(6):953-973 [FREE Full text] [doi: 10.1007/s11524-016-0081-6] [Medline: 27752825]
16. de Cocker K, Chastin SF, de Bourdeaudhuij I, Imbo I, Stragier J, Cardon G. Citizen science to communicate about public health messages: the reach of a playful online survey on sitting time and physical activity. Health Commun 2019 Jun;34(7):720-725. [doi: 10.1080/10410236.2018.1433955] [Medline: 29412005]
17. Rosenthal MS. What we might accomplish by engaging in our local communities. Prev Med Rep 2015;2:13-14 [FREE Full text] [doi: 10.1016/j.pmedr.2014.11.007] [Medline: 26844045]
18. Thompson B, Molina Y, Viswanath K, Warnecke R, Prelip ML. Strategies to empower communities to reduce health disparities. Health Aff (Millwood) 2016 Aug 1;35(8):1424-1428 [FREE Full text] [doi: 10.1377/hlthaff.2015.1364] [Medline: 27503967]

19. Kuo AK, Summers NM, Vohra S, Kahn RS, Bibbins-Domingo K. The promise of precision population health: reducing health disparities through a community partnership framework. Adv Pediatr 2019 Aug;66:1-13. [doi: 10.1016/j.yapd.2019.03.002] [Medline: 31230687]
20. Cacari-Stone L, Wallerstein N, Garcia AP, Minkler M. The promise of community-based participatory research for health equity: a conceptual model for bridging evidence with policy. Am J Public Health 2014 Sep;104(9):1615-1623. [doi: 10.2105/AJPH.2014.301961] [Medline: 25033119]
21. BeLue R, Carmack C, Myers KR, Weinreb-Welch L, Lengerich EJ. Systems thinking tools as applied to community-based participatory research: a case study. Health Educ Behav 2012 Dec;39(6):745-751. [doi: 10.1177/1090198111430708] [Medline: 22467637]
22. Friel S, Pescud M, Malbon E, Lee A, Carter R, Greenfield J, et al. Using systems science to understand the determinants of inequities in healthy eating. PLoS One 2017;12(11):e0188872 [FREE Full text] [doi: 10.1371/journal.pone.0188872] [Medline: 29190662]
23. Silvertown J. A new dawn for citizen science. Trends Ecol Evol 2009 Sep;24(9):467-471. [doi: 10.1016/j.tree.2009.03.017] [Medline: 19586682]
24. Katapally TR, Bhawra J, Leatherdale ST, Ferguson L, Longo J, Rainham D, et al. The SMART study, a mobile health and citizen science methodological platform for active living surveillance, integrated knowledge translation, and policy interventions: longitudinal study. JMIR Public Health Surveill 2018 Mar 27;4(1):e31 [FREE Full text] [doi: 10.2196/publichealth.8953] [Medline: 29588267]
25. Dickinson JL, Zuckerberg B, Bonter DN. Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 2010 Dec;41(1):149-172. [doi: 10.1146/annurev-ecolsys-102209-144636]
26. Theobald EJ, Ettinger AK, Burgess HK, DeBey L, Schmidt N, Froehlich H, et al. Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. Biol Cons 2015 Jan;181:236-244. [doi: 10.1016/j.biocon.2014.10.021]
27. Bonney R, Cooper C, Ballard H. The theory and practice of citizen science: launching a new journal. Citiz Sci: Theory Practice 2016 May 20;1(1):1. [doi: 10.5334/cstp.65]
28. Farmer RD, Lawrenson R. Lecture Notes: Epidemiology and Public Health Medicine. Massachusetts: Wiley-Blackwell; 2004.
29. Israel BA, Coombe CM, Cheezum RR, Schulz AJ, McGranaghan RJ, Lichtenstein R, et al. Community-based participatory research: a capacity-building approach for policy advocacy aimed at eliminating health disparities. Am J Public Health 2010 Nov;100(11):2094-2102. [doi: 10.2105/AJPH.2009.170506] [Medline: 20864728]
30. Carroll-Scott A, Toy P, Wyn R, Zane JI, Wallace SP. Results from the data & democracy initiative to enhance community-based organization data and research capacity. Am J Public Health 2012 Jul;102(7):1384-1391 [FREE Full text] [doi: 10.2105/AJPH.2011.300457] [Medline: 22594748]
31. Freudenberg N, Tsui E. Evidence, power, and policy change in community-based participatory research. Am J Public Health 2014 Jan;104(1):11-14. [doi: 10.2105/AJPH.2013.301471] [Medline: 24228677]
32. Minkler M, Vasquez VB, Chang C, Miller J, Rubin V, Blackwell AG, et al. University of Washington. 2008. Promoting Healthy Public Policy through Community-Based Participatory Research: Ten Case Studies URL: https://depts.washington.edu/ccph/pdf_files/CBPR_final.pdf [accessed 2019-08-21]
33. Minkler M, Vásquez VB, Tajik M, Petersen D. Promoting environmental justice through community-based participatory research: the role of community and partnership capacity. Health Educ Behav 2008 Feb;35(1):119-137. [doi: 10.1177/1090198106287692] [Medline: 16861594]
34. Vásquez VB, Minkler M, Shepard P. Promoting environmental health policy through community based participatory research: a case study from Harlem, New York. J Urban Health 2006 Jan;83(1):101-110 [FREE Full text] [doi: 10.1007/s11524-005-9010-9] [Medline: 16736358]
35. Vásquez VB, Lanza D, Hennessey-Lavery S, Facente S, Halpin HA, Minkler M. Addressing food security through public policy action in a community-based participatory research partnership. Health Promot Pract 2007 Oct;8(4):342-349. [doi: 10.1177/1524839906298501] [Medline: 17728199]
36. Petersen DM, Minkler M, Vásquez VB, Kegler MC, Malcoe LH, Whitecrow S. Using community-based participatory research to shape policy and prevent lead exposure among native American children. Prog Community Health Partnersh 2007;1(3):249-256. [doi: 10.1353/cpr.2007.0028] [Medline: 20208287]
37. Tajik M, Minkler M. Environmental justice research and action: a case study in political economy and community-academic collaboration. Int Q Community Health Educ 2006;26(3):213-231. [doi: 10.2190/IQ.26.3.b] [Medline: 17827092]
38. Gregory AJ, Atkins JP. Community operational research and citizen science: two icons in need of each other? Eur J Oper Res 2018 Aug;268(3):1111-1124. [doi: 10.1016/j.ejor.2017.07.037]
39. Hinckson E, Schneider M, Winter SJ, Stone E, Puhan M, Stathi A, et al. Citizen science applied to building healthier community environments: advancing the field through shared construct and measurement development. Int J Behav Nutr Phys Act 2017 Sep 29;14(1):133 [FREE Full text] [doi: 10.1186/s12966-017-0588-6] [Medline: 28962580]
40. Gray S, Jordan R, Crall A, Newman G, Hmelo-Silver C, Huang J, et al. Combining participatory modelling and citizen science to support volunteer conservation action. Biol Cons 2017 Apr;208:76-86. [doi: 10.1016/j.biocon.2016.07.037]

41. Frerichs L, Lich KH, Dave G, Corbie-Smith G. Integrating systems science and community-based participatory research to achieve health equity. Am J Public Health 2016 Feb;106(2):215-222. [doi: 10.2105/AJPH.2015.302944] [Medline: 26691110]
42. Mabry PL, Olster DH, Morgan GD, Abrams DB. Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. Am J Prev Med 2008 Aug;35(2 Suppl):S211-S224 [FREE Full text] [doi: 10.1016/j.amepre.2008.05.018] [Medline: 18619402]
43. Hammond RA. Complex systems modeling for obesity research. Prev Chronic Dis 2009 Jul;6(3):A97 [FREE Full text] [Medline: 19527598]
44. Leischow SJ, Best A, Trochim WM, Clark PI, Gallagher RS, Marcus SE, et al. Systems thinking to improve the public's health. Am J Prev Med 2008 Aug;35(2 Suppl):S196-S203 [FREE Full text] [doi: 10.1016/j.amepre.2008.05.014] [Medline: 18619400]
45. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health 2012 Apr;33:357-376 [FREE Full text] [doi: 10.1146/annurev-publhealth-031210-101222] [Medline: 22224885]
46. El-Sayed AM, Galea S. Systems Science and Population Health. New York: Oxford University Press; 2017.
47. Rosenhead J, Mingers J. Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty and Conflict. Second Edition. Chichester, UK: Wiley; 2001.
48. Midgley G. Systemic Intervention: Philosophy, Methodology, and Practice. New York: Kluwer Academic/Plenum Publishers; 2000.
49. Gillen EM, Lich KH, Yeatts KB, Hernandez ML, Smith TW, Lewis MA. Social ecology of asthma: engaging stakeholders in integrating health behavior theories and practice-based evidence through systems mapping. Health Educ Behav 2014 Feb;41(1):63-77. [doi: 10.1177/1090198113486804] [Medline: 23709516]
50. Kitson NA, Price M, Lau FY, Showler G. Developing a medication communication framework across continuums of care using the Circle of Care Modeling approach. BMC Health Serv Res 2013 Oct 17;13:418 [FREE Full text] [doi: 10.1186/1472-6963-13-418] [Medline: 24134454]
51. Bonney R, Cooper CB, Dickinson JL, Kelling S, Phillips T, Rosenberg K, et al. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioSci 2009;59(11):977-984. [doi: 10.1525/bio.2009.59.11.9]
52. Crain R, Cooper C, Dickinson JL. Citizen science: a tool for integrating studies of human and natural systems. Annu Rev Environ Resour 2014 Oct 17;39(1):641-665. [doi: 10.1146/annurev-environ-030713-154609]
53. Pryor A, Carpenter C, Townsend M. Outdoor education and bush adventure therapy: a socio-ecological approach to health and wellbeing. Aust J Outdoor Environ Educ 2005 Apr 1;9(1):3-13. [doi: 10.1007/BF03400807]
54. Maller C, Townsend M, Pryor A, Brown P, St Leger L. Healthy nature healthy people: 'contact with nature' as an upstream health promotion intervention for populations. Health Promot Int 2006 Mar;21(1):45-54. [doi: 10.1093/heapro/dai032] [Medline: 16373379]
55. Robinson T. Applying the socio-ecological model to improving fruit and vegetable intake among low-income African Americans. J Community Health 2008 Jul 2;33(6):395-406. [doi: 10.1007/s10900-008-9109-5] [Medline: 18594953]
56. Fink D, Damoulas T, Dave J. Adaptive Spatio-Temporal Exploratory Models: Hemisphere-Wide Species Distributions From Massively Crowdsourced eBird Data. In: Proceedings of the Twenty-seventh AAAI Conference on Artificial Intelligence. 2013 Presented at: AAAI'13; July 14–18, 2013; Bellevue, Washington, USA p. 1284-1290 URL: https://pdfs.semanticscholar.org/f885/6ff340a3e029041377bfe5db0db73659d8d6.pdf
57. Fink D, Hochachka WM, Zuckerberg B, Winkler DW, Shaby B, Munson MA, et al. Spatiotemporal exploratory models for broad-scale survey data. Ecol Appl 2010 Dec;20(8):2131-2147. [doi: 10.1890/09-1340.1] [Medline: 21265447]
58. Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong WK, Kelling S. Data-intensive science applied to broad-scale citizen science. Trends Ecol Evol 2012 Feb;27(2):130-137. [doi: 10.1016/j.tree.2011.11.006] [Medline: 22192976]
59. Xue Y, Dilkina B, Damoulas T, Fink D, Gomes CP, Kelling S. Improving Your Chances: Boosting Citizen Science Discovery. In: Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing. 2013 Presented at: HCOMP'13; November 7-9, 2013; Palm Springs, California, USA p. 198-206 URL: https://www.cs.cornell.edu/~yexiang/publications/hcomp13_birdassist/xue.pdf
60. Hufford M, Taylor B. Edgework in boundary crossings: assessing foundations for public ecology in the Appalachian region. In: Craynon JR, editor. Environmental Considerations in Energy Production. Washington, DC: Mining Development Services; 2013:99-110.
61. Tidball KG, Krasny ME. What role for citizen science in disaster conflict. In: Dickinson JL, Bonney R, Fitzpatrick JW, editors. Citizen Science: Public Participation in Environmental Research. Ithaca, New York: Cornell University Press; 2012:226-234.
62. Osgood ND. Frontiers in health modeling. In: El-Sayed AM, Galea S, editors. Systems Science and Population Health. Oxford, United Kingdom: Oxford University Press; 2017:191-216.
63. Wald DM, Longo J, Dobell AR. Design principles for engaging and retaining virtual citizen scientists. Conserv Biol 2016 Jun;30(3):562-570. [doi: 10.1111/cobi.12627] [Medline: 26365126]

64. Lunden I. Techcrunch. 2015. 6.1B Smartphone Users Globally By 2020, Overtaking Basic Fixed Phone Subscriptions URL: http://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/ [accessed 2019-02-24]
65. Catalyst: Digital Marketing Services to Fuel Results. 2015. With Growth Comes Change: The Evolving Mobile Landscape in 2015 URL: http://catalyst.ca/2015-canadian-smartphone-market/ [accessed 2019-02-24]
66. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
67. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. URL: https://www.smartstudysask.com [accessed 2019-07-12]
68. Manjoo F. The New York Times. 2017. Tech's Frightful Five: They've Got Us URL: https://www.nytimes.com/2017/05/10/technology/techs-frightful-five-theyve-got-us.html [accessed 2019-03-09]
69. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. The SMART Platform URL: https://www.smartstudysask.com/about [accessed 2019-07-12]
70. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. SMART Indigenous Youth URL: https://www.smartstudysask.com/smart-first-nations-youth [accessed 2019-07-12]
71. Métis Nation of Ontario. 2013. Canada's Most Vulnerable: Improving Health Care for First Nations, Inuit, and Métis Seniors URL: http://www.metisnation.org/media/422632/senior_ab_report_2013_en_final.pdf [accessed 2019-07-04]
72. O'Donnell V, Wallace S. Statistics Canada. 2011. First Nations, Métis and Inuit Women URL: https://www150.statcan.gc.ca/n1/pub/89-503-x/2010001/article/11442-eng.htm [accessed 2019-07-04]
73. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. SMART Adults URL: https://www.smartstudysask.com/copy-of-smart-youth [accessed 2019-07-12]
74. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. SMART Youth URL: https://www.smartstudysask.com/smart-youth [accessed 2019-07-12]
75. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. SMART India URL: https://www.smartstudysask.com/smart-india [accessed 2019-07-12]
76. Katapally T. SMART: A Mobile Health and Citizen Science Platform. 2019. SMART Response URL: https://www.smartstudysask.com/copy-of-smart-adults [accessed 2019-07-12]
77. Katapally TR, Chu LM. Methodology to derive objective screen-state from smartphones: a SMART platform study. Int J Environ Res Public Health 2019 Jun 27;16(13):E2275 [FREE Full text] [doi: 10.3390/ijerph16132275] [Medline: 31252617]
78. Martinez-Moyano IJ. Documentation for model transparency. Syst Dyn Rev 2012 May 22;28(2):199-208. [doi: 10.1002/sdr.1471]
79. Cooper C. Citizen Science: How Ordinary People are Changing the Face of Discovery. New York: The Overlook Press; 2016.
80. Scott WR. Approaching adulthood: the maturing of institutional theory. Theor Soc 2008 Jun 11;37(5):427-442. [doi: 10.1007/s11186-008-9067-z]
81. Weick KE, Quinn RE. Organizational change and development. Annu Rev Psychol 1999;50:361-386. [doi: 10.1146/annurev.psych.50.1.361] [Medline: 15012461]
82. Freitag A, Pfeffer MJ. Process, not product: investigating recommendations for improving citizen science 'success'. PLoS One 2013;8(5):e64079 [FREE Full text] [doi: 10.1371/journal.pone.0064079] [Medline: 23691154]
83. Fuller J, Hermeston W, Passey M, Fallon T, Muyambi K. Acceptability of participatory social network analysis for problem-solving in Australian Aboriginal health service partnerships. BMC Health Serv Res 2012 Jun 10;12:152 [FREE Full text] [doi: 10.1186/1472-6963-12-152] [Medline: 22682504]
84. Alabri A, Hunter J. Enhancing the Quality and Trust of Citizen Science Data. In: Proeedings of the Sixth International Conference on e-Science. 2011 Jan 20 Presented at: eScience'11; December 7-10, 2010; Brisbane, QLD, Australia URL: https://ieeexplore.ieee.org/document/5693902 [doi: 10.1109/eScience.2010.33]
85. Hunter J, Alabri A, van Ingen C. Assessing the quality and trustworthiness of citizen science data. Concurr Comp-Pract E 2012 Sep 13;25(4):454-466. [doi: 10.1002/cpe.2923]
86. Tulloch AI, Possingham HP, Joseph LN, Szabo JK, Martin TG. Realising the full potential of citizen science monitoring programs. Biol Conserv 2013 Sep;165:128-138. [doi: 10.1016/j.biocon.2013.05.025]
87. Bonter DN, Cooper CB. Data validation in citizen science: a case study from project FeederWatch. Front Ecol Environ 2012 Aug;10(6):305-307. [doi: 10.1890/110273]
88. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. Front Ecol Environ 2013 Apr;11(3):156-162. [doi: 10.1890/120103]
89. Szabo JK, Vesk PA, Baxter PW, Possingham HP. Regional avian species declines estimated from volunteer-collected long-term data using list length analysis. Ecol Appl 2010 Dec;20(8):2157-2169. [doi: 10.1890/09-0877.1]
90. Worthington JP, Silvertown J, Cook L, Cameron RA, Dodd ME, Greenwood RM, et al. Evolution MegaLab: a case study in citizen science methods. Methods Ecol Evol 2012;3(2):3030-3039. [doi: 10.1111/j.2041-210X.2011.00164.x]

91. Wiggins A, Bonney R, Graham E, Henderson S, Kelling S, LeBuhn G, et al. DataONE. 2013. Data Management Guide for Public Participation in Scientific Research URL: https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf [accessed 2019-07-12]
92. Büchi M, Just N, Latzer M. Modeling the second-level digital divide: A five-country study of social differences in Internet use. New Media & Society 2016 Jul 10;18(11):2703-2722. [doi: 10.1177/1461444815604154]
93. Tsetsi E, Rains SA. Smartphone Internet access and use: Extending the digital divide and usage gap. Mobile Media & Communication 2017 Jun 13;5(3):239-255. [doi: 10.1177/2050157917708329]
94. Hilbert M. The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014. Telecommunications Policy 2016 Jun;40(6):567-581. [doi: 10.1016/j.telpol.2016.01.006]
95. Galloway L. BBC. 2019. Five Countries on the Frontline of Tech URL: http://www.bbc.com/travel/story/20190630-five-countries-on-the-frontline-of-tech [accessed 2019-07-07]
96. Office of the United Nations High Commissioner for Human Rights. 2011. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom Of Opinion and Expression, Frank La Rue URL: https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf [accessed 2019-07-12]
97. BBC. 2010. Internet Access is 'a Fundamental Right' URL: http://news.bbc.co.uk/2/hi/technology/8548190.stm [accessed 2019-03-08]
98. Budget 2019 - Canada. 2019. Building a Better Canada: Universal High-Speed Internet URL: https://www.budget.gc.ca/2019/docs/nrc/infrastructure-infrastructures-internet-en.html [accessed 2019-07-12]

Edited by G Eysenbach; submitted 18.03.19; peer-reviewed by D Wald, O Leal Neto; comments to author 28.04.19; revised version received 20.07.19; accepted 28.07.19; published 30.08.19.

Please cite as:
Katapally TR
The SMART Framework: Integration of Citizen Science, Community-Based Participatory Research, and Systems Science for Population Health Science in the Digital Age
JMIR Mhealth Uhealth 2019;7(8):e14056
URL: http://mhealth.jmir.org/2019/8/e14056/
doi:10.2196/14056
PMID:31471963

Review

Using Telemedicine to Diagnose Surgical Site Infections in Low- and Middle-Income Countries: Systematic Review

Charlotte E J Sandberg[1], BSc; Stephen R Knight[2], BSc, MBChB; Ahmad Uzair Qureshi[3], PhD, FCPS, MRCS, PGDME, MCPS; Samir Pathak[4,5,6], BSc, MD, MSc, MBChB

[1]Faculty of Health Sciences, University of Bristol, Bristol, United Kingdom

[2]Centre for Informatics, Usher Institute, Edinburgh, United Kingdom

[3]Department of Surgery, Services Institute of Medical Sciences, Lahore, Pakistan

[4]Bristol Centre for Surgical Research, Population Health Sciences, Bristol Medical School, Bristol, United Kingdom

[5]NIHR Bristol Biomedical Research Centre, Bristol, United Kingdom

[6]University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

Corresponding Author:
Charlotte E J Sandberg, BSc
Faculty of Health Sciences
University of Bristol
First Floor
5 Tyndall Avenue
Bristol, BS8 1UD
United Kingdom
Phone: 44 117 928 9000
Email: cs14961@bristol.ac.uk

Abstract

Background: A high burden of preventable morbidity and mortality due to surgical site infections (SSIs) occurs in low- and middle-income countries (LMICs), and most of these SSIs occur following discharge. There is a high loss to follow-up due to a wide geographical spread of patients, and cost of travel can result in delayed and missed diagnoses.

Objective: This review analyzes the literature surrounding the use of telemedicine and assesses the feasibility of using mobile phone technology to both diagnose SSIs remotely in LMICs and to overcome social barriers.

Methods: A literature search was performed using Medline, Embase, CINAHL, PubMed, Web of Science, the Cochrane Central Register of Controlled Trials and Google Scholar. Included were English language papers reporting the use of telemedicine for detecting SSIs in comparison to the current practice of direct clinical diagnosis. Papers were excluded if infections were not due to surgical wounds, or if SSIs were not validated with in-person diagnosis. The primary outcome of this review was to review the feasibility of telemedicine for remote SSI detection.

Results: A total of 404 articles were screened and three studies were identified that reported on 2082 patients across three countries. All studies assessed the accuracy of remote diagnosis of SSIs using predetermined telephone questionnaires. In total, 44 SSIs were accurately detected using telemedicine and an additional 14 were picked up on clinical follow-up.

Conclusions: The use of telemedicine has shown to be a feasible method in remote diagnosis of SSIs. Telemedicine is a useful adjunct for clinical practice in LMICs to decrease loss to postsurgical follow-up.

(JMIR Mhealth Uhealth 2019;7(8):e13309) doi:10.2196/13309

KEYWORDS

surgical site infection; wound infection; developing country; low- and middle-income countries; telemedicine; postoperative; follow-up

Introduction

Surgical site infections (SSIs) cause preventable morbidity and mortality. The etiology of SSIs is multifactorial and is due to various perioperative factors such as: antibiotic use, cleanliness of surgical equipment and postoperative wound care [1]. The greatest burden is seen in low- and middle-income countries

(LMICs) [2], and therefore investigating factors to reduce the incidence of SSIs should be a clinical priority.

Primary prevention of infection plays a significant role in decreasing morbidity and mortality; however, complete eradication of SSIs is not possible. Therefore, to minimize the morbidity of SSIs, timely management is required. This proves challenging in LMICs, as most SSIs are diagnosed after discharge [3]. The surgical population experiences a high loss to follow-up, ranging between 32-75% [3,4,5], which increases the proportion of infections that go undetected and untreated. The wide geographic spread, as a result of poor healthcare infrastructure, is a significant factor in the high loss to follow-up. Specifically, patients report long journey times, with high associated costs and loss of income as the key reasons behind nonattendance [6].

While this highlights a wider problem surrounding access to healthcare facilities in LMICs, one practical solution to improve patient follow-up after discharge is the use of telemedicine. This is possible because LMICs have seen exponential expansion of mobile phone infrastructure in recent years, allowing for greater implementation of telemedicine [7].

Previous studies reporting the use of telemedicine in LMICs have described improvement in cancer clinic attendance rates [8] and long-term surgical follow-up of cleft palate surgeries [9]. Currently, the only systematic review available that details the use of telemedicine for surgical follow-up for SSI detection is based solely within developed countries [10].

A number of studies describe the use of telemedicine in LMICs using telephone questionnaires [3-5]; however, to the authors' knowledge, a review comprehensively summarizing the use of telemedicine within an LMIC setting has yet to be performed. Thus, the authors aimed to review the feasibility of telemedicine for diagnosis of SSIs in LMICs and the associated financial costs.

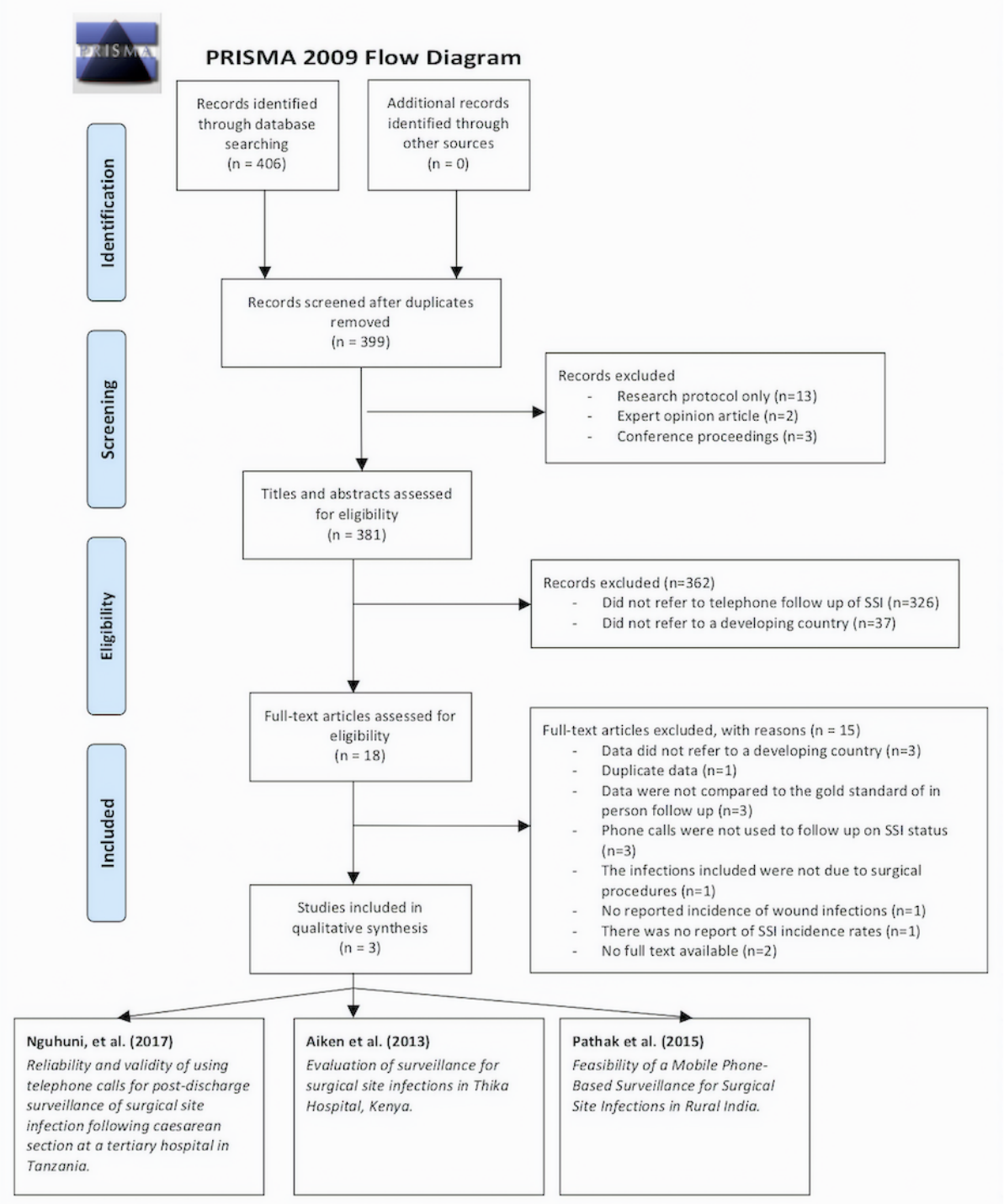
Methods

Overview

The electronic search was conducted according to the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines [11] (Figure 1). The search was conducted independently by two authors (CS and SK) and included the databases Medline, Embase, CINAHL, PubMed, Web of Science, the Cochrane Central Register of Controlled Trials and Google Scholar.

MeSH search terms used were: ["developing country*" OR "LEDC" OR "resource poor setting*" OR "lower economically developed" OR "LMIC" OR "Low* and middle income" OR "low* middle income" OR "less developed" OR "global south" OR "low* income" OR "third world" OR "global health" OR "rural"] AND ["surgical site infection*" OR "surgical wound infection*" OR "SSI" OR "Surgical wound dehiscence" OR "Wound infection*" OR "Complication*" OR "Postsurgical" OR "postoperative complication*" OR "wound healing*"] AND ["phone" OR "telephone" OR "cellphone" OR "text message*" OR "phone call" OR "telephone call" OR "text" OR "telemedicine" OR "smartphone" OR "interviews as topic" OR "telephone interview*"]. Bibliographic references for each article were directly searched to identify additional studies not found in the primary electronic search. The last search date was June 2019.

Figure 1. Preferred reporting items for systematic reviews and meta-analysis flow diagram.



Definitions

LMICs are defined as belonging to the World Health Organization (WHO) classification of a low or low-middle income country [12] at the time of study. Low-income countries are defined as a country with a gross national income (GNI) per capita of 1025 United States Dollars (USD) or less, and low-middle income countries are defined as having a GNI per capita between 1026-4035 USD [12]. Telemedicine is defined as the use of telecommunication devices for remote delivery of medical care [13].

SSIs are defined in line with the Centers for disease control and prevention (CDC) criteria [14], stipulating infection must have occured within 30 days of primary incision and can only involve the skin and soft tissues surrounding the incision site. At least one of any of the following signs of infection must also be present: purulent discharge, positive organism cultures, heat, erythema, local edema or pain.

Inclusion Criteria

Literature reviews and original research published in English in peer-reviewed journals were eligible for inclusion. Studies included reported the use of telephone communication for detection and diagnosis of SSIs, in comparison to the current

practice of direct clinical diagnosis. All operations requiring a skin incision and overnight stay were included. All types of anesthesia were included.

Exclusion Criteria

Animal studies, case reports, grey literature, trial protocols, editorials and expert opinion articles were excluded. Two papers were excluded due to full texts being unavailable. Conference reports and proceedings were excluded due to the high probability of incomplete data. Articles not in the English language were also excluded. Studies where telemedicine was not validated by in-person physician diagnosis were excluded. Similarly, studies that did not report SSI incidence rate or did not report on infections from surgical wounds were excluded. In the instance that multiple studies referred to the same dataset, the publication with the most complete dataset was used.

Outcome Measures

The primary outcome measure was the number of SSIs diagnosed using telephone interviews as a proportion of the total number of SSIs diagnosed. Secondary outcome measures included determining the proportion of patients who were able to be contacted using telephone follow-up, as well as the cost of the intervention itself for the healthcare provider and the patient.

Data Extraction

Data were extracted independently by two researchers (CS and SK) from the included studies using a standardized *pro forma*. All differences in data extraction were highlighted and discussed between the two authors. If a consensus was not reached, SP was consulted and given the deciding vote. The data were categorized as study demographics (ie, geographical location, number of included patients, follow-up time, number of follow-ups, method of telephone use), patient demographics (ie, study population, type of surgical procedures) and infection demographics (ie, SSI diagnosis criteria, SSI incidence).

Outcome data was collected in groups relating to patient characteristics (ie, number of patients with a mobile phone, number of patients participating in the telemedicine intervention, proportion of patients who were successfully followed up with using telemedicine), SSI characteristics (ie, total number of SSIs, proportion of SSIs diagnosed using telemedicine), and economic outcomes relating to the cost of telemedicine intervention.

Quality Assessment

Objective measurement of quality and risk of bias was independently analyzed using the Methodological index for nonrandomized studies (MINORS) criteria [15] by two authors (CS and SK). In addition, the effectiveness and ability of each study to be replicated in a wider setting was assessed using the WHO mobile health evidence and reporting assessment (mERA) checklist [16].

Results

Study Selection

The search criteria retrieved 404 articles, of which seven were duplicates. Following screening, a total of 18 articles were identified for full text review. Three full text articles met the inclusion criteria.

Included studies were performed in Kenya [5], India [4] (low-middle income countries) and Tanzania [3] (low income country). A total of 2082 patients were included, with 172 SSIs reported overall and a mean SSI rate of 8.75%. All studies assessed the use of telephone calls to diagnose SSIs (Table 1).

Table 1. Included study characteristics.

Characteristics	Aiken et al	Nguhuni et al	Pathak et al
Publication year	2013	2017	2015
Geographic location	Kenya	Tanzania	India
Number of included patients	1172	374	536
Study population	All patients undergoing surgical procedure	Obstetric patients	All patients undergoing surgical procedure
Included surgical procedures	Caesarean section, laparotomy, hernia repair, orthopedic lower limb surgery, salpingectomy, cystectomy	Caesarean section only	Hernia repair, colonic and urogenital surgery
Method of telemedicine intervention	Telephone calls with predetermined questionnaire	Telephone calls with predetermined questionnaire	Telephone calls with predetermined questionnaire
Follow-up frequency	2	3	1
Follow up performed	Days 14 & 28	Days 6, 12 & 28	Day 30
Length of follow up	30 days	30 days	30 days
SSI[a] diagnosis criteria	CDC[b] criteria for diagnosis of SSI	CDC criteria for diagnosis of SSI	Local SSI criteria
SSI rate	7.90%	12.0%	6.34%

[a]SSI: surgical site infection

[b]CDC: centers for disease control and prevention

Table 2. Summary of findings.

Findings	Aiken et al	Nguhuni et al	Pathak et al
Patients providing a telephone number, %	>90	84	100
Patients with ≥1 successful telephone follow up who had provided a telephone number, %	Not stated	87	71
Patients contacted by phone and seen in clinic, %	Not stated	73	69
Patients undergoing surgical procedure	89	484	10
Patients with SSIs[a] that participated in telephone screening, %	Not stated	72	Not stated
SSIs detected over phone or clinic review at day 7	Not stated	11/14	Not stated
SSIs detected over phone or by clinical review at day 14	Not stated	7/11	Not stated
SSIs detected over phone or by clinical review at day 30	Not stated	0/0	Not stated
Total SSIs detected over phone or by clinical review	16 / 23	18 / 25	10 / 10

[a]SSI: surgical site infection

Accuracy of Telephone Follow-Up Diagnosis

Of all the cases, 44 SSIs were diagnosed over the phone [3,4,5], and all were confirmed as SSIs on clinical follow-up. However, 14 additional SSIs were not detected over the phone and only diagnosed on clinical follow-up (Table 2).

Number of Patients Able to be Contacted by Phone

A total of 172 SSIs were recorded across the three studies [3,4,5]; however, only 58 of these SSIs occurred in populations who took part in telephone follow-up. The remaining 114 SSIs were diagnosed in follow-up clinics and involved patients not contacted via telemedicine.

Within the three studies, between 84-100% of patients provided a telephone number [3,4,5]; however, 13-29% of these patients were unable to be contacted. Loss to follow-up occurred due to patients not returning to clinic after telemedicine communication. This resulted in an inability to confirm SSI diagnosis.

Cost Analysis

Only one study [3] commented on the cost of intervention. Phone calls were recorded as lasting between 3-5 minutes, and it was estimated that each phone call cost approximately 0.50 USD. As each patient was followed up with on days 7, 14 and 28, this would be equivalent to 1.50 USD per patient (0.23% of Tanzanian gross domestic product [GDP] per capita at the time of calculation). In a reflection of the cost for in-person outpatient appointments, 2008 data supplied by the WHO documents one hospital outpatient appointment in Tanzania as costing a minimum of 1.27 USD [17], which is 2.54 times more expensive than if telephone follow up were to be used.

Quality Assessment of Studies

Each of the studies was assessed using the MINORS criteria [15] to determine their quality. As they were all comparative prospective studies, the maximum score was 24. The individual study scores were between 14-19 (Table 3).

The MINORS criteria assess whether the paper has a clearly stated aim, if it includes all eligible patients, if it collected data according to a preestablished protocol, if there was an appropriate endpoint to the aim of the study, if there was an unbiased assessment of the study endpoint, if it had an appropriate follow-up period, whether it had a loss to follow up less than 5%, if it prospectively calculated the study size, whether or not it utilized an adequate control group, if the study group was managed during the same time period as the control, if there was a baseline equivalence of the two groups and if there is adequate statistical analysis.

When each text was compared against the WHO mERA checklist (Table 4), they scored between eight to nine points out of a possible 16. The mERA checklist requires authors to present the availability of infrastructure to support the intervention and justify its use; to describe how the intervention can integrate into existing health information systems and how this will be delivered; to describe formative research on the intervention; to provide user feedback; to provide a description of how people are informed of the program; to mention barriers to access and costs assessment; and, finally, they must evaluate the limitations of delivery at scale and of adapting the intervention to other populations.

Table 3. Methodological index for nonrandomized studies criteria for each paper.

Criteria	Studies					
	Aiken et al [5]		Nguhuni et al [3]		Pathak et al [4]	
	Score	Description	Score	Description	Score	Description
1[a]	2	—[b]	2	—	2	—
2[c]	2	—	1	Did not state reason some patients were excluded	1	No inclusion criteria stated
3[d]	1	References a pilot study	0	Doesn't report a pilot or protocol	0	No pilot or protocol mentioned
4[e]	2	SSIs[f] diagnosed in reference to the standardized CDC-NHNS[g] risk criteria	2	SSIs diagnosed in reference to the standardized CDC-NHNS risk criteria	0	No criteria for SSI diagnosis explained
5[h]	1	Not stated whether those in clinic were blinded to the telephone outcomes	1	Not stated whether those in clinic were blinded to the telephone outcomes	1	Not stated whether those in clinic were blinded to the telephone outcomes
6[i]	2	30 days	2	30 days	2	30 days
7[j]	0	The loss to follow up was not commented on	1	Loss to follow up was 13%, but 26% of those who had been contacted by phone did not return to clinic	2	No loss to follow up
8[k]	2	Specificity, sensitivity	2	Specificity, sensitivity, positive predictive values, negative predictive values	1	Not recorded
9[l]	2	Compared against clinical diagnosis	2	Compared against clinical diagnosis	2	Compared against clinical diagnosis
10[m]	2	Patients were their own control	2	Patients were their own control	2	Patients were their own control
11[n]	0	No Table 1 referenced	2	—	0	No data on group equivalence mentioned
12[o]	2	—	2	—	1	No statistics mentioned for SSI detection via mobile phone
Total score[p]	18	—	19	—	14	—

[a]Has a clearly stated aim.

[b]Not applicable/no description given.

[c]Includes all eligible patients.

[d]Collected data according to a preestablished protocol.

[e]There was an appropriate endpoint to the aim of the study.

[f]SSI: surgical site infection.

[g]CDC-NHNS: Centers for Disease Control and Prevention-National Healthcare Safety Network.

[h]There was an unbiased assessment of the study endpoint.

[i]Had an appropriate follow-up period.

[j]Had a loss to follow up less than 5%.

[k]Prospectively calculated the study size.

[l]Utilized an adequate control group.

[m]Study group was managed during the same time period as the control.

[n]Baseline equivalence of the study group and control group.

[o]Adequate statistical analysis.

[p]Out of 24.

Table 4. Mobile health evidence and reporting assessment checklist.

Criteria	Aiken et al [5]	Nguhuni et al [3]	Pathak et al [4]
Infrastructure (population level)	No	No	No
Technology platform	Yes	Yes	Yes
Interoperability or HIS[a] context	Yes	Yes	Yes
Intervention delivery	Yes	Yes	Yes
Intervention content	Yes	Yes	Yes
Usability/content testing	No	No	No
User feedback	No	No	No
Access of individual participants	No	No	Yes
Cost assessment	No	Yes	No
Adoption inputs/ program entry	Yes	Yes	Yes
Limitations for delivery at scale	No	No	Yes
Contextual adaptability	No	No	No
Replicability	Yes	Yes	Yes
Data security	No	No	No
Compliance with national guidelines or regulatory statutes	No	No	No
Fidelity of the intervention	Yes	Yes	Yes
Total (out of 16)	7	8	9

[a]HIS: health information systems.

Discussion

Primary Findings

More operative interventions are being undertaken in LMICs, however, higher rates of postoperative complications in LMICs are seen when compared with high income countries. Among these complications, SSIs have been found to be the most common [2]. The aim of this review was to establish the feasibility of using telemedicine to increase SSI detection and reduce the associated complications.

This review demonstrates that using telemedicine is a feasible intervention. All studies reported high telemedicine access and all SSIs diagnosed using telemedicine were found to be accurate. However, telemedicine alone underdiagnosed SSIs, as additional superficial infections were picked up on during in-person clinical review.

Additionally, in terms of cost effectiveness, telemedicine intervention was found to be 2.54 times less expensive than hospital outpatient appointments, using 2008 data supplied by the WHO recording one hospital outpatient appointment in Tanzania as costing a minimum of $1.27 [17].

Furthermore, based on an extrapolation of similar study data that showed a follow-up rate of 54% [18], the use of telemedicine could have detected an additional 8.28 SSIs otherwise lost to follow-up. This amounts to 36.59 USD per disability adjusted life years (DALY), where the DALY is an undiagnosed infection. This is considered a highly cost-effective intervention as per the WHO Choosing interventions that are cost-effective (CHOICE) guidelines [19], since the intervention is less than one third of Tanzania's 2015 GDP per capita (290.66 USD) [20]. However, as only one paper discussed cost-effectiveness, firm conclusions on the cost of telemedicine intervention as a whole cannot be extrapolated from these data alone.

The included studies were able to reach up to 87% of patients for follow-up using telemedicine. This is much higher than the reported follow-up in a comparable Tanzanian study that did not use telemedicine, where only 46% of patients returned in person to clinics [18].

These results also reflect the success of LMIC telemedicine interventions in other medical fields. One study on follow-up of cancer patients in Nigeria found an additional 78.4% of patients completed all follow-up appointments if they were contacted by phone instead of being required to return to the hospital [8]. Telemedicine has also shown good efficacy in increasing patient compliance with returning for hospital follow-up, due to increased interaction with health care providers through text and call reminders [9,21].

This study strengthens the existing evidence for the incorporation of more telemedicine services in LMICs, as a practical way of delivering healthcare in remote settings. In Tanzania, two phone calls at day 7 and day 14 would have picked up over 90% of SSIs [3], evidence that this intervention can be a useful addition to clinicians' practices. However, across the three studies, 25% of SSIs were missed by telemedicine alone. This shows that, currently, the intervention may have a low sensitivity that can limit its application.

Limitations

The most notable limitation of this review was that only three studies were eligible for inclusion in the final analysis. This may have been a result of only considering articles in the English language; however, this is not thought to bias overall results [22,23]. Furthermore, there was high heterogeneity between studies, due to studies being undertaken in different geographical areas, involving different surgical procedures, and in populations without standardized SSI diagnosis criteria between the three studies. As a result, meta-analysis was not considered applicable.

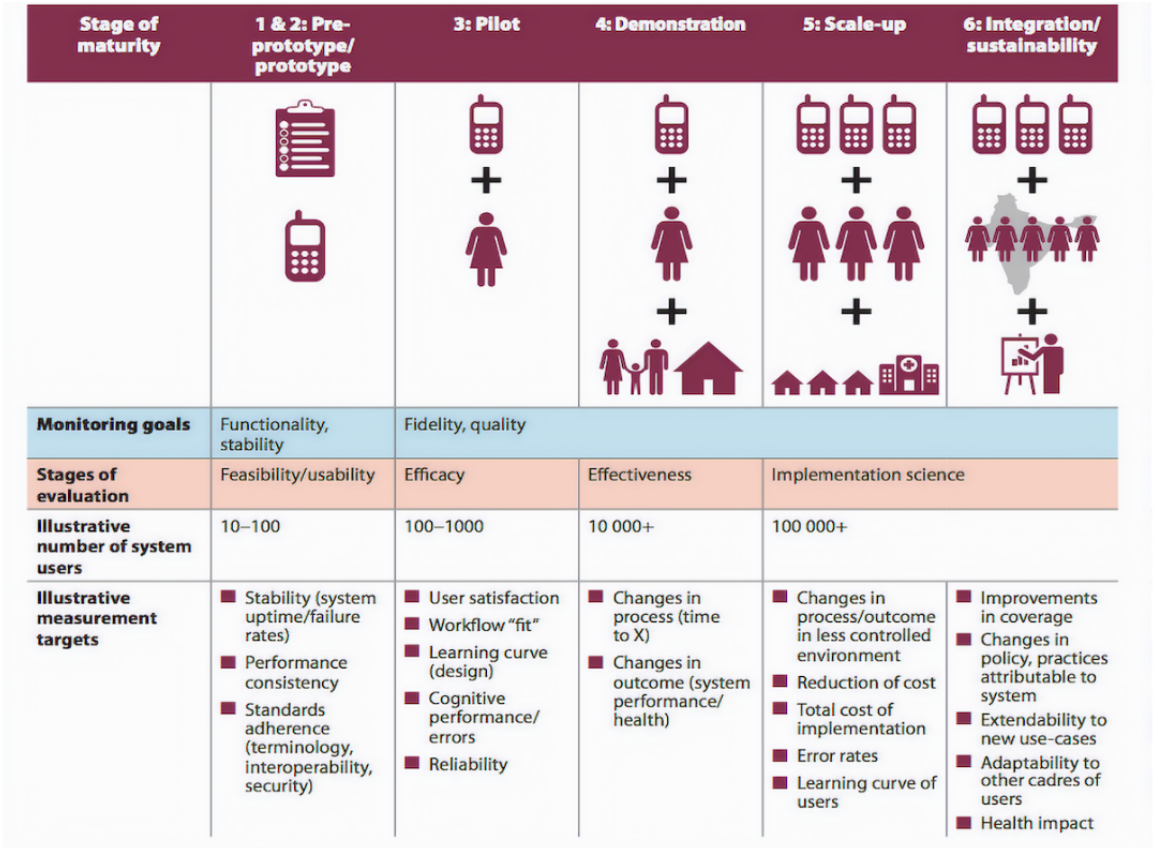
Additionally, within all three papers there is potential reporting bias due to a lag period of 2-7 days between phone diagnoses of SSI and clinical follow up. As SSI status is able to change within this window, this may have affected results. Furthermore, none of the studies confirmed whether diagnosing clinicians were blinded to the telephone results. There is further potential bias with regards to the subjectivity of CDC diagnosis criteria, such as whether a wound is considered erythematous or hot. Although this is explored by Nguhuni et al [3] and Aiken et al [5], Pathak et al [4] do not comment upon this.

A potential consequence of these biases is the low reported incidence of SSIs across the three studies (6.3% [4], 7.9% [5] and 12% [3]), of which 93% were reported as superficial. This is markedly lower than the SSI incidence reported in the GlobalSurg 2 study [24], which found LMICs to report an average SSI rate of 14-23.2%, with 81.1% being superficial. This discrepancy may be due to caesarean sections being the most common procedure represented within the study data, as these tend to be elective operations, with clean wounds and minimal contamination. Furthermore, patients undergoing caesarean section are not representative of the wider surgical population, as they are usually otherwise healthy young women. This limits the generalizability of the data analyzed.

Regarding the included data, while a total of 2719 patients were eligible for inclusion and up to 90% of patients provided telephone numbers [3] across the studies, only 573 patients were recorded as having successful telephone follow-ups with clinical confirmation and were therefore able to be included in the final analysis. Consequently, this may also limit the results. Moreover, telemedicine access is likely to be correlated to socioeconomic status and literacy, which was not commented upon in any of the included studies. It could therefore be argued that while telemedicine addresses the need for access to healthcare, it does so unequally, increasing social inequality.

None of the studies included in this review commented upon subsequent interventions or outcomes due to increased SSI detection, therefore the impact of the intervention is currently unclear. Future work should incorporate this and be undertaken using the WHO guidelines on monitoring and evaluation of digital health interventions [16] (Figure 2). Initially, pilot studies should be used to better measure the efficacy of telemedicine use, then they should be followed up by scaling-up of the intervention to measure the effectiveness of the intervention itself as well as changes in patient outcomes in both trial and less controlled environments.

Figure 2. World Health Organization Mobile health evidence and reporting assessment schematic depiction of the six stages of the intervention maturity life-cycle, from preprototype to national level deployment.



Furthermore, the studies analyzed in this review only considered the use of telephone consultation; however, as mobile data access increases, ease of access to modalities such as image capture and live video streaming will also increase. This may help increase accuracy of diagnosis, as clinicians are able to simultaneously view the surgical site and question the patient. Therefore, ongoing study should not be limited to previously used telemedicine technology but instead focus on the most appropriate in terms of ease of use, accuracy of diagnosis, time and cost.

Conclusion

The use of telephone surveillance has been shown to be a feasible option for follow-up in comparison to in-person diagnosis, with evidence that it is also cost effective. The main advantage of its use is that it can help identify SSIs that would otherwise be unreported, thus increasing the potential prompt management of time sensitive infections. However, currently, the quantitative impact of this intervention has yet to be calculated, so more research is needed to address this. Future studies should also aim to improve the accuracy of diagnosis via telemedicine through strengthening questioning methods and exploring other modalities such as live video streaming. With application of robust telephone screening programs, there is evidence that this has the potential to be a great addition to the tools used to improve the high SSI morbidity and mortality seen in LMICs.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at University Hospitals Bristol National Health Service (NHS) Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Authors' Contributions

CS was responsible for the initial concept, study design, data collection, data interpretation and analysis, as well as drafting the article. SK was involved in the data collection, data interpretation and analysis, as well as the critical revision of the article. AQ was involved in the data analysis and interpretation, and the critical revision of the article. Finally, SP was responsible for the study design, data interpretation and analysis, critical revision of the article, and oversaw supervision of the overall project. All authors reviewed and approved the final version of the manuscript prior to submission.

References

1. Owens CD, Stoessel K. Surgical site infections: epidemiology, microbiology and prevention. J Hosp Infect 2008 Nov;70 Suppl 2:3-10. [doi: 10.1016/S0195-6701(08)60017-1] [Medline: 19022115]
2. Allegranzi B, Bagheri Nejad S, Combescure C, Graafmans W, Attar H, Donaldson L, et al. Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. Lancet 2011 Jan 15;377(9761):228-241. [doi: 10.1016/S0140-6736(10)61458-4] [Medline: 21146207]
3. Nguhuni B, De Nardo P, Gentilotti E, Chaula Z, Damian C, Mencarini P, et al. Reliability and validity of using telephone calls for post-discharge surveillance of surgical site infection following caesarean section at a tertiary hospital in Tanzania. Antimicrob Resist Infect Control 2017;6:43 [FREE Full text] [doi: 10.1186/s13756-017-0205-0] [Medline: 28503302]
4. Pathak A, Sharma S, Sharma M, Mahadik VK, Lundborg CS. Feasibility of a Mobile Phone-Based Surveillance for Surgical Site Infections in Rural India. Telemed J E Health 2015 Mar 06;21(11):946-949 [FREE Full text] [doi: 10.1089/tmj.2014.0199] [Medline: 25748641]
5. Aiken AM, Wanyoro AK, Mwangi J, Mulingwa P, Wanjohi J, Njoroge J, et al. Evaluation of surveillance for surgical site infections in Thika Hospital, Kenya. J Hosp Infect 2013 Feb;83(2):140-145 [FREE Full text] [doi: 10.1016/j.jhin.2012.11.003] [Medline: 23332563]
6. Boam T, Hueschelrath A, Tho L, da Costa TM, McMurrey A, Gollogly J. Managing Soft Tissue Sarcomas in a Developing Health System. World J Surg 2016 Jul;40(7):1542-1549. [doi: 10.1007/s00268-016-3455-8] [Medline: 26920408]
7. World Bank. World development report: Digital Dividends. Washington DC. USA: The World Bank; 2016.
8. Odigie VI, Yusufu LMD, Dawotola DA, Ejagwulu F, Abur P, Mai A, et al. The mobile phone as a tool in improving cancer care in Nigeria. Psychooncology 2012 Mar;21(3):332-335. [doi: 10.1002/pon.1894] [Medline: 22383275]
9. Foong DYJ, Butler DP, Vanna K, Sok Leng T, Gollogly J. A mobile phone initiative to increase return for speech therapy follow-up after cleft palate surgery in the developing world. J Plast Reconstr Aesthet Surg 2012 Sep;65(9):e260-e261. [doi: 10.1016/j.bjps.2012.04.026] [Medline: 22578649]
10. Gunter RL, Chouinard S, Fernandes-Taylor S, Wiseman JT, Clarkson S, Bennett K, et al. Current Use of Telemedicine for Post-Discharge Surgical Care: A Systematic Review. J Am Coll Surg 2016 May;222(5):915-927. [doi: 10.1016/j.jamcollsurg.2016.01.062] [Medline: 27016900]

11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
12. World Bank Data Team. Worldbank.org. New country classifications by income level | The Data Blog The World Bank: The World Bank; 2016:2016-2007.
13. Institute of Medicine (US) Committee on Evaluating Clinical Applications of Telemedicine. In: Field MJ, editor. Telemedicine: A Guide to Assessing Telecommunications in Health Care. Washington, DC: National Academies Press; 1996. URL: https://www.ncbi.nlm.nih.gov/books/NBK45448/ [doi: 10.17226/5296]
14. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control 2008 Jun;36(5):309-332. [doi: 10.1016/j.ajic.2008.03.002] [Medline: 18538699]
15. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. ANZ J Surg 2003 Sep;73(9):712-716. [Medline: 12956787]
16. WHO. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. In: WHO. Geneva: World Health Organization; 2016.
17. WHO. 2010. WHO | Health service delivery costs URL: http://www.who.int/choice/cost-effectiveness/inputs/en/
18. Fehr J, Hatz C, Soka I, Kibatala P, Urassa H, Smith T, et al. Risk factors for surgical site infection in a Tanzanian district hospital: a challenge for the traditional National Nosocomial Infections Surveillance system index. Infect Control Hosp Epidemiol 2006 Dec;27(12):1401-1404. [doi: 10.1086/509855] [Medline: 17152042]
19. Hutubessy R, Chisholm D, Edejer TT. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. Cost Eff Resour Alloc 2003 Dec 19;1(1):8 [FREE Full text] [doi: 10.1186/1478-7547-1-8] [Medline: 14687420]
20. World BNAD. World Bank national accounts data. In: World Bank. Online: The World Bank; 2018.
21. Lin H, Chen W, Luo L, Congdon N, Zhang X, Zhong X, et al. Effectiveness of a short message reminder in increasing compliance with pediatric cataract treatment: a randomized trial. Ophthalmology 2012 Dec;119(12):2463-2470. [doi: 10.1016/j.ophtha.2012.06.046] [Medline: 22921386]
22. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002 Feb;31(1):115-123. [Medline: 11914306]
23. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol 2000 Sep;53(9):964-972. [Medline: 11004423]
24. GlobalSurg Collaborative. Surgical site infection after gastrointestinal surgery in high-income, middle-income, and low-income countries: a prospective, international, multicentre cohort study. Lancet Infect Dis 2018 May;18(5):516-525 [FREE Full text] [doi: 10.1016/S1473-3099(18)30101-4] [Medline: 29452941]

Abbreviations

CDC: centers for disease control and prevention
CHOICE: Choosing interventions that are cost-effective
DALY: Disability adjusted life years
GDP: gross domestic product
GNI: Gross national income
LMIC: Low-middle income country
mERA: Mobile health evidence and reporting assessment
NHS: National Health Service
NIHR: National Institute for Health Research
SSI: surgical site infection
USD: United States dollar
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.01.19; peer-reviewed by F Kaliyadan, A Tsanas, E Caiani; comments to author 28.04.19; revised version received 07.06.19; accepted 17.06.19; published 19.08.19.

Please cite as:
Sandberg CEJ, Knight SR, Qureshi AU, Pathak S
Using Telemedicine to Diagnose Surgical Site Infections in Low- and Middle-Income Countries: Systematic Review
JMIR Mhealth Uhealth 2019;7(8):e13309
URL: http://mhealth.jmir.org/2019/8/e13309/
doi:10.2196/13309
PMID:31429414

Review

Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitations

Alina Trifan[1,2*], PhD; Maryse Oliveira[1,2*], MSc; José Luís Oliveira[1,2], PhD

[1]Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal

[2]Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal

[*]these authors contributed equally

Corresponding Author:
Alina Trifan, PhD
Department of Electronics, Telecommunications and Informatics
University of Aveiro
Campus Universitario de Santiago
Aveiro,
Portugal
Phone: 351 234370500
Email: alina.trifan@ua.pt

Abstract

Background: Technological advancements, together with the decrease in both price and size of a large variety of sensors, has expanded the role and capabilities of regular mobile phones, turning them into powerful yet ubiquitous monitoring systems. At present, smartphones have the potential to continuously collect information about the users, monitor their activities and behaviors in real time, and provide them with feedback and recommendations.

Objective: This systematic review aimed to identify recent scientific studies that explored the passive use of smartphones for generating health- and well-being–related outcomes. In addition, it explores users' engagement and possible challenges in using such self-monitoring systems.

Methods: A systematic review was conducted, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, to identify recent publications that explore the use of smartphones as ubiquitous health monitoring systems. We ran reproducible search queries on PubMed, IEEE Xplore, ACM Digital Library, and Scopus online databases and aimed to find answers to the following questions: (1) What is the study focus of the selected papers? (2) What smartphone sensing technologies and data are used to gather health-related input? (3) How are the developed systems validated? and (4) What are the limitations and challenges when using such sensing systems?

Results: Our bibliographic research returned 7404 unique publications. Of these, 118 met the predefined inclusion criteria, which considered publication dates from 2014 onward, English language, and relevance for the topic of this review. The selected papers highlight that smartphones are already being used in multiple health-related scenarios. Of those, physical activity (29.6%; 35/118) and mental health (27.9; 33/118) are 2 of the most studied applications. Accelerometers (57.7%; 67/118) and global positioning systems (GPS; 40.6%; 48/118) are 2 of the most used sensors in smartphones for collecting data from which the health status or well-being of its users can be inferred.

Conclusions: One relevant outcome of this systematic review is that although smartphones present many advantages for the passive monitoring of users' health and well-being, there is a lack of correlation between smartphone-generated outcomes and clinical knowledge. Moreover, user engagement and motivation are not always modeled as prerequisites, which directly affects user adherence and full validation of such systems.

(JMIR Mhealth Uhealth 2019;7(8):e12649) doi:10.2196/12649

KEYWORDS

smartphone; mobile phone; mhealth; digital health; digital medicine; digital phenotyping; health care; mHealth; self-management; systematic review

Introduction

Background

Modern mobile phones have long transcended their basic use as communication tools. At present, a smartphone is equally a digital camera, a pedometer, a fitness tracker, or a virtual assistant, among others. Smartphones are familiar, unobtrusive, and discrete devices in today's society. Their various embedded sensors along with their high ubiquity have turned them into a valuable accessory in multiple areas of research. One such area is passive sensing or self-monitoring for either predicting or classifying health-related behaviors of smartphone users [1].

Behavioral patterns such as app usage, social interactions, and a user's activity log or contextual information such as user's location or Wi-Fi connectivity are just a few examples of smartphone data that can be modeled into passive indicators of a user's health or well-being [2,3]. A smartphone's numerous embedded sensors such as digital camera, microphone, global positioning system (GPS), accelerometer, gyroscope, Wi-Fi, Bluetooth, light and sound sensors, along with their programmable platforms, enable the passive collection of user data, thus making smartphones particularly promising self-monitoring tools.

Objectives

This systematic review aims to overview current existing literature about the passive sensing technologies and data of smartphones used to monitor users' health status. Passive sensing does not require any explicit user involvement but rather relies on the ubiquity of smartphones for gathering meaningful data in the background, without any biases that could be introduced by users' categorical participation. In this review, we assess recent studies on the use of smartphones as a tool for providing passive health insights, which do not use any other kind of complementary sensing or monitoring tools. Moreover, we are interested in highlighting possible limitations or system validation concerns that have been identified in the studies included in the review.

Methods

Search Strategy

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and is registered in the PROSPERO database (identifier *CRD* 4201912447). The objective of this paper was to review the literature regarding the functionality of passive sensing of modern smartphones. As such, we focused on finding the most suitable keywords for retrieving recent studies that focus on this topic. We conducted a bibliographic search on the following Web-based databases: PubMed, IEEE Xplore, ACM Digital Library, and Scopus.

The search query used for this purpose was as follows: (smartphone OR mobile) AND (sensing OR monitoring) AND well-being AND (health OR mhealth)

This strategy retrieved 7602 publications. Papers published between January 2014 and March 2019 were included in the search. We first removed duplicate titles by an automatic script and then assessed the remaining titles for relevance for the topic. The studies that passed this first assessment were further evaluated based on their abstract. The final decision on the inclusion of a study was based on its full-text evaluation.

Inclusion and Exclusion Criteria

The titles, authors, and publication dates of the manuscripts resulting from the search were provided in a list that was further ordered by author names. Manuscripts written by the same author group and that refer to the same methodology or application were analyzed for the sake of identifying the most recent or complete publication. Having identified one such manuscript per author group, the remaining articles written by the same author group were discarded, as they would contain similar content and thus add some redundancy to the final results of the review. Other inclusion criteria were as follows:

Relevance for the Chosen Topic

Study focus is passive sensing. Therefore, studies in which users have to explicitly manipulate the smartphone were not considered. Publications that considered smartphone as the sole sensing device were included.

Publication Date

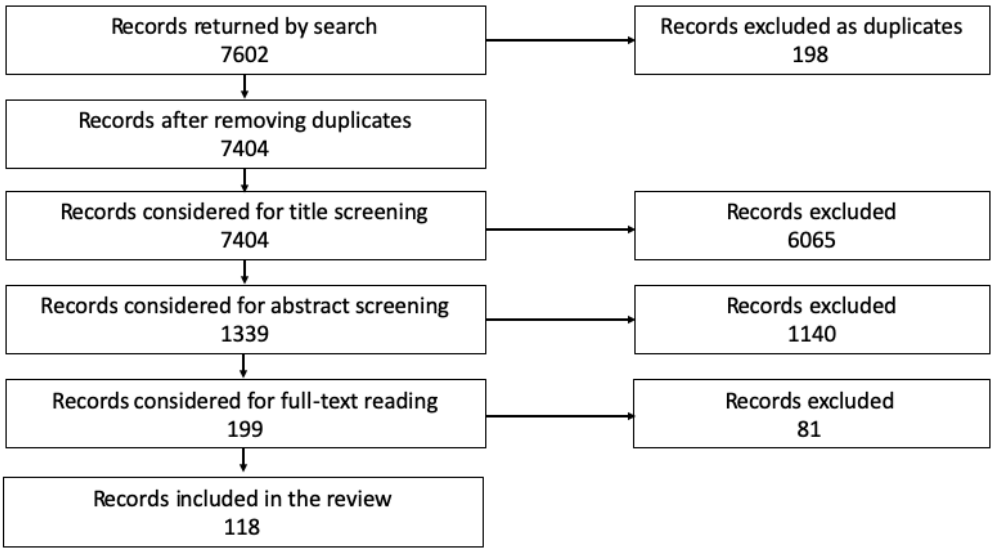
Papers published from January 1, 2014 to April 1, 2019 were included in the review. Due to the fast evolution of smartphone technologies, what existed a few years ago may be obsolete now. Therefore, we decided to include only recent manuscripts based on current technologies.

Exclusion criteria were as follows: (1) publication language other than English; (2) use of other sensing devices or external sensors; (3) user interaction with the sensing system—this review focuses on passive sensing, where users should neither be aware of the sensing process nor willingly interact with the device for this purpose; (4) unavailability of the full text of a manuscript through the library services in our research institute; (5) out of scope for this review's target; (6) lack of results—position papers were excluded; and reviews.

Study Selection

On the basis of the aforementioned selection criteria, the query results were evaluated based on their titles first and then abstracts. The full text of remaining papers was read and analyzed critically to select the ones that best fulfilled the main purpose of this review. Figure 1 shows a PRISMA flow diagram [4] of the bibliographic search.

Figure 1. Flowchart describing the selection of the studies for the review.



Data Collection and Analysis

The first 2 authors of this review performed individual assessments of the papers to be included in the review. These reviewers identified possible bias in each paper, based on the Cochrane Collaboration's risk of bias tool [5]. Finally, observations were combined into 1 spreadsheet for discussion. In case of disagreement, the third author provided advice on the final decision regarding the inclusion of a manuscript. No papers were discarded because of bias.

Study Limitations

The search query used for the retrieval of studies for this review resulted in 7602 papers. These papers were evaluated by 2 reviewers only, which may have caused biases in the selection and screening of search results considering the topic of the review. However, when in doubt, the 2 reviewers involved the third author for an objective opinion. Another limitation of this review is the fact that all the information presented and summarized here was manually collected.

Results

Overview

A total of 7602 manuscripts were retrieved through the systematic search methodology described above. After removal of duplicates, we obtained 7404 studies. Of the 7404 titles inspected, 1339 were considered suitable for abstract assessment. Out of these, 199 abstracts were considered as potential candidates for the review, which led to 199 full-text retrievals and assessments. Finally, 119 manuscripts were included in the review. Table 1 shows the number of returned and selected papers from different Web-based databases.

The exclusion of a large number of papers after title assessment is because of the broadness of the search query used for study retrieval. The query was not applied to a specific field or section of a paper (eg, title or abstract), rather we looked for the terms in the query anywhere in the text of the manuscript. This led to the retrieval of a large number of papers related to the Internet of Things, smart homes, wearable monitoring systems, and robotics, as well as a considerable number of systematic reviews. The abstract evaluation further refined the number of candidate studies as, on one hand, many revealed the use of external sensors as complements of smartphones in the sensing process. On the other hand, many studies exposed explicit human interaction with the monitoring system, which would no longer satisfy 1 of the inclusion criteria for this review, passive sensing. Table 2 summarizes the percentage of excluded papers in the last step of our evaluation, based on the exclusion criteria described above.

Among the studies included for the review, we can verify that the number of published papers related to passive sensing and monitoring of health conditions using smartphones has increased over the years. More particularly, the number has doubled from 2014 to 2017 as shown in Table 3. This advocates for the research interest on the topic and strengthens the motivation of this review.

Below we provide an overview of some of the study characteristics including their main purpose, target audience, number and types of participants, and the sensing methods used. We also compiled the health conditions that have been studied and monitored using the various smartphone sensors.

Table 1. Number of returned and selected papers from different databases.

Studies	PubMed, n	IEEE Explore, n	ACM Digital Libraries, n	Scopus, n
Returned	2994	1604	409	2595
Included in review	44	41	10	23

Table 2. Distribution of rejected papers resulting from the full-text assessment.

Reason for exclusion	Excluded studies, n
Full text not available	18
Review paper	7
Off-topic	12
Preliminary work	6
User interaction required	22
Use of external sensors	17
Language (not English)	2
Same application–different study	1

Table 3. Number of unique returned papers by year.

Year	Studies per year, n
2014	16
2015	19
2016	30
2017	28
2018	19
2019	6

Focus and Target Population of Included Studies

As shown in Table 3, the interest in sensing capabilities of a smartphone with the aim of improving users' health and well-being has been increasing over the last few years. Among the selected papers for this systematic review, physical activities and mental health are 2 of the most studied health dimensions, along with sociability, students' academic performance monitoring, and general well-being, as shown in Table 4.

Of the selected papers, 29.6% (35/118) are dedicated to the detection of users' physical activities. Most of them aimed to recognize basic daily activities such as walking [6-22], standing or sitting [6-12,15,17-19,21-24], jogging or running [6-13,17,24], going up and down the stairs [6-10,15], lying down [10,11,15], and driving a bike [6,12,13] or vehicle [13,14,25]. In addition, 1 study tried to infer riding up and down an elevator [15], 1 assessed different activities including being stationary, limping, shuffling, and skipping [13], and 1 detected shopping and dining activities [14]. Physical activities were also explored in the sense of detecting and counting steps [26,27], distinguishing physical activity from lifestyle activities such as eating [28,29], assessing mobility in the elderly to avoid sedentary lives [30], studying its relationship with happiness including nonexercise activities [31-33], or even measuring and predicting the walking speed and distance of patients with pulmonary diseases [34].

Another health-related issue well studied in the selected papers is mental health disorders. Some of the mental health-related issues, factors, or diseases that have been investigated using smartphones are as follows: stress conditions [35-37], bipolar disorder [38-42], anxiety [42,43], schizophrenia [44,45], depression [46-49], psychotic relapse [50], mood [51-54], and affect, which have been detected, for example, using photos taken by the camera in smartphones [55,56]. A novel approach for understanding users' emotions is the study of the typing behavior and texting speed of the users [57]. The influence of users' exposure to natural outdoor environments on mental health has also been investigated through passive sensing [58]. Two other studies developed their monitoring solution including a recommendation system to support patients with depression to cope with their diagnosis [59,60]. Mental health systems have also been used as a tool by caregivers to access the summary of situations experienced by patients with depression [61] or to alert physicians and families if an abnormal behavior is detected in patients with mood disorders [62].

Sociability has been less studied, but it is an equally important health dimension of people's overall well-being. It is known to have a considerable impact on the stress and anxiety levels of individuals. In fact, healthy relationships between colleagues may improve their productivity [3], united families are happier [29,63], and students cope better with their studies when surrounded by friends [64]. One way of analyzing this health dimension is by exploring interaction patterns and near locations [65,66]. An interesting approach for using the sensing capabilities of smartphones to infer the risk-taking propensity of users has been proposed in 2 studies [67,68]. One recommendation study in particular includes a feature that informs caregivers that their patients feel lonely and need additional examination [69].

Only 5.9% (7/118) of the selected papers chose to infer users' sleep by detecting sleep patterns, irregular nights, and sleep start and end times [2,70-74]. One study focused on the correlation between sleep patterns and schizophrenia [75].

The health areas described above were investigated on an individual basis by some of the selected papers, but several other studies explored more than just 1 area to infer insights on users' general well-being. Such systems are developed to detect the physical activities, sleep patterns, sociability levels, and location of users to either better understand and improve their behaviors or to promote awareness and self-reflection [76-80].

Self-monitoring systems have become very helpful in supporting older people with their health conditions and in the early diagnosis of abnormal conditions in the elderly. For example, the easy monitoring of cardiac parameters with smartphones, only using the users' photographs of the finger or face, can provide a first pulse rate estimation, and users can quickly understand if something is wrong and needs additional examination [81,82]. Similarly, incidents of fall events and tremors are prone to increase in older people. Fall detection systems can quickly alert when a fall occurs, decrease the time spent on the floor, and reduce the fear of falling among the elderly [1]. On the other hand, the early diagnosis of hand tremors by passive sensing is an important contribution in the diagnosis and treatment of Parkinson disease [83-86].

Finally, 10.1% (12/118) of the selected studies developed monitoring systems specifically dedicated to students, mainly to understand how their behaviors (physical activities, sleep, and social interactions) affect their academic performance [87-89], mental health [46-48], social anxiety [43], mobility, and behaviors [18,90-93]. One study [94] presented an approach for predicting the students' food purchase within their proximity to provide them with recommendations about healthier options.

Considering the selected papers and their described study focus, we can categorize them by disease or lifestyle monitoring. In fact, 4 studies aimed to monitor health conditions related to a specific disease, such as detecting sleep abnormalities in patients with schizophrenia or hand tremors in those with Parkinson disease, and another 12 opted to use smartphones to sense users' daily lives to improve their general health and well-being. Among the studies that targeted a specific population, 6.9% (8/118) were on monitoring students' lives [64,69,87,88,90,92-94] and 27.1% (33/118) on people with mental health conditions, such as depression or schizophrenia [55,60-62,75]. Senior population and workers were targeted by 3 studies each [1,3,9,30,35,69]. Among the remaining studies, 2 aimed to monitor patients with Parkinson disease [83,84], 1 targeted pulmonary patients [34], and 1 targeted family members [29]. It should be noted that among studies that aimed to monitor diseases, almost all of them targeted a specific population.

Table 4. Study fields of the selected papers.

Study topic	Studies per topic, n
General well-being	12
Fall detection	6
Sleep	7
Sociability	9
Mental health	33
Physical activity	35
Heart rate	2
Hand tremors and Parkinson's	4
Respiratory issues	3
Students' well-being	10

Smartphone Technologies

Today's off-the-shelf smartphones are equipped with many passive and powerful sensing technologies, which allow the continuous collection of various health-related data. Among the smartphone physical sensors, accelerometer is the most used sensor because of its low privacy and power consumption. In fact, 56.7% of the selected papers (67/118) took advantage of this sensor to gather users' data, mostly related to physical activities.

The GPS is another commonly explored physical sensor in smartphones as it is part of most commercially available smartphones. Of the studies included, 40.6% (48/118) collected useful GPS data about users' location and movements. This sensor was either used alone or along with Wi-Fi, Bluetooth, or accelerometer. Besides the users' location, Bluetooth was highly used to infer levels of sociability. In fact, of the 6 papers that used Bluetooth, 4 aimed to detect users' physical encounters.

Microphone and gyroscope are other well-studied sensors in passive systems and have been explored in 20.3% (24/118) and 16.9% (20/118) of the selected papers, respectively. A microphone is used to infer loneliness, sleep, and fall events, and a gyroscope is essentially used to detect basic physical activities.

In addition to the information collected by physical sensors, proposed solutions also collect a set of useful health-related data about the use of smartphones and users' usage pattern. The most common ones are related to communication events including calls and text messages and smartphone usage such as screen events, light values, time spent on the phone, and device settings. Furthermore, battery level and status and app usage, used in 5.9% (7/118) of the selected papers, allow the collection of useful data about sleep.

Other health-related data can be collected from physical sensors and smartphone data, for example, camera, temporal context, and magnetometer, as shown in Table 5. The table overviews the sensors and smartphone data that are used in the selected papers.

Table 6 provides an overview of the use of smartphone sensors and data in the selected papers and different health areas.

Table 5. Source of the health-related data in percentage. SMS: short message service; API: application program interface; GPS: global positioning system.

Source of data	Studies, n
Camera	3
Google APIs	3
Battery level & stats	5
Magnetometer	6
Bluetooth	6
SMS & calls	13
Gyroscope	14
Microphone	17
Wi-Fi	15
Smartphone & app usage	19
GPS	48
Accelerometer	35
Others	8

Table 6. Summary of the smartphone sensors used in the reviewed papers.

Studied behavior	Smartphone sensors/data
General well-being	Microphone [77,95]; accelerometer [32,76,77,79,95-97]; smartphone usage [54,67,77,79]; app usage [54,78], activity recognition API[a] [78]; text messages, calls, Wi-Fi [78,79]; GPS[b] [67,78-80,95,97]; Bluetooth, magnetometer, gyroscope, battery level and status [79], camera [98]
Fall detection	Audio features (microphone) [1], accelerometer [99-102], GPS[99]
Sleep	App and smartphone usage [2,71,73,103]; Wi-Fi, temporal context, battery level and status [2,70,71]; accelerometer [2,71,75]; GPS, calls, text messages, activity recognition API [70]; microphone [71,74]
Sociability (loneliness, relationships)	Bluetooth [3,65,66,69]; accelerometer, gyroscope, microphone [29,104]; GPS [29,43,64,65,68,104]; Wi-Fi [29,66,69]; calls, text messages, social app usage [43,65,68,69]; emails [69]
Mental health (depression, emotions, stress level, bipolar disorder, schizophrenia)	GPS [36-39,42,44,46,47,50-52,58-61,72,105-107]; smartphone and app usage [36,39,41,42,52,53,59,60,72,106,108-110]; accelerometer [35,36,39-42,44,51,57,58,60,72,106]; cell-ID/calls [45,49,51,72,105-107]; text messages [42,45,51,55,105,107]; Wi-Fi [42,44,47,51,60]; Bluetooth [44]; microphone [36,40,44,45,51,52,62,106]; camera [55,56]; keyboard [57]; temporal context [60]; battery usage [37]; Bluetooth [37]; Google location services API, activity recognition API [61,63]
Physical activities recognition (mobility, steps counting)	Accelerometer [6-13,15-19,21-28,30,31,33,34,38,111-114]; gyroscope [6,12,15,16,18,21,22,25,27,30,33,111,113,114]; magnetometer [6,16,18,21,27,111,113]; GPS [13,14,17-20,28,115-117]; barometer [15,18,111]; gravity sensor [26], microphone [18,28]; Wi-Fi access points [17,18,28]
Heart rate measurements	Camera [81,82]
Hand tremor	Accelerometer [83,84]; gyroscope [84]
Oxygen, breath, and voice analysis	Accelerometer [118]; microphone [119,120]
Parkinson disease	GPS [86]; gyroscope [85]; accelerometer [85]
Students' monitoring (behaviors, performance)	GPS [48,87,88,91]; microphone [87,88,90,91,93]; Wi-Fi [87,88,91,94]; accelerometer [87,90,91,93]; smartphone usage [87,90,91]; temporal context [88]; app usages, text messages, calls [48,89,90]; battery level and status [90,91]; location, weather data [92]; gyroscope, Bluetooth [91]; Google activity recognition [89]

[a]API: application programming interface.

[b]GPS: global positioning system.

One of the main advantages of the use of smartphones in health monitoring is the possibility to passively collect data. Passive data collection means that user interaction or participation is not intentional, and all sensing data come from the ubiquitous sensors of the smartphone. Of the 118 selected papers, 50 used collected data from only 1 sensor, mostly accelerometer to detect physical activities. GPS and camera were also used alone in 7 different papers. On the other hand, of the selected papers that investigated the use of several sensors, 23 used accelerometer that was essentially used along with gyroscope, GPS, Wi-Fi, and microphone to detect physical activities and general users' behaviors.

In the spectrum of smartphone technologies, one of the main challenges that can affect the health-related collection of data when developing monitoring systems is the choice of the operating system. In fact, there are some differences and difficulties in development for Android or IOS systems, the 2 most used phone operating systems worldwide. Android is currently the most popular system and has the advantage of being convenient from the programming point of view [7]. Scanning rates of sensors are found to be superior with this operating system [3]. Furthermore, IOS hampers third-party apps to run endlessly in background, which may make the data collection difficult [91]. Of the selected papers, 56.7% (67/118) developed their system only for Android smartphones, 6 developed for both Android and IOS, and 45 did not provide any information about the chosen operating system.

System Validation

To ensure that users of smartphone-based passive monitoring systems engage with their use requires a strong validation before releasing such systems for mass usage. Three aspects related to the validation of systems can be highlighted: the dimension of the sample of participants, the study duration, and the ground truth data that are used to compare and evaluate the results.

Validation of monitoring systems is an important phase as it can provide researchers and developers with relevant feedback and information about the accuracy and efficiency of the systems. The developed systems are tested by a sample of participants for a specific duration. Of the selected papers, about 71.1% (84/118) asked less than 50 participants to use and test their developed systems. Only few studies tested their monitoring systems with more participants: in 16.1% (19/118) of the studies, the systems were tested by 51 to 450 participants, and only 2.5% (3/118) used more than 10,000 participants in the validation phase. Although most of the papers gave information about the number of participants on their studies, 12 out of 118 (10.1%) did not provide any relevant information (see Table 7). Another aspect to be noted is that, of the papers with information about the participants, 21 out of 118 asked students to test their developed systems [2,56,64,69,79,87,88,90-92,94]. This may be an indicator of the willingness of younger adults to engage in this area.

Study duration is also an important feature to be considered. Of the selected papers, about 20% (24 out of 118) did not provide any relevant information about the study duration. Of those studies with a specific study duration, 16.1% (19/118) lasted between 1 to 3 weeks or between 4 to 8 weeks, 12.7% (15/118) lasted between 8 to 35 weeks, and 7.6% (9/118) lasted for more than 36 weeks. Some of the papers that tried to detect physical activities chose to ask the participants to perform specific activities to test their developed systems without having a specific duration (29%, 35/118) [1,6-13,15,26,27,30,34,84] (see Table 8).

Only 1 of the selected papers did not provide any information about the number of participants and the study duration but mentioned that they had used 4 different smartphones to infer nearness based on users' daily activities and social interactions over time and space [66].

Ground truth data allow the comparison and validation of the data collected by smartphones. Of the selected papers, about 59.3% (70/118) indicated the type of data used as ground truth, and the remaining studies did not provide any relevant information. The most used method is self-reports and questionnaires that can be performed either by a physician or provided by the participants. This method is very useful when testing monitoring systems because self-reports can be prompted to the users in their smartphones without involving any additional efforts. On the other hand, this method presents some disadvantages because the users may not always respond accurately, and results turn out to be biased. In the studies selected for this review, the questionnaire method has been essentially used to collect information about the users' mental health [36,39,55,56,64,77,87,92], sleep [2,58,77], stress levels [35,53,55,108], and physical activities [51,77,109]. Some of the studies chose to use self-reports recognized in the health area, as for example the Patient Health Questionnaire about depression [59,60], the Pittsburgh Sleep Quality Index [75], the Unified Parkinson's Disease Rating Scale [84], and the Beck's Depression Inventory [80]. To collect ground truth data, dedicated devices can also be used as an alternative to questionnaires: actigraph [34], fitness devices [27,34], electrocardiogram [82], and video clips [15,30] to record participants' physical activities. The actual pulse rate of the participants has also been collected when trying to measure cardiac parameters using the smartphone [81]. Of the selected papers, 3 asked the participants to manually label the data about them used during the study [9,65,88].

Table 7. Number of participants within the selected papers.

Participants, n	Studies per participant range, n
≤50	84
51–450	19
>10,000	3
Not specified	12

Table 8. Study duration of the selected papers.

Experiment duration, weeks	Studies, n
1 – 3	19
4 – 8	15
9 – 35	19
≥36	9
Not specified	57

Limitations and Validation Concerns

Users' motivations, interests, and concerns about monitoring systems may influence their adherence on using available solutions. Some of them are related to physiological aspects such as improving behaviors or monitoring health conditions such as cardiac parameters, and others are related to more technical aspects of the systems. Selected papers in this review had more concerns about technical limitations of the proposed solutions as they may affect the users' interest and adherence to monitoring systems.

As described previously, 56.7% (67/118) papers decided to develop their systems with Android as it is simpler to develop third-party apps and because it is the most common operating system worldwide attracting more people to use the proposed systems.

Battery levels and privacy are 2 main themes approached in some of the selected papers. In fact, if these 2 aspects do not fulfill the users' expectations, they may not use the available solutions. Of the selected papers, about 36.4% (43/118) improved the use of smartphone battery or demonstrated some concerns about its levels and hope to improve this performance in future work. The most used solution to maintain reasonable levels of battery was to decrease the sampling rates of sensors [2,13,35,70,77,91,93]. Other studies chose to pause the sampling when the battery was low [51] or to only do a unique sampling per day [65]. Finally, only 1 study [11] used accelerometer to classify activities because this sensor does not use much battery. Related to privacy, 25.4% (30/118) evidenced that privacy issues may drop users' adherence. For example, users may want their data to be securely stored as explained and implemented in 2 studies [34,87]. Other studies chose to not store any user information on the smartphone or in the cloud [51,78], to hash all the relevant information about the user [2,3,65,78,87] or to only use the accelerometer as it raises few privacy concerns [35].

Another possible limitation of these studies is that if a developed system is tested by a sample of young adults, it may not be adapted to senior people, and results may not be accurate [1,15]. Some of the proposed models were developed and tested only with a specific population and may be too personalized, thus leading to inaccurate results when the systems are used by other populations [88,94]. Other papers pointed out the fact that personalized models produced better results than general models [2,35,70,76]. Summing up, about 16.1% (19/118) raised some concerns about the accuracy of the developed models when used on different populations. This percentage can be explained by the fact that 43.2% (51/118) of the selected papers chose to develop their systems to specific populations, and no concerns were raised by the developed models.

One of the main advantages in using a smartphone in health monitoring is its unobtrusiveness. However, almost half the selected papers required the smartphones to be on a specific body position, such as in the pocket trouser, in the handbag, or in the hand. Other studies required the smartphones to be placed in the users' vicinity [1,2,59] or to keep it always on to make sure that the system works correctly [3,75]. These conditions may nullify the use of smartphones as it turns it into an obtrusive device for users.

Finally, considering that the main purpose of health monitoring systems is to improve users' behaviors, health, and well-being, 37.2% (44/118) of the selected papers referred the importance of a recommendation and feedback system to make sure that users are aware of their behaviors to be able to improve them. In fact, such system features may lead to improvements in users' daily lives and health when providing useful information to users, for example, improvements in subjects' depression levels [60]. However, users are not willing to receive too many recommendations, as described in 1 study [55], and notifications should be sent to users only when necessary, for example, when symptoms are detected [83].

Table 9 presents a list of the selected papers that referred the described technical aspects that can have an impact on the users' adherence to the systems.

Table 9. List of the selected papers that referred possible limitations either in the validation of the systems or in their use.

Concerns	Reference
Battery levels	[2,3,11,13,15,29,31,35,51,61,65,66,70,77-79,83,91,93]
Privacy	[2,3,34,35,51,61,65,69,70,78,79,87,91]
Developed models	[1,2,15,35,70,76,88,94]
Smartphone body position	[1-3,6-13,15,26,27,29,30,34,58,59,75,83,84]
Recommendations and feedback	[29-31,51,55,57,59-61,65,77,78,83,88,90-94]

Discussion

Comparison With Prior Work

The reviewed studies illustrate the potential of monitoring several health dimensions using only data collected from the smartphone to support users in improving their health and well-being. Several strategies for data collection were demonstrated for different health areas offering researchers several options to develop passive sensing solutions. We provide an overview of the limitations of such health-related monitoring systems reviewing the specific use of smartphone technologies to monitor, understand, and improve users' well-being through several health dimensions. As far as we know, this is the first review that investigates the use of smartphone sensing technologies and data in health monitoring and discusses the limitations and concerns on using such systems.

Many reviewed papers focused on specific conditions such as mental health (bipolar disease, schizophrenia, major depressive disease, and mood disorder) [121-125], stress [126], cardiology [127], sleep [128], weight control through physical activities [129], management of chronic diseases in older adults [130], or in a more general way, health and well-being with particular representation of mental health and sleep [131], and psychological research (social interactions, activities, and mobility patterns) [132]. Regarding the technologies and devices used in the reviews, smartphone is the most commonly used [121-132], but only a few studies used it to collect data from its sensors [121,123,126,128,129]. In other cases, smartphones are used to prompt ecological momentary assessments to users [123,124,126], provide smartphone apps [122-125,128], or send some recommendations by short messaging service to the users [129]. Reviewed papers also consider wearable devices [122,123,125,128,130] and other devices and technologies such as tablets, fitness trackers, smartphone-connected devices, accessories, and desktop resources [123,127-130].

Compared with these reviewed papers, this review does not target a specific condition or a sensor. Our ambition was to identify all health-related aspects that can be monitored with a smartphone and to understand how far we are from using such systems as an alternative or a complement to standard clinical procedures.

Current Challenges

Although the use of smartphones in health monitoring demonstrates to be a promising study field, available solutions still face some limitations that need to be overcome to make sure that users are comfortable and confident in using such systems. In fact, in some situations, monitoring systems may be perceived as uncomfortable, burdensome, and intrusive to users.

Regular users expect monitoring systems to be able to provide useful information and recommendations about their behaviors [133]. Given a health-related feedback, users are prone to improve their lifestyle and habits in relation with physical activities, well-being, sociability, and mental health [134,135].

Several technological aspects of health monitoring systems using smartphones should be taken into account. Among them, the most interesting one is the possibility to passively and continuously collect health-related data about users without changing their daily lives, thus turning smartphones into an unobtrusive and less burdensome tool compared with other health devices. In addition, smartphones are portable, cheaper, and more convenient than other devices and stay with the users throughout the day, which makes them a familiar tool to users [135]. Moreover, these passive systems can be used to share behavioral and health-related data with health professionals and peers. Recommendations, interventions, feedback, and reminders can be integrated to inform the users about their current state and eventually improve it [133,136].

Despite these advantages, users may still have some concerns about the use of smartphones in health monitoring. Nowadays, users decide very quickly on whether they are going to use a smartphone app or not; therefore, the developed systems should fully meet their expectations. The first aspect that the users normally evaluate is the design of apps. In addition, they hope that the developed system is easy to use and that they will not spend too much time to understand how it works. Concerns about the battery and privacy are also often raised. In fact, users expect that their battery level will not drop significantly given that these systems usually run in background continuously. Users may also discard apps because of privacy issues. Data collected using smartphones are private and should not be shared without permission or maliciously accessed. Generally, users accept to share their data with physicians or within a group of people with the same goal but are not comfortable with sharing it on social media sites, as an example. In addition, users are comfortable with apps using password access but are not willing to spend too much effort in creating accounts. Moreover, inconsistent or inappropriate results or advice may lead to the removal of a certain app. Still related to technical aspects, users expect that the app will not consume excessive space and memory and that it can run in background without affecting other smartphone functionalities [133,136].

Another important point is that users are willing to receive a reasonable number of notifications about their current state, mostly positive recommendations. The possibility to choose the frequency and timing of notifications is a feature that is interesting to them [133]. On the other hand, users are also interested in setting personal goals and achieving them. This shows that a challenge or gamification feature is prone to increase the users' engagement [133,136].

Considering the described challenges and possible concerns, the developed systems referred in this systematic review still face some limitations that need to be overcome to meet users' expectations and needs. First of all, validation of monitoring systems is one of the most important phases, and the systems should be tested with a sample of population highly representative of the target population for a sufficient period to collect enough data and produce results as accurate as possible. Among the selected papers, 71.1% (84/118) asked up to only 50 participants to test the developed system, and about 17.7% (21/118) of the selected papers tested their system for 1 to 3 weeks, which seems to be a short period to ensure reasonable results to make sure users are confident on using available solutions. In addition, some of the proposed systems developed

models too personalized for specific populations, which may produce inaccurate results when using the system with other populations. Furthermore, the main advantage of using smartphones as a data collector is its unobtrusiveness. However, 43.2% (51/118) of the selected papers require users to keep the smartphone near them or use it on a specific body position such as hand, chest, or trouser pocket. Privacy and battery levels are other 2 aspects that need to be considered when developing monitoring systems and that make users more confident when using such systems. In fact, users insist on maintaining a good battery level despite the use of several smartphone sensors and expect that their data will be securely stored.

This review points out that smartphones may have the potential to collect health-related data and provide useful feedback to users about their health conditions. Despite the growing interest and ongoing maturation, monitoring systems may still need to be improved to attract a more diversified type of users and meet their expectations. Besides above-mentioned needs and concerns, more questions may be raised by the use of smartphones in health monitoring. In fact, at present, smartphones are used worldwide, but younger population are more comfortable using them. Health monitoring systems may be very useful to older populations, but smartphones may not be an easy and adaptable tool to them. In addition, these systems may attract more people with diagnosed diseases and specific goals, such as monitoring behaviors, controlling pulse rate, or improving their fitness, than to people with no specific goal in mind. Finally, a disadvantage of such systems is that when the users are familiar with them or have achieved their personal goals, they may not use the developed system anymore.

Conclusions

In recent years, the capabilities of smartphones have made it possible to detect and monitor health-related behaviors of their users. Smartphones are easy to use, unobtrusive, familiar, and cheap compared with more traditional monitoring methods and come with many sensors that allow the continuous collection of health-related data, without directly interfering with users' daily activities.

As demonstrated by this systematic review, the monitoring of health and well-being of users using a smartphone and its sensors is a promising field, hence the growing interest and ongoing maturation. Although there are a couple of predominant fields in which smartphone passive sensing contributes to the well-being of its users, considerable other domains remain underexplored. In addition, most studies focus on the prevailing use of some of the most common sensors, such as GPS or accelerometer, whereas only a handful of studies have so far explored user patterns in interaction with smartphones.

Smartphones have emerged as a good monitoring tool as they are unobtrusive, discrete, and omnipresent in today's society and allow to continuously collect data about their users. Smartphones facilitate the diagnosis and treatment of some diseases as the care manager may have access to additional data sensed by them. Nevertheless, available solutions still present some limitations, such as privacy and battery issues, that have to be overcome to meet the users' expectations. Finally, another aspect worth mentioning is that researchers and developers are focused on understanding what might motivate users to use such monitoring systems and arouse their confidence and long-term adherence.

Acknowledgments

This work was supported by the Integrated Programme of SR&TD "SOCA" (Reference no CENTRO-01-0145-FEDER-000010), cofunded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.

Conflicts of Interest

None declared.

References

1. Cheffena M. Fall detection using smartphone audio features. IEEE J Biomed Health Inform 2016 Dec;20(4):1073-1080. [doi: 10.1109/JBHI.2015.2425932] [Medline: 25915965]
2. Huang K, Ding X, Xu J, Guanling C, Ding W. Monitoring sleep and detecting irregular nights through unconstrained smartphone sensing. : IEEE; 2015 Presented at: UIC-ATC-ScalCom; August 10-14, 2015; Beijing, China p. 36-45. [doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.30]
3. Boonstra TW, Werner-Seidler A, O'Dea B, Larsen ME, Christensen H. Smartphone app to investigate the relationship between social connectivity and mental health. Conf Proc IEEE Eng Med Biol Soc 2017 Dec;2017:287-290. [doi: 10.1109/EMBC.2017.8036818] [Medline: 29059866]
4. 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]
5. 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]
6. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. Sensors (Basel) 2014 Jun 10;14(6):10146-10176 [FREE Full text] [doi: 10.3390/s140610146] [Medline: 24919015]

7. Wang C, Zhang W. Activity recognition based on smartphone and dual-tree complex wavelet transform. : IEEE; 2015 Presented at: 2015 8th International Symposium on Computational Intelligence and Design (ISCID); December 12-15, 2015; Hangzhou, China p. 267-270.
8. Arif M, Bilal M, Kattan A, Ahamed SI. Better physical activity classification using smartphone acceleration sensor. J Med Syst 2014 Sep;38(9):95. [doi: 10.1007/s10916-014-0095-0] [Medline: 25000988]
9. Spinsante S, Angelici A, Lundström J, Espinilla M, Cleland I, Nugent C. A mobile application for easy design and testing of algorithms to monitor physical activity in the workplace. Mob Inf Syst 2016;2016:1-17. [doi: 10.1155/2016/5126816]
10. Yang HC, Li YC, Liu ZY, Qiu J. Harlib: A human activity recognition library on android. : IEEE; 2014 Presented at: 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing(ICCWAMTIP); December 19-21, 2014; Chengdu, China p. 313-315.
11. Aguiar B, Silva J, Rocha T, Carneiro S, Sousa I. Monitoring physical activity and energy expenditure with smartphones. : IEEE; 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain p. 664-667.
12. Li P, Wang Y, Tian Y, Zhou TS, Li JS. An automatic user-adapted physical activity classification method using smartphones. IEEE Trans Biomed Eng 2017 Dec;64(3):706-714. [doi: 10.1109/TBME.2016.2573045] [Medline: 27249822]
13. Zheng W, Yoshishara Y, Noel T, Tang D, Kubota N. Energy-efficient activity recognition on smartphone. : IEEE; 2016 Presented at: Third International Conference on Computing Measurement Control and Sensor Network (CMCSN); May 20-22, 2016; Matsue, Japan p. 1-4.
14. Wan N, Lin GE, Wilson G. Addressing location uncertainties in GPS-based activity monitoring: a methodological framework. Trans GIS 2017 Aug;21(4):764-781 [FREE Full text] [doi: 10.1111/tgis.12231] [Medline: 28943777]
15. del Rosario MB, Wang K, Wang J, Liu Y, Brodie M, Delbaere K, et al. A comparison of activity classification in younger and older cohorts using a smartphone. Physiol Meas 2014 Nov;35(11):2269-2286. [doi: 10.1088/0967-3334/35/11/2269] [Medline: 25340659]
16. Merchán-Baeza JA, González-Sánchez M, Cuesta-Vargas AI. Using smartphones to collect quantitative data on lower limb functionality in people who have suffered a stroke. J Stroke Cerebrovasc Dis 2018 Dec;27(12):3555-3562. [doi: 10.1016/j.jstrokecerebrovasdis.2018.08.012] [Medline: 30241925]
17. Incel OD, Ozgovde A. Arservice: a smartphone based crowd-sourced data collection and activity recognition framework. Procedia Comput Sci 2018;130:1019-1024. [doi: 10.1016/j.procs.2018.04.142]
18. Mafrur R, Nugraha IG, Choi D. Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. Hum Cent Comput Inf Sci 2015 Oct 12;5(1):31. [doi: 10.1186/s13673-015-0049-7]
19. Lee K, Kwan MP. Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results. Comput Environ Urban Syst 2018 Jan;67:124-131. [doi: 10.1016/j.compenvurbsys.2017.09.012]
20. Solanas A, Batista E, Borras F, Martínez-Ballesté A, Patsakis C. Wandering analysis with mobile phones: On the relation between randomness and wandering. : IEEE; 2015 Presented at: International Conference on Pervasive and Embedded Computing and Communication Systems (PECCS); February 11-13, 2015; Angers, Loire Valley, France p. 168-173.
21. Capela NA, Lemaire ED, Baddour N, Rudolf M, Goljar N, Burger H. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. J Neuroeng Rehabil 2016 Jan 20;13:5 [FREE Full text] [doi: 10.1186/s12984-016-0114-0] [Medline: 26792670]
22. Hnoohom N, Mekruksavanich S, Jitpattanakul A. Human activity recognition using triaxial acceleration data from smartphone and ensemble learning. : IEEE; 2017 Presented at: 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS); December 4-7, 2017; Jaipur, India p. 408-412.
23. Bort-Roig J, Puig-Ribera A, Contreras RS, Chirveches-Pérez E, Martori JC, Gilson ND, et al. Monitoring sedentary patterns in office employees: validity of an m-health tool (Walk@Work-App) for occupational health. Gac Sanit 2018;32(6):563-566 [FREE Full text] [doi: 10.1016/j.gaceta.2017.05.004] [Medline: 28923337]
24. Juen J, Cheng Q, Prieto-Centurion V, Krishnan JA, Schatz B. Health monitors for chronic disease by gait analysis with mobile phones. Telemed J E Health 2014 Nov;20(11):1035-1041 [FREE Full text] [doi: 10.1089/tmj.2014.0025] [Medline: 24694291]
25. Chu H, Raman V, Shen J, Kansal A, Bahl V, Choudhury RR. I am a smartphone and I know my user is driving. 2014 Presented at: Sixth International Conference on Communication Systems and Networks (COMSNETS); January 6-10, 2014; Bangalore, India p. 1-8.
26. Tang Z, Guo Y, Chen X. Self-adaptive step counting on smartphones under unrestricted stepping modes. 2016 Presented at: IEEE 40th Annual Computer Software and Applications Conference (COMPSAC); June 10-14, 2016; Atlanta, USA p. 788-797.
27. Khedr K, El-Sheimy N. A smartphone step counter using IMU and magnetometer for navigation and health monitoring applications. Sensors (Basel) 2017 Nov 8;17(11):2573 [FREE Full text] [doi: 10.3390/s17112573] [Medline: 29117143]
28. Lutrek M, Cvetkovi B, Mirchevska V, Kafal O, Romero AE, Stathis K. Recognising lifestyle activities of diabetic patients with a smartphone. : IEEE; 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); May 20-23, 2015; Istanbul, Turkey p. 317-324.

29. Gu F, Niu J, He Z, Jin X. Familypal: An effective system for detecting family activities based on smartphone. 2017 Presented at: IEEE 15th International Conference on Industrial Informatics (INDIN); July 24-26, 2017; Emden, Germany p. 155-160.
30. Madhushri P, Dzhagaryan A, Jovanov E, Milenkovic A. A smartphone application suite for assessing mobility. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:3117-3120. [doi: 10.1109/EMBC.2016.7591389] [Medline: 28268970]
31. Lathia N, Sandstrom G, Mascolo C, Rentfrow PJ. Happier people live more active lives: using smartphones to link happiness and physical activity. PLoS One 2017;12(1):e0160589 [FREE Full text] [doi: 10.1371/journal.pone.0160589] [Medline: 28052069]
32. Liu CT, Chan CT. Exercise performance measurement with smartphone embedded sensor for well-being management. Int J Environ Res Public Health 2016 Dec 11;13(10):1001 [FREE Full text] [doi: 10.3390/ijerph13101001] [Medline: 27727188]
33. Kelly D, Curran K, Caulfield B. Automatic prediction of health status using smartphone-derived behavior profiles. IEEE J Biomed Health Inform 2017 Dec;21(6):1750-1760. [doi: 10.1109/JBHI.2017.2649602] [Medline: 28092582]
34. Juen J, Cheng Q, Schatz B. A natural walking monitor for pulmonary patients using mobile phones. IEEE J Biomed Health Inform 2015 Jul;19(4):1399-1405. [doi: 10.1109/JBHI.2015.2427511] [Medline: 25935052]
35. Garcia-Ceja E, Osmani V, Mayora O. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. IEEE J Biomed Health Inform 2016 Dec;20(4):1053-1060. [doi: 10.1109/JBHI.2015.2446195] [Medline: 26087509]
36. 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]
37. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. J Med Internet Res 2018 Dec 30;20(7):e10131 [FREE Full text] [doi: 10.2196/10131] [Medline: 30061092]
38. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G, et al. Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder. Assessment 2016 Dec;23(4):472-483 [FREE Full text] [doi: 10.1177/1073191116656794] [Medline: 27358214]
39. 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 6;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
40. Maxhuni A, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. Pervasive Mob Comput 2016 Sep;31:50-66. [doi: 10.1016/j.pmcj.2016.01.008]
41. Wang R, Aung MS, Abdullah S, Brian R, Campbell AT, Choudhury T, et al. Crosscheck: toward passive sensing detection of mental health changes in people with schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp'16; September 12-16, 2016; Heidelberg, Germany p. 886-897.
42. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
43. Boukhechba M, Huang Y, Chow P, Fua K, Teachman BA, Barnes LE. Monitoring social anxiety from mobilitycommunication patterns. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: UbiComp'17; September 11-15, 2017; Maui, Hawaii, USA p. 749-753.
44. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer AE, Mistler LA, et al. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. Psychiatr Serv 2016 Dec 1;67(5):558-561 [FREE Full text] [doi: 10.1176/appi.ps.201500130] [Medline: 26695497]
45. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172. [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]
46. Farhan AA, Lu J, Bi J, Russell A, Wang B, Bamis A. Multi-view bi-clustering to identify smartphone sensing features indicative of depression. 2016 Presented at: IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Arlington, USA p. 264-273.
47. Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, et al. Fusing location data for depression prediction. IEEE Trans Big Data 2018:1-1. [doi: 10.1109/TBDATA.2018.2872569]
48. Boukhechba M, Daros AR, Fua K, Chow PI, Teachman BA, Barnes LE. DemonicSalmon: monitoring mental health and social interactions of college students using smartphones. Smart Health 2018 Dec;9-10:192-203. [doi: 10.1016/j.smhl.2018.07.005]
49. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Dec;36(1):72-81. [doi: 10.1002/da.22822] [Medline: 30129691]

50. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MS, et al. CrossCheck: integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatr Rehabil J 2017 Dec;40(3):266-275 [FREE Full text] [doi: 10.1037/prj0000243] [Medline: 28368138]
51. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathi N, Sandstrom GM. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia p. 103-112.
52. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]
53. McNamara L, Ngai E. SADHealth: a personal mobile sensing system for seasonal health monitoring. IEEE Syst J 2018 Mar;12(1):30-40. [doi: 10.1109/JSYST.2016.2521805]
54. Zhang H, Gashi S, Kimm H, Hanci E, Matthews O. Moodbook: An application for continuous monitoring of social media usage and mood. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. 2018 Presented at: UbiComp'18; October 8-12, 2018; Singapore p. 1150-1155.
55. Jeong S, Breazeal CL. Improving smartphone users' affect and wellbeing with personalized positive psychology interventions. In: Proceedings of the Fourth International Conference on Human Agent Interaction. 2016 Presented at: HAI'16; October 4-7, 2016; Singapore p. 131-137.
56. Wang R, Campbell AT, Zhou X. Using opportunistic face logging from smartphone to infer mental health: challenges and future directions. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: UbiComp/ISWC'15 Adjunct; September 7-11, 2015; Osaka, Japan p. 683-692.
57. Shapsough S, Hesham A, Elkhorazaty Y, Zualkerman IA, Aloul F. Emotion recognition using mobile phones. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom); September 14-16, 2016; Munich, Germany p. 1-6.
58. Triguero-Mas T, Donaire-Gonzalez D, Seto E, Valentín A, Martínez D, Smith G, et al. Natural outdoor environments and mental health: Stress as a possible mechanism. Environ Res 2017 Dec;159:629-638. [doi: 10.1016/j.envres.2017.08.048] [Medline: 28938204]
59. 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]
60. 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]
61. Teles AS, Rocha A, da Silva E Silva FJ, Lopes JC, O'Sullivan D, van de Ven P, et al. Enriching mental health mobile assessment and intervention with situation awareness. Sensors (Basel) 2017 Jan 10;17(1):127 [FREE Full text] [doi: 10.3390/s17010127] [Medline: 28075417]
62. Ng JKY, Wang J, Lam KL, Kam CH, Han S. Capturing and analyzing pervasive data for smart health. 2017 Presented at: IEEE 28th International Conference on Advanced Information Networking and Applications; May 13-16, 2014; Taipei, Taiwan p. 985-992.
63. Sahiti K, Kalanadhabhatta LM, Bhunia SS, Singhal A, Majethia R. Smartphone-based qualitative analyses of social activities during family time. In: Proceedings of the First International Workshop on Human-centered Sensing, Networking, and Systems. 2017 Presented at: HumanSys'17; November 5, 2017; Delft, Netherlands p. 47-54.
64. Huang Y, Xiong H, Leach K, Zhang Y, Chow P, Fua K, et al. Assessing social anxiety using gps trajectoriespoint-of-interest data. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp'16; September 12-16, 2016; Heidelberg, Germany p. 898-90e.
65. Luo L, Yang J, Bao X, Yan Z, Jiang Y. SWAN: A novel mobile system to track and analyze social well-being. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: UbiComp/ISWC'15 Adjunct; September 7-11, 2015; Osaka, Japan p. 703-712.
66. Rute S, Firdose S, Lopes LA, Moreira W, Mendes P. NSense: A people centric, non-intrusive opportunistic sensing tool for contextualizing nearness. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom); September 14-16, 2016; Munich, Germany p. 1-6.
67. Singh VK, Goyal R, Wu S. Riskalyzer: inferring individual risk-taking propensity using phone metadata. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-21. [doi: 10.1145/3191766]
68. Bati GF, Singh VK. “Trust Us”: Mobile Phone Use Patterns Can Predict Individual Trust Propensity. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21-26, 2018; Montreal, Canada p. 330.

69. Pulekar G, Agu E. Autonomously sensing loneliness and its interactions with personality traits using smartphones. 2016 Presented at: IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT); November 9-11, 2016; Cancun, Mexico p. 134-137.
70. Saeb S, Cybulski TR, Schueller SM, Kording KP, Mohr DC. Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. J Med Internet Res 2017 Dec 18;19(4):e118 [FREE Full text] [doi: 10.2196/jmir.6821] [Medline: 28420605]
71. Montanini L, Sabino N, Spinsante S, Gambio E. Smartphone as unobtrusive sensor for real-time sleep recognition. 2018 Presented at: IEEE International Conference on Consumer Electronics (ICCE); January 12-14, 2018; Las Vegas, USA p. 1-4.
72. Sarda A, Munuswamy S, Sarda S, Subramanian V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: crosssectional observational study. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11041 [FREE Full text] [doi: 10.2196/11041] [Medline: 30694197]
73. Lin HY, Wong BY, Lin SH, Chiu YC, Pan YC, Lee YH. Development of a mobile application (App) to delineate "digital chronotype" and the effects of delayed chronotype by bedtime smartphone use. J Psychiatr Res 2019 Mar;110:9-15. [doi: 10.1016/j.jpsychires.2018.12.012] [Medline: 30611008]
74. Nakano H, Hirayama K, Sadamitsu Y, Toshimitsu A, Fujita H, Shin S, et al. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. J Clin Sleep Med 2014 Jan 15;10(1):73-78 [FREE Full text] [doi: 10.5664/jcsm.3364] [Medline: 24426823]
75. Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017 Oct 16;3(1):37 [FREE Full text] [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
76. DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B. Well-being tracking via smartphone-measured activity and sleep: cohort study. JMIR Mhealth Uhealth 2017 Oct 5;5(10):e137 [FREE Full text] [doi: 10.2196/mhealth.7820] [Medline: 28982643]
77. Lane ND, Lin M, Mohammod M, Yang X, Lu H, Cardone G, et al. BeWell: sensing sleep, physical activities and social interactions to promote wellbeing. Mobile Netw Appl 2014 Jan 9;19(3):345-359. [doi: 10.1007/s11036-013-0484-5]
78. Aslam H, Mukhtar H, Seemi F, Belaid D. Harnessing smartphones as a personal informatics tool towards self-awareness and behavior improvement. 2016 Presented at: IEEE 14th Inernational Conference on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech); August 8-12, 2016; Auckland, New Zealand p. 467-474.
79. Mo X, Shi D, Yang R, Li H, Tong ZH, Wang F. A framework of fine-grained mobile sensing data collection and behavior analysis in an energy-configurable way. 2015 Presented at: IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity); December 19-21, 2015; Chengdu, China p. 391-398.
80. DeMasi O, Recht B. A step towards quantifying when an algorithm cancannot predict an individual's wellbeing. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: UbiComp'17; September 11-15, 2017; Maui, Hawaii, USA p. 763-771.
81. Siddiqui SA, Zhang Y, Feng Z, Kos A. A pulse rate estimation algorithm using ppg and smartphone camera. J Med Syst 2016 May;40(5):126. [doi: 10.1007/s10916-016-0485-6] [Medline: 27067432]
82. Huang RY, Dung LR. Measurement of heart rate variability using off-the-shelf smart phones. Biomed Eng Online 2016 Jan 29;15:11 [FREE Full text] [doi: 10.1186/s12938-016-0127-8] [Medline: 26822804]
83. García-Magariño I, Medrano C, Plaza I, Oliván B. A smartphone-based system for detecting hand tremors in unconstrained environments. Pers Ubiquit Comput 2016 Sep 8;20(6):959-971. [doi: 10.1007/s00779-016-0956-2]
84. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C. A smartphone-based tool for assessing Parkinsonian hand tremor. IEEE J Biomed Health Inform 2015 Nov;19(6):1835-1842. [doi: 10.1109/JBHI.2015.2471093] [Medline: 26302523]
85. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Dec;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
86. 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]
87. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp'14; September 13-17, 2014; Seattle, USA p. 3-14.
88. Wang R, Harari G, Hao P, Zhou X, Cambpell AT. Smartgpa: how smartphones can assess and predict academic performance of college students. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan p. 295-306.

89. Baras K, Soares L, Paulo N, Barros R. 'Smartphine': supporting students' well-being according to their calendar and mood. 2016 Presented at: International Multidisciplinary Conference on Computer and Energy Science (SpliTech); July 13-15, 2016; Split, Croatia p. 1-7.
90. Tseng VW, Merrill M, Wittleder F, Abdullah S, Aung MH, Choudhury T. Assessing mental health issues on college campuses: preliminary findings from a pilot study. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp'16; September 12-16, 2016; Heidelberg, Germany p. 1200-1208.
91. Hossain A, Poellabauer C. Challenges in building continuous smartphone sensing applications. 2016 Presented at: IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); October 17-19, 2016; New York, USA p. 1-8.
92. Vhaduri S, Munch A, Poellabauer C. Assessing health trends of college students using smartphones. 2016 Presented at: IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT); November 9-11, 2016; Cancun, Mexico p. 70-73.
93. Harari GM, Gosling SD, Wang R, Chen F, Chen Z, Campbell AT. Patterns of behavior change in students over an academic term: a preliminary study of activity and sociability behaviors using smartphone sensing methods. Comput Hum Behav 2017 Feb;67:129-138. [doi: 10.1016/j.chb.2016.10.027]
94. Chen F, Wang R, Zhou X, Campbell AT. Department of Computer Science - Dartmouth College. 2014. My smartphone knows I am hungry URL: https://home.cs.dartmouth.edu/~xia/papers/wpa14.pdf
95. Barbosa R, Nunes D, Figueira A, Aguiar H, Silva JS, Gonzalez F, et al. An architecture for emotional smartphones in Internet of Things. 2016 Presented at: IEEE Ecuador Technical Chapters Meeting (ETCM); October 12-14, 2016; Guayaquil, Equador p. 1-5.
96. Vathsangam H, Sukhatme GS. Using phone-based activity monitors to promote physical activity in older adults: a pilot study. 2014 Presented at: IEEE Healthcare Innovation Conference (HIC); October 8-10, 2014; Seattle, USA p. 42-47.
97. Kelly D, Donnelly S, Caulfield B. Smartphone derived movement profiles to detect changes in health status in COPD patients - a preliminary investigation. Conf Proc IEEE Eng Med Biol Soc 2015 Dec;2015:462-465. [doi: 10.1109/EMBC.2015.7318399] [Medline: 26736299]
98. Nambi AU, Bannur S, Mehta I, Kalra H, Virmani A, Padmanabhan VN, et al. HAMS: Driver and Driving Monitoring using a Smartphone. In: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. 2018 Presented at: MobiCom'18; October 29-November 2, 2018; New Delhi, India p. 840-842.
99. Kau LJ, Chen CS. A smart phone-based pocket fall accident detection, positioning, and rescue system. IEEE J Biomed Health Inform 2015 Jan;19(1):44-56. [doi: 10.1109/JBHI.2014.2328593] [Medline: 25486656]
100. Sie MR, Lo SC. The design of a smartphone-based fall detection system. 2015 Presented at: IEEE 12th International Conference on Networking, Sensing and Control; April 9-11, 2015; Taipei, Taiwan p. 456-461.
101. Vermeulen J, Willard S, Aguiar B, De Witte LP. Validity of a smartphone-based fall detection application on different phones worn on a belt or in a trouser pocket. Assist Technol 2015;27(1):18-23. [doi: 10.1080/10400435.2014.949015] [Medline: 26132221]
102. Sansrimahachai W, Toahchoodee M, Piakaew R, Vijitphu T, Jeenboonmee S. Real-time fall risk assessment system based on acceleration data. 2017 Presented at: International Conference on Orange Technologies (ICOT); December 8-10, 2017; Singapore p. 33-36.
103. Chow KK, Leong BD, Lee BY. Imagining consequences of excessive smartphone use via a character-based mobile application. Int J Ment Health Addiction 2018 Nov 20;16(6):1420-1434. [doi: 10.1007/s11469-018-9984-7]
104. Fasoulis A, Virvou M, Tsihrintzis G, Patsakis C, Alepis E. Sensus vox: Sentiment mapping through smartphone multi-sensory crowdsourcing. 2018 Presented at: IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI); November 5-7, 2018; Portland, Oregon, USA p. 438-444.
105. Bati GF, Singh VK. Are you altruistic? Your mobile phone could tell. 2017 Presented at: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); August 4-8, 2017; San Francisco, CA, USA p. 1-6.
106. Wang R, Scherer EA, Walsh M, Wang W, Aung MH, Ben-Zeev D, et al. Predicting symptom trajectories of schizophrenia using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 5;22(2):32-37. [doi: 10.1145/3276145.3276157]
107. Asensio-Cuesta S, Sánchez-García Á, Conejero JA, Saez C, Rivero-Rodriguez A, García-Gómez JM. Smartphone sensors for monitoring cancer-related quality of life: app design, EORTC QLQ-C30 mapping and feasibility study in healthy subjects. Int J Environ Res Public Health 2019 Dec 5;16(3):461 [FREE Full text] [doi: 10.3390/ijerph16030461] [Medline: 30764535]
108. Ferdous R, Osmati V, Mayora O. Smartphone app usage as a predictor of perceived stress levels at workplace. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); May 20-23, 2015; Istanbul, Turkey p. 225-228.

109. Rapeepisarn T, Tatiyanupanwong S, Kornvisitvatin B, Tangsripairoj S. iRelief: an android application for smartphone syndrome prevention and treatment. 2016 Presented at: IEEE 5th ICT International Student Project Conference (ICT-ISPC); May 27-28, 2016; Nakhon Pathom, Thailand p. 121-124.
110. Kanjo E, Kuss DJ, Ang CS. NotiMind: utilizing responses to smart phone notifications as affective sensors. IEEE Access 2017;5:22023-22035. [doi: 10.1109/ACCESS.2017.2755661]
111. Gu F, Kealy A, Khoshelham K, Shang J. User-independent motion state recognition using smartphone sensors. Sensors (Basel) 2015 Dec 4;15(12):30636-30652 [FREE Full text] [doi: 10.3390/s151229821] [Medline: 26690163]
112. Weiss GM, Lockhart JW, Pulickal TT, McHugh T, Ronan IH, Timko JL. Actitracker: a smartphone-based activity recognition system for improving health and well-being. 2016 Presented at: IEEE International Conference on Data Science and Advanced Analytics (DSAA); October 17-19, 2016; Montreal, Canada p. 682-688.
113. Mukherjee A, Misra S, Mangrulkar P, Rajarajan M, Rahulamathavan Y. SmartARM: a smartphone-based group activity recognition and monitoring scheme for military applications. 2017 Presented at: IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS); December 17-20, 2017; Bhubaneswar, Odisha, India p. 1-6.
114. Wu T, Wang L, Zheng Z, Wu S, Ma J, Tao X, et al. Carmus: Towards a general framework for continuous activity recognition with missing values on smartphones. 2018 Presented at: IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); July 23-27, 2018; Tokyo, Japan p. 850-859.
115. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan p. 1293-1304.
116. Wan N, Lin G. Classifying human activity patterns from smartphone collected GPS data: A fuzzy classification and aggregation approach. Trans GIS 2016 Dec;20(6):869-886 [FREE Full text] [doi: 10.1111/tgis.12181] [Medline: 28603451]
117. Goodspeed R, Yan X, Hardy J, Vydiswaran VG, Berrocal VJ, Clarke P, et al. Comparing the data quality of global positioning system devices and mobile phones for assessing relationships between place, mobility, and health: field study. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e168 [FREE Full text] [doi: 10.2196/mhealth.9771] [Medline: 30104185]
118. Cheng Q, Juen J, Schatz BR. Using mobile phones to simulate pulse oximeters: gait analysis predicts oxygen saturation. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. 2014 Presented at: BCB'14; September 20-23, 2014; Newport Beach, CA, USA p. 331-340.
119. Azam MA, Shahzadi A, Khalid A, Anwar SM, Naeem U. Smartphone based human breath analysis from respiratory sounds. Conf Proc IEEE Eng Med Biol Soc 2018 Dec;2018:445-448. [doi: 10.1109/EMBC.2018.8512452] [Medline: 30440430]
120. Jannetts S, Schaeffler F, Beck J, Cowen S. Assessing voice health using smartphones: bias and random error of acoustic voice parameters captured by different smartphone types. Int J Lang Commun Disord 2019 Mar;54(2):292-305. [doi: 10.1111/1460-6984.12457] [Medline: 30779425]
121. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
122. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017 Nov 12;7(11):150 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
123. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
124. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Dec 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
125. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Dec;18(2):163-169 [FREE Full text] [Medline: 27489456]
126. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self- assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Dec 13;19(2):e41 [FREE Full text] [doi: 10.2196/jmir.6397] [Medline: 28193600]
127. Nguyen HN, Silva JN. Use of smartphone technology in cardiology. Trends Cardiovasc Med 2016 May;26(4):376-386. [doi: 10.1016/j.tcm.2015.11.002] [Medline: 26686266]
128. Ko PR, Kientz JA, Choe EY, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
129. Thomas JG, Bond DS. Review of innovations in digital health technology to promote weight control. Curr Diab Rep 2014;14(5):485. [doi: 10.1007/s11892-014-0485-1] [Medline: 24664797]
130. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
131. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Dec;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]

132. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: practical considerations and challenges. Perspect Psychol Sci 2016 Dec;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
133. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
134. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
135. Harari GM, Müller S, Mishra V, Wang R, Campbell AT, Rentfrow PJ, et al. An evaluation of students' interest in and compliance with self- tracking methods: recommendations for incentives based on three smartphone sensing studies. Soc Psychol Personal Sci 2017 Jul 18;8(5):479-492. [doi: 10.1177/1948550617712033]
136. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]

Abbreviations

GPS: global positioning system
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 30.10.18; peer-reviewed by A Pereira, D Revere, T Wang, E Kohls; comments to author 04.02.19; revised version received 24.05.19; accepted 28.05.19; published 23.08.19.

Please cite as:
Trifan A, Oliveira M, Oliveira JL
Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitations
JMIR Mhealth Uhealth 2019;7(8):e12649
URL: http://mhealth.jmir.org/2019/8/e12649/
doi:10.2196/12649
PMID:31444874

Review

Evidence on the Use of Mobile Apps During the Treatment of Breast Cancer: Systematic Review

Flávia Oliveira Almeida Marques Cruz[1,2], MSc; Ricardo Alencar Vilela[1,3,4], MSc; Elaine Barros Ferreira[1], PhD; Nilce Santos Melo[1], PhD; Paula Elaine Diniz Dos Reis[1], PhD

[1]School of Health Sciences, University of Brasilia, Brasilia, Brazil

[2]Centro Universitário do Distrito Federal, Brasilia, Brazil

[3]Hospital Universitário de Brasília, Brasilia, Brazil

[4]Grupo CONFIAR, Goiânia, Brazil

Corresponding Author:
Flávia Oliveira Almeida Marques Cruz, MSc
School of Health Sciences
University of Brasilia
Asa Norte
Brasilia, 70910-900
Brazil
Phone: 55 613107 3300
Fax: 55 613107 3300
Email: flavinhaoliveira5@hotmail.com

Abstract

Background: Cancer is a major cause of morbidity, disability, and mortality worldwide, and breast cancer is the most common cause of death in women. Different modalities of cancer treatment can have adverse effects that reduce the quality of life of patients and lead to treatment interruptions, if not managed properly. The use of mobile technologies has brought innovative possibilities for improving health care. Mobile apps can help individuals manage their own health and well-being and may also promote healthy lifestyles and information access.

Objective: The aim of this study was to identify available evidence on the use of mobile apps to provide information and facilitate communication regarding self-care management related to the adverse effects of toxicities owing to breast cancer therapy.

Methods: This systematic review includes studies which were identified using a search strategy adapted for each electronic database: CINAHL, Cochrane Library, LILACS, LIVIVO, PubMed, SCOPUS, and Web of Science. In addition, a gray literature search was performed using Google Scholar. All the electronic database searches were conducted on April 17, 2019. Two investigators independently reviewed the titles and abstracts of the studies identified and then read the full text of all selected papers. The quality of the included studies was analyzed by the Cochrane Collaboration Risk of Bias Tool and the Methodological Index for Non-Randomized Studies.

Results: A total of 9 studies which met the eligibility criteria—3 randomized clinical trials and 6 nonrandomized studies published in English from 2010 to 2018—were considered for this systematic review; 396 patients with breast cancer, as well as 40 experts in the medical and nursing fields, and 3 software engineers were included.

Conclusions: The evidence from the studies included in this systematic review is currently limited but suggests that mobile apps for women with breast cancer might be an acceptable information source that can improve patient well-being; they can also be used to report symptoms and adverse treatment-related effects and promote self-care. There is a need to test more evidence-based apps in future randomized clinical trials.

(JMIR Mhealth Uhealth 2019;7(8):e13245) doi:10.2196/13245

KEYWORDS

mobile applications; health education; nursing care; review; educational technology; breast neoplasms

Introduction

Background

Cancer is a major cause of morbidity, disability, and mortality worldwide, affecting more than 18 million people each year [1]. In Brazil alone, about 600,000 new cases of the disease are estimated for 2019. Among women, the most common types of cancer are those of the breast, intestines, cervix, lungs, and thyroid. Worldwide, breast cancer is the most common cause of death in women, with approximately 626,000 estimated victims in 2018 [1,2].

The different modalities of cancer treatment can have adverse effects that may reduce the quality of life of patients and lead to treatment interruptions if not managed properly [3]. Considering the need to handle side effects in a population with a life-threatening disease, promoting symptom management–related knowledge remains a high priority for patients and a challenge for health professionals [4-6].

The landscape of mobile app usage has been growing and evolving. Apps can be used to offer services related to entertainment, media, education, shopping, finances, travel, health, and so on [7]. The use of mobile technologies presents innovative possibilities for improving health care. Mobile apps can help individuals manage their own health and well-being and may also promote healthy lifestyles and information access. According to international estimates, by 2018, nearly 2 billion smartphone and tablet users were using health-related apps [8].

Considering the important role they can play in patient education, disease self-management, and remote monitoring of patients, the use of smartphones is receiving more attention in the health domain every day. A systematic review investigating smartphone-based health care technologies concluded that many medical apps have been developed and are widely used by health professionals and patients alike [9].

In a study evaluating 185 mobile apps related to breast disease in the main app stores (Apple iTunes, Google Play, BlackBerry World, and Windows Phone), the authors found that most (n=139) concerned breast cancer [10]. A recent cross-sectional review of 599 apps [11] has verified the state of the practice regarding breast cancer-related mobile apps to characterize health apps from app stores (iOS and Android). These studies have identified a lack of evidence related to the involvement of medical experts in the creation and development of such apps. They have, therefore, highlighted the need to identify high-quality apps to increase consumer confidence in their use during health care [10,11].

A systematic review of apps targeting patients with breast and prostate cancer involved 5 studies and a total of 644 patients. The purposes of the apps were related to the main psychological variables in psycho-oncological care: quality of life and anxiety and depression symptoms [12]. Another systematic review identified 29 studies on mobile health apps targeting only patients with breast cancer. More than half of the studies addressed apps in an intervention for prevention, early detection of breast cancer, or survivors of the disease [13]. Both systematic reviews found that rigorous trials regarding the subject are lacking, despite the existence of studies related to cancer-focused apps. Future investigations should continue to explore and test the impact of mobile health apps on the treatment of breast cancer [12,13].

None of these recent reviews exclusively evaluated apps for women with breast cancer during cancer treatment. Therefore, given the magnitude of the disease burden, the needs of this population, the increasing use of mobile apps in the health domain, and the need to identify quality apps, it is necessary to enhance knowledge about the mobile apps available to provide information and improve the course of treatment of women with breast cancer.

Objective

This systematic review aimed to identify available evidence on the use of mobile apps to provide information and facilitate communication regarding self-care management related to the adverse effects of toxicities owing to breast cancer therapy.

Methods

Protocol and Registration

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Checklist [14]. The protocol was registered at the International Prospective Register of Systematic Reviews under number CRD42018083548 [15].

Eligibility Criteria

In this systematic review, we included the following: (1) studies about mobile apps, defined as any computer programs or software installed on mobile electronic devices to provide information and facilitate communication regarding self-care management and adverse effects related to toxicities owing to breast cancer therapy, (2) studies that performed validation of content or evaluated the usability or effectiveness of apps, and (3) studies that collected the opinions of patients, clinicians, or experts about apps developed for patients with breast cancer. There were no restrictions on the year of publication or language.

Studies were excluded for the following reasons: (1) if they focused on mobile apps related to other types of cancer, (2) if they focused on electronic technologies, but not mobile apps, such as telephone services, text messages, videotapes, audiotapes, audiovisual materials in DVDs, websites, games, or online programs for desktop computers, (3) if they concerned the post-treatment period, (4) if their objective was to evaluate mobile apps intended for health professionals but not patients, (5) if they focused on emotional, cognitive, and behavioral strategies, and (6) if they took the form of reviews, letters, conference summaries, book chapters, or studies that only described the development of mobile apps.

Information Sources and Search Strategy

Studies were identified using an individual search strategy for each of the following electronic databases: CINAHL, Cochrane Library, LILACS, LIVIVO, PubMed, SCOPUS, and Web of Science (Multimedia Appendix 1). The reference lists of selected

papers were hand searched for potentially relevant studies that might have been missed in the electronic database searches. In addition, a gray literature search was performed using Google Scholar.

Duplicated references were removed by using appropriate software (EndNote Basic, Thomson Reuters). All the electronic database searches were conducted on April 17, 2019.

Study Selection

Study selection was completed in 2 phases by using an online app (Rayyan, Qatar Computing Research Institute). In phase 1, 2 investigators (FOAMC and PEDR) independently screened the titles and abstracts of all citations retrieved from electronic databases and identified papers that appeared to meet the inclusion criteria. In phase 2, the same investigators independently read the full text of all selected papers and excluded studies that did not meet the inclusion criteria. Any disagreements in the first or second phases were resolved by discussion and consensus between the 2 reviewers. In case a consensus could not be reached, a third investigator (EBF) became involved to make a final decision. Studies that were excluded after full-text assessment and the reasons for their exclusion are listed in Multimedia Appendix 2.

Data Collection Process and Items

Two investigators (FOAMC and PEDR) independently collected data from the selected papers: population characteristics (groups, n, mean age, and treatment focus on app), study characteristics (author(s), country and year of publication, and objective), intervention characteristics (purpose of the app, operation, description, and operating system), and outcome characteristics (primary outcomes and main conclusions). Any disagreement was resolved by discussion and mutual agreement. A third author (EBF) was involved when required to make a final decision. If the required data were not complete, attempts were made to contact the authors to retrieve any pertinent information.

Risk of Bias in Individual Studies

Two investigators (FOAMC and PEDR) independently conducted the risk of bias assessment for the selected papers. Again, any disagreement was resolved by discussion and mutual agreement. A third author (EBF) was involved when required to make a final decision.

To assess the risk of bias of the included randomized controlled trials, including judgments about sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data, and selective reporting, the Cochrane Collaboration Risk of Bias Tool [16] was used. The risk of bias was assessed as low, high, or unclear. We also used the Methodological Index for Non-Randomized Studies (MINORS) [17] for the nonrandomized studies; this was done to analyze the study aim and appropriate endpoints, inclusion of participants, data collection, and follow-up period, as well as the calculation of the study size and loss to follow-up. For the comparative study, the characteristics of the groups and the statistical analyses were also verified.

Synthesis of Results

The heterogeneity across studies was evaluated by considering clinical (treatment-related differences), methodological (design and risk of bias), and statistical (outcome measures) characteristics. Therefore, owing to the heterogeneity among the included studies, a quantitative synthesis was not undertaken. Congruent with the review objectives, the results of the included studies were analyzed and reported according to the characteristics of the mobile apps, their assessment, and satisfaction with their use.

Results

Study Selection

The literature search initially yielded 3416 papers from 7 electronic databases. After duplicate removal, the titles and abstracts of 2396 papers were screened, and 25 potentially relevant studies were selected for full-text reading; 16 papers were excluded (Multimedia Appendix 2) and 9 papers met all the eligibility criteria and were considered for this systematic review [18-26]. Figure 1 shows a flow diagram of the study identification, screening, and inclusion processes.

Study Characteristics

All studies were published in English from 2010 to 2018 and evaluated mobile apps for women with breast cancer during treatment. In total, 4 studies included patients undergoing chemotherapy [18-21], 1 concerned postoperative patients [22], and another focused on both chemotherapy and surgery [23]. One study included patients undergoing chemotherapy and radiation therapy [24], and another included patients undergoing adjuvant endocrine therapy with aromatase inhibitors [25]. Only one study allowed any type of therapy for breast cancer [26].

In this systematic review, 396 patients with breast cancer were included, as well as 40 medical and nursing experts and 3 software engineers. The clinical nursing experts' average career length was 17.0 years, with the duration ranging from 7 to 36 years in one of the studies [23]. One study included 2 breast surgeons, an oncologist, a radiation oncologist, a plastic surgeon, a gynecologist, a clinical geneticist, and 3 specialized breast cancer nurses [26].

A study measured health literacy through a validated instrument, showing that 72.5% of participants had high health literacy [25]. Regarding educational level, in 1 study, 56.1% of the participants had completed elementary or junior middle school [21]. Concerning familiarity with mobile apps, 1 study found that 67% of the participants frequently used them, while 33% were relatively inexperienced before participating in the study [26].

The main characteristics of the studies included are presented in Tables 1, 2, and 3.

Figure 1. Flow diagram of literature search and selection process (adapted from Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA]).

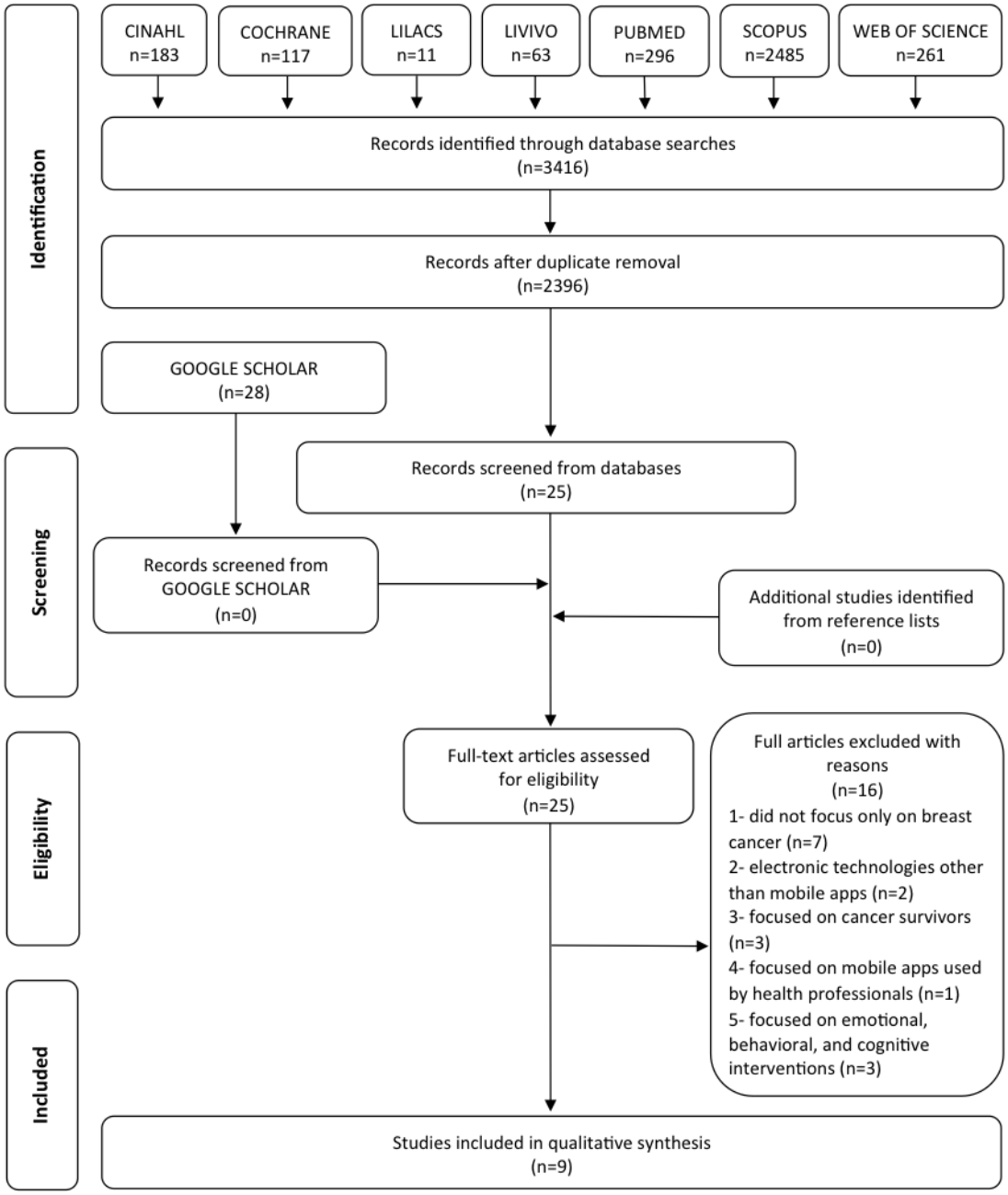


Table 1. Summary of population characteristics of included papers (n=9).

Year, author, country[a]	Groups	N	Mean age (years)	App's treatment focus
2016, Egbring et al, Switzerland [18]	CG[b]: regular physician support; EG1[c]: mobile app without physician review; EG2: mobile app and physician review on scheduled visits	CG: 41 patients; EG1: 45 patients; EG2: 41 patients	CG: 56; EG1: 50; EG2: 53	Chemotherapy
2018, Graetz et al, United States [25]	CG: mobile app without weekly reminders to use it; EG: mobile app with weekly reminders to use it	CG: 23 patients; EG: 21 patients	CG: 59.3; EG: 60.6	Adjuvant endocrine therapy with aromatase inhibitors
2016, Hwang, Canada [22]	CG: conventional follow-up; EG: e-monitoring in addition to conventional follow-up	CG: 37 patients; EG: 35 patients	CG: 65.5; EG: 60.1	Surgery
2010, Klasnja et al, United States [24]	Two groups of patients evaluated the app for three weeks	5 patients	50	Chemotherapy (n=3) and radiation therapy (n=2)
2017, Liu et al, China [23]	EvG1[d]: clinical nursing experts; EvG2: medical and nursing experts and software engineers	EvG1: 19 nurses; EvG2: 8 experts	EvG1: 38.9; EvG2: not available	Surgery and chemotherapy
2016, Young-Afat et al, The Netherlands [26]	EvG1: patients with breast cancer; EvG2: physicians and specialized nurses	EvG1: 15 patients; EvG2: 10 experts	EvG1: 51; EvG2: not available	Any type of cancer therapy
2017, Zhu et al, China [19]	EvG1: specialized nurses and an oncologist; EvG2: patients with breast cancer	EvG1: 6 experts; EvG2: 6 patients	EvG1: not available; EvG2: above 50	Chemotherapy
2018, Zhu et al, China [20]	Two groups (FTFI[e] and TI[f]) of patients that evaluated the app	13 patients	49.5	Chemotherapy
2018, Zhu et al, China [21]	CG: only usual care; EG: e-support program and usual care	A: 57 patients; B: 57 patients	A: 46.2; B: 48.2	Chemotherapy

[a]Country of the study coordinator.

[b]CG: control group.

[c]EG: experimental group.

[d]EvG: evaluator group.

[e]FTFI: face-to-face interviews.

[f]TI: telephonic interviews.

Table 2. Summaries and intervention characteristics of included papers (n=9).

Year, author, country[a]	Study characteristics	Intervention characteristics	
	Objective	Purpose of the app	Operation
2016, Egbring et al, Switzerland [18]	To evaluate the effects of a mobile app on patient-reported daily functional activity	Improvement in the patient-reported functional activity and adverse effects of chemotherapy	The app allows patients to record their daily functional activity and perceived symptoms during chemotherapy with indications of severity. Patients can edit a list of their preselected symptoms or select any of the 48 symptoms available.
2018, Graetz et al, United States [25]	To evaluate the feasibility of a Web-based symptom-reporting app for patients with early-stage breast cancer using AIs[b]	Improvement in symptom burden and medication adherence	The ability to report symptoms and AI medication use, with built-in alerts sent to a patient's care team on the basis of the predetermined thresholds.
2016, Hwang, Canada [22]	To determine if unscheduled visits for care and hospital readmission can be prevented by e-monitoring and to assess patient satisfaction with the app	Provision of care for postoperative wounds	The app allows for electronic wound monitoring. The patient takes photos of the wound on postoperative days 1, 3, 7, and 14 and attaches them to electronic messages sent to the surgeon, who must answer within 24 hours.
2010, Klasnja et al, United States [24]	To refine the functional requirements of a mobile app to assist patients with cancer during treatment	Provision of health information to manage care-related issues in unanchored settings	The app has modules including daily check-ins to track well-being and symptoms; calendar events (eg, consultations with clinicians); logs to monitor medications, pain, and surgery drains; and notes (ie, text, photo, and audio) for quick capture of care-related information.
2017, Liu et al, China [23]	To develop and evaluate the structure and contents of a smartphone app for women with breast cancer	Provision of information support regarding disease, treatment, medication, exercise, nutrition, symptoms, examination, and social support	The app has 5 main function modules: personalized information recommendation, category knowledge center, headline information browsing, newest information browsing, and information searching.
2016, Young-Afat et al, The Netherlands [26]	To evaluate patient experience and satisfaction, physicians' and nurses' opinions, and scientific potential of a supportive breast cancer app	To be beneficial in clinical practice and research	The app has 4 main functionalities: repository for information (audio recorded and imaging), symptom registration, timeline of treatment trajectory, and personalized information about breast cancer and treatment.
2017, Zhu et al, China [19]	To develop and evaluate the content and functionality of a mobile app for women with breast cancer undergoing chemotherapy	Provision of social, emotional, and information support	The app has 4 components: learning (information related to breast cancer and symptom management), discussion (anonymous support group), ask the expert (online consultation), and personal stories (stories of breast cancer survivors).
2018 A, Zhu et al, China [20]	To explore participants' perceptions of the strengths and weaknesses of the BCS[c], and their suggestions for program improvement	Provision of social, emotional, and information support	The app has 4 components: learning (information related to breast cancer and symptom management), discussion (anonymous support group), ask the expert (online consultation), and personal stories (stories of breast cancer survivors).

Year, author, country[a]	Study characteristics	Intervention characteristics	
	Objective	Purpose of the app	Operation
2018 B, Zhu et al, China [21]	To determine the effectiveness of the BCS program to address women's self efficacy, symptoms, and quality of life during chemotherapy	Provision of social, emotional, and information support	The app has 4 components: learning (information related to breast cancer and symptom management), discussion (anonymous support group), ask-the expert (online consultation), and personal stories (stories of breast cancer survivors).

[a]Country of the study coordinator.

[b]AIs: aromatase inhibitors.

[c]BCS: Breast Cancer e-Support Program.

Table 3. Summary of interventions and outcome characteristics of included papers (n=9).

Year, author, country[a]	Intervention characteristics		Outcome characteristics	
	Description (developer)	Operating system	Primary outcomes	Main conclusions
2016, Egbring et al, Switzerland [18]	Mobile app to record daily functional activity and adverse effects of chemotherapy	iOS[b] and Android	Functional activity and adverse effects of chemotherapy	Patient well-being and reporting of the adverse effects of chemotherapy can be improved by using a mobile app under the supervision of the treating physician.
2018, Graetz et al, United States [25]	App that allows patients to share information in real time with their cancer care team outside of clinic visits	Information not available	Symptom burden and medication adherence	The use of an app with weekly reminders significantly improved short-term AI[c] adherence, which may reduce the symptom burden of women with breast cancer.
2016, Hwang, Canada [22]	Smartphone app that allows for communication between the patient and the surgeon (Medeo)	Information not available	Unscheduled visits for care, hospital readmission, and patient satisfaction	Electronic wound monitoring was associated with significantly less unscheduled care, including hospital readmission and visits to the emergency department or walk-in clinic, a high degree of patient satisfaction, and a possible reduction in cost to the health care system.
2010, Klasnja et al, United States [24]	Mobile app to assist patients in managing care-related information (HealthWeaver Mobile)	Android only	Participants' perceptions of HealthWeaver Mobile	The possibility of taking photos and audio notes was highly valued by participants. The app was seen not only as a valuable way to capture information quickly but also as a means of accessing information, especially through calendar events.
2017, Liu et al, China [23]	Smartphone app to provide personalized information support (Information Assistant)	Information not available	Participants' evaluation of Information Assistant	A few useful pieces of information, photos, and videos have been added, allowing patients to gain maximum benefits from the app. It is able to deliver high-quality information support.
2016, Young-Afat et al, The Netherlands [26]	Supportive breast cancer mobile app to assist in clinical practice and research (OWise)	iOS and Android	Participants' evaluation of OWise	Benefits for patients and their medical teams, especially because of the option to make audio recordings of consultations and the availability of personalized information.
2017, Zhu et al, China [19]	Mobile app to promote women's self-efficacy and social support (BCS[d])	iOS and Android	Participants' evaluation of BCS	More information has been added in the app tutorial, as well as the information related to food choices, sexual activity, pregnancy, and the interpretation of laboratory results, making the app useful, attractive, and easy to use.
2018, Zhu et al, China [20]	Mobile app to promote women's self-efficacy and social support (BCS)	iOS and Android	Participants' perceptions of BCS	Potential of BCS to support women during chemotherapy. Its benefits can be maximized by incorporation into routine care.

Year, author, country[a]	Intervention characteristics		Outcome characteristics	
	Description (developer)	Operating system	Primary outcomes	Main conclusions
2018, Zhu et al, China [21]	Mobile app to promote women's self-efficacy and social support (BCS)	iOS and Android	Self-efficacy	The BCS program demonstrated its potential as an effective and easily accessible intervention to promote women's self-efficacy, symptom interference, and quality of life during chemotherapy.

[a]Country of the study coordinator.

[b]iOS: iPhone Operating System.

[c]AIs: Aromatase Inhibitors.

[d]BCS: Breast Cancer e-Support Program.

Risk of Bias Within Studies

On the basis of the methodological quality assessment using MINORS [17], the total scores of the validation studies ranged from 12 to 16 [19,20,23,24,26], while the nonrandomized comparative study reached a total score of 21 [22], as shown in Table 4.

On the basis of the methodological quality assessment using the Cochrane Collaboration Risk of Bias Tool [16], one of the studies [21] presented a low risk of bias in all domains evaluated. Only one of the studies [18] presented high risk in 2 domains, both regarding blinding; the authors indicated that the absence of blinding may have affected the outcomes evaluated. In one study, there were no reports regarding blinding of outcome assessment [26], as shown in Figure 2.

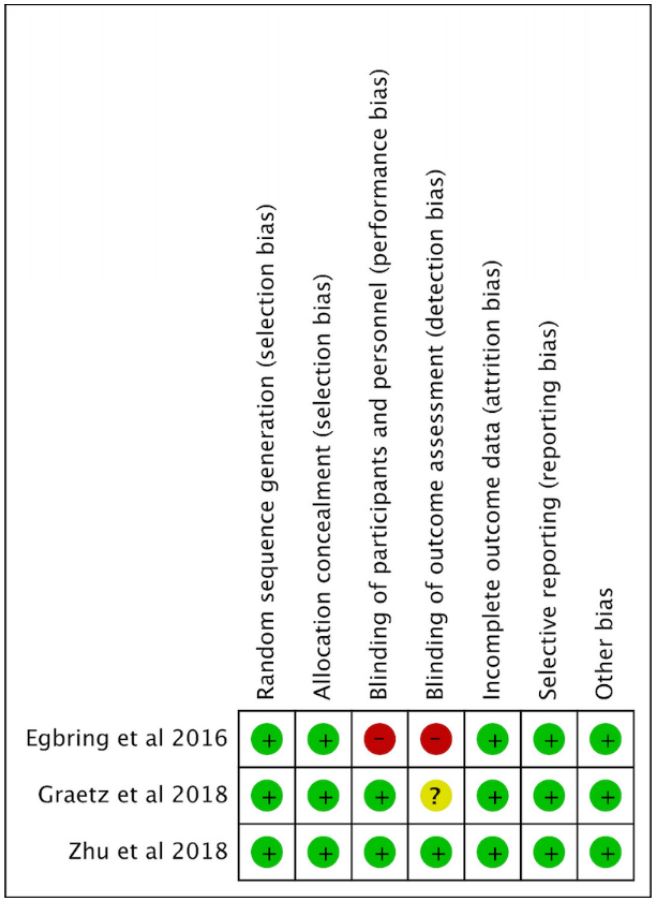
Table 4. Methodological appraisal of selected studies on the basis of Methodological Index for Nonrandomized Studies (MINORS).

Criteria[a]	Hwang, 2016 [22]	Klasnja et al 2010 [24]	Liu et al 2017 [23]	Young-Afat et al 2016 [26]	Zhu et al 2017 [19]	Zhu et al 2018 [20]
1. A clearly stated aim	2	2	2	2	2	2
2. Inclusion of consecutive patients	2	2	2	2	2	0
3. Prospective collection of data	2	2	2	2	2	2
4. Endpoints appropriate to aim of study	2	2	2	2	2	2
5. Unbiased assessment of study endpoint	2	2	2	2	2	2
6. Follow-up period appropriate for aim of study	2	2	2	2	0	2
7. Loss to follow-up less than 5%	2	0	2	0	2	2
8. Prospective calculation of study size	1	0	2	2	2	2
9. An adequate control group	2	N/A[b]	N/A	N/A	N/A	N/A
10. Contemporary groups	1	N/A	N/A	N/A	N/A	N/A
11. Baseline equivalence of groups	1	N/A	N/A	N/A	N/A	N/A
12. Adequate statistical analyses	2	N/A	N/A	N/A	N/A	N/A
Total score	21	12	16	14	14	14

[a]Items were scored 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate). The global ideal total score is 16 for noncomparative studies and 24 for comparative studies. See list of references for full source information on papers. MINORS index is described in Slim et al [14].

[a]Not applicable.

Figure 2. Methodological appraisal of selected studies on the basis of Cochrane Collaboration Risk of Bias Tool.



Synthesis of Results

All the analyzed studies carried out the evaluation of mobile apps for women during the treatment of breast cancer. On the basis of the objective, different types of apps were developed. In general, all involved the provision of useful and quality information to patients as a way of improving management of adverse treatment effects through the promotion of self-care at home.

Some studies reported the importance of using the specific app under the supervision of health professionals, either online [19,22,25] or in person [18]. Two studies evaluated an app capable of communicating with health care professionals, providing remote electronic monitoring [22,25]. One study developed a pilot randomized controlled trial to assess the feasibility of an app with or without weekly reminders, allowing the sharing of health information on a real-time basis with the patient's oncology care team. Participants who used the weekly reminders feature had a higher app usage rate (74% vs 38%, $P<.05$) during the intervention and reported higher drug adherence than those who did not opt for this feature (100% vs 72%, $P<.05$) [25].

Another study reported a significantly lower number of unexpected medical consultations, including hospital readmissions and visits to the emergency department, in the e-monitoring group than the conventional follow-up control group (3% vs 22%, $P<.05$), as the app allows the professional to conduct an electronic consultation if necessary, avoiding complications. Almost all e-monitoring patients felt that the app led to improved care (95%) and would recommend it to a friend or colleague (90%).

Three studies in this systematic review [19-21] concerned an app with an *ask the expert* module, capable of providing online consultations, which was found to be useful, attractive, and easy to use [19]. The authors developed and evaluated the content and functionality of the mobile app [19], tested it to determine its effectiveness regarding women's self efficacy, symptoms, and quality of life during chemotherapy [21], and explored participants' perceptions of app strengths and weaknesses [20].

Through qualitative interviews, it emerged that the participants considered the app to be useful for improving knowledge and promoting emotional well-being and would recommend it to other women undergoing chemotherapy [20]. In addition, when tested, the app had significantly better health outcomes at 3 months regarding self-efficacy (21.05; 95% CI 1.87 to 40.22; $P=.03$; $d=0.53$), symptom interference (−0.73; 95% CI −1.35 to −0.11; $P=.02$; $d=-0.51$), and quality of life (6.64; 95% CI 0.77 to 12.50; $P=.03$, $d=0.46$) compared with participants who received usual care [21].

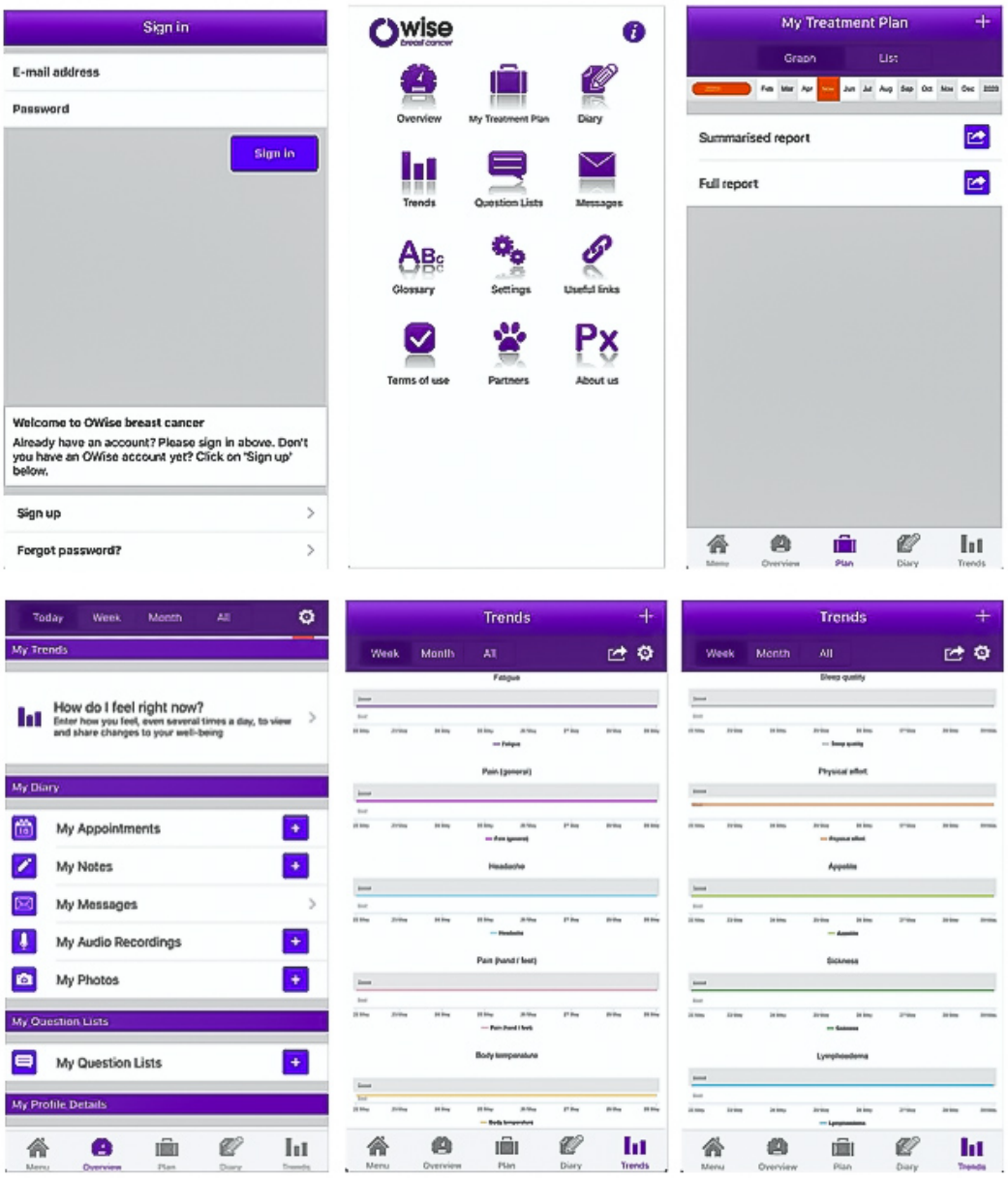
Another study reported the importance of an app for patients who require face-to-face follow-up with health care professionals on scheduled visits. Through the app, the patient can make a recording of the perceived symptoms at home, improving the report of the adverse effects management for the health care professional during consultation [18].

The focus of some studies was on using mobile apps to improve the retention of information about the management of adverse effects and, consequently, to improve self-care and patient well-being. One study evaluated an app with a module dedicated to the latest research findings, technologies, and methods for breast cancer care, thereby helping patients obtain up-to-date information [23].

Moreover, one study assessed a mobile app that, in addition to offering patient guidance through the provision of knowledge, includes a calendar to highlight important dates related to events, treatments, and consultations. This app incorporates texts, photos, and audio, making it possible for patients to capture and store important information related to their care. This app allows greater interaction between the patient and the app, a fact that was valued by the study participants [24].

However, this was not the only study on an app enabling the inclusion of audio. In another study, all patients (n=15) used the audio recording function to record consultations with their nurses and physicians, and 14 (93%) patients found this feature useful. Patients considered the audio aspect and the personalization of information about disease and treatment as the most useful features of the app. Almost all physicians and nurses (90%) also found the recording function useful and would recommend the app to their patients [26]. Figure 3 shows screenshots of this smartphone app, named OWise.

Figure 3. Screenshots of the OWise.



Discussion

Summary of Evidence

This is a systematic review about the available evidence regarding women's use of mobile apps during breast cancer treatment. A total of 9 papers, which covered the development of various types of mobile apps, were included [18-26]. The apps encompassed the transmission of useful information and the management of the adverse effects of breast cancer treatment through various features.

While there is growing interest in using patient reporting tools to improve symptom monitoring during cancer treatment, there is a lack of evidence about the impact of this strategy on clinical outcomes [26,27]. A mobile app has the potential to restore the daily functional activity of patients with early-stage breast cancer. This benefit is most significant when the patient uses the app under the supervision of health professionals as the review of the treatment during consultation is an opportune moment for a sincere discussion of symptoms reported by the patients in the app [18].

Electronic follow-up via apps may help avoid unplanned visits to the emergency room and hospital readmissions, reducing health care costs [22]. These results corroborate those of a study that examined health-related quality of life, emergency room visits, and hospitalizations among patients receiving chemotherapy for advanced solid tumors. The use of tablet computers to self-report symptoms during cancer treatment was associated with significant clinical benefits, decreased emergency room admissions, and reduced hospitalizations [27].

Web-based self-management support systems and apps involving health professionals are important to ensure the communication of high-quality information between patients and the health team empowering the patients to increase their self-care and improve their own health [28] and also the communication between patients and staff outside of health institutions as a way of ensuring continuity of care [22], enhancing knowledge, improving confidence level, and promoting emotional well-being [20]. Besides that, the involvement of health professionals in monitoring treatment effects through an app outside of clinic visits also might be a cost-effective way to improve symptom management and health outcomes [25].

Calendar events, photos, messages, and audio are a few of the mechanisms that can increase the usability of apps and improve the acquisition of information related to illness and treatment [18,24,26]. These features may improve self-care at home and restore the daily functional activity of patients. In addition, the audio recording function to record consultations with health professionals does not increase the duration of the consultation [26].

Characteristics related to communication and interaction were also demonstrated by a recent study that aimed to investigate the behavior of patients with prostate cancer when using a mobile app during radiation therapy [29]. As found in the studies included in this systematic review, there was improvement in the reporting and self-management of symptoms during treatment. In addition, the app provided continuous access to links to important information and, like one of the apps evaluated by a study included in this review, helped patients obtain up-to-date information regarding their disease and treatment [23].

In this context, health apps are increasingly being used in clinical care and might have significant informative potential. However, they are often inserted into clinical care before the necessary research to confirm the benefits for patients and health professionals has been conducted [26].

Limitations

Some methodological limitations of this systematic review should be taken into account. Most studies validated apps for mobile devices during their development, focusing on their content and functionality. More data from randomized clinical trials are needed to assess the effects of using mobile apps during the treatment of women with breast cancer, as is being done with the Breast Cancer e-Support program. In addition, the evaluated apps have different functionalities and are geared toward different treatments. Therefore, their heterogeneity prevents the data from being satisfactorily grouped, hindering a quantitative analysis.

Conclusions

The evidence from the studies included in this systematic review is currently limited but suggests that mobile apps for women with breast cancer might be an acceptable information source and lead to improved patient well-being. They can also be used to report symptoms and adverse treatment-related effects and promote self-care. However, the real utility of mobile apps for women undergoing breast cancer treatment is still uncertain. More evidence-based apps must be tested in future randomized clinical trials.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategies.
[PDF File (Adobe PDF File), 58KB - mhealth_v7i8e13245_app1.pdf]

Multimedia Appendix 2
Full articles excluded (n=16) from review with reasons.

[PDF File (Adobe PDF File), 81KB - mhealth_v7i8e13245_app2.pdf]

References

1. 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]
2. de Oliveira Santos M. National Cancer Institute. 2018. Estimate 2018: Cancer Incidence in Brazil URL: http://www1.inca.gov.br/rbc/n_64/v01/pdf/15-review-estimate-2018-cancer-incidence-in-brazil.pdf [accessed 2019-08-07]
3. DeVita Jr VT, Lawrence TS, Rosenberg SA, editors. DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology. Tenth Edition. Philadelphia, Pennsylvania: LWW; 2015.
4. Kearney N, Kidd L, Miller M, Sage M, Khorrami J, McGee M, et al. Utilising handheld computers to monitor and support patients receiving chemotherapy: results of a UK-based feasibility study. Support Care Cancer 2006 Jul;14(7):742-752. [doi: 10.1007/s00520-005-0002-9] [Medline: 16525792]
5. Kinnane N, Thompson L. Evaluation of the addition of video-based education for patients receiving standard pre-chemotherapy education. Eur J Cancer Care (Engl) 2008 Jul;17(4):328-339. [doi: 10.1111/j.1365-2354.2007.00846.x] [Medline: 18652000]
6. Davoodi S, Mohammadzadeh Z, Safdari R. Mobile phone based system opportunities to home-based managing of chemotherapy side effects. Acta Inform Med 2016 Jun;24(3):193-196 [FREE Full text] [doi: 10.5455/aim.2016.24.193-196] [Medline: 27482134]
7. dos Anjos R, Reis C, Cláudio A, Neto D. O Laboratório de Bancos de Dados. 2014. Proposta Para Teste de Usabilidade Para Aplicações Móveis no Contexto de Computação Ubíqua URL: http://www.lbd.dcc.ufmg.br/colecoes/wtdqs/2014/002.pdf [accessed 2019-08-07]
8. Food and Drug Administration. 2015. Mobile Medical Applications. Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/medical-devices/digital-health/mobile-medical-applications [accessed 2019-08-07]
9. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
10. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - what's the evidence? Breast 2014 Oct;23(5):683-689. [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
11. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
12. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. JMIR Mhealth Uhealth 2017 Dec 4;5(12):e187 [FREE Full text] [doi: 10.2196/mhealth.8741] [Medline: 29203459]
13. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
14. 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]
15. Cruz F, Reis PE, Ferreira EB, Melo N. PROSPERO - University of York. 2018. Mobile Apps Related to the Treatment of Breast Cancer: A Systematic Review URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018083548 [accessed 2019-08-07]
16. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. Hoboken, New Jersey: Wiley-Blackwell; 2011.
17. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003 Sep;73(9):712-716. [doi: 10.1046/j.1445-2197.2003.02748.x] [Medline: 12956787]
18. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 6;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
19. Zhu J, Ebert L, Xue Z, Shen Q, Chan SW. Development of a mobile application of breast cancer e-support program for women with breast cancer undergoing chemotherapy. Technol Health Care 2017;25(2):377-382. [doi: 10.3233/THC-161292] [Medline: 28085020]
20. Zhu J, Ebert L, Guo D, Yang S, Han Q, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 1): qualitative study of women's perceptions. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e85 [FREE Full text] [doi: 10.2196/mhealth.9311] [Medline: 29643056]
21. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]

22. Hwang H. Electronic wound monitoring after ambulatory breast cancer surgery: improving patient care and satisfaction using a smart phone app. B C Med J 2016;58(8):448-453 [FREE Full text]
23. Liu Y, Geng Z, Wu F, Yuan C. Developing 'information assistant': a smartphone application to meet the personalized information needs of women with breast cancer. Stud Health Technol Inform 2017;245:156-160. [doi: 10.1177/1357633X15594753] [Medline: 29295072]
24. Klasnja P, Hartzler A, Powell C, Phan G, Pratt W. Health weaver mobile: designing a mobile tool for managing personal health information during cancer care. AMIA Annu Symp Proc 2010 Nov 13;2010:392-396 [FREE Full text] [Medline: 21347007]
25. Graetz I, McKillop CN, Stepanski E, Vidal GA, Anderson JN, Schwartzberg LS. Use of a web-based app to improve breast cancer symptom management and adherence for aromatase inhibitors: a randomized controlled feasibility trial. J Cancer Surviv 2018 Aug;12(4):431-440 [FREE Full text] [doi: 10.1007/s11764-018-0682-z] [Medline: 29492753]
26. Young-Afat DA, van Gils CH, Bruinvels DJ, van der Pol CC, Witkamp AJ, Sijtsema S, et al. Patients' and health care providers' opinions on a supportive health app during breast cancer treatment: a qualitative evaluation. JMIR Cancer 2016 Jun 7;2(1):e8 [FREE Full text] [doi: 10.2196/cancer.5334] [Medline: 28410170]
27. 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. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
28. Ekstedt M, Børøsund E, Svenningsen IK, Ruland CM. Reducing errors through a web-based self-management support system. Stud Health Technol Inform 2014;201:328-334. [doi: 10.3233/978-1-61499-415-2-328] [Medline: 24943563]
29. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Nyman MH, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]

Abbreviations

MINORS: Methodological Index for Non-Randomized Studies

Edited by G Eysenbach; submitted 27.12.18; peer-reviewed by P Santana-Mancilla, J Shenson, F Alahdab, D McDonnell; comments to author 14.03.19; revised version received 17.05.19; accepted 31.05.19; published 27.08.19.

Please cite as:
Cruz FOAM, Vilela RA, Ferreira EB, Melo NS, Reis PEDD
Evidence on the Use of Mobile Apps During the Treatment of Breast Cancer: Systematic Review
JMIR Mhealth Uhealth 2019;7(8):e13245
URL: http://mhealth.jmir.org/2019/8/e13245/
doi:10.2196/13245
PMID:31456578

Original Paper

RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices

Yatharth Ranjan[1], MSc; Zulqarnain Rashid[1], PhD; Callum Stewart[1], MSc; Pauline Conde[1], BSc; Mark Begale[2], MSc; Denny Verbeeck[3], MSc; Sebastian Boettcher[4], MSc; The Hyve[5]; Richard Dobson[1,6], PhD; Amos Folarin[1,6], PhD; The RADAR-CNS Consortium[7]

[1]The Institute of Psychiatry, Psychology & Neuroscience (IoPPN), Department of Biostatistics & Health Informatics, King's College London, London, United Kingdom

[2]Vibrent Health, Fairfax, VA, United States

[3]Janssen Pharmaceutica NV, Turnhoutseweg, Beerse, Belgium

[4]Epilepsy Center, Department of Neurosurgery, University of Hospital Freiburg, Freiburg, Germany

[5]The Hyve, MJ Utrecht, Netherlands

[6]Institute of Health Informatics, University College London, London, United Kingdom

[7]The RADAR-CNS Consortium, London, United Kingdom

Corresponding Author:
Amos Folarin, PhD
The Institute of Psychiatry, Psychology & Neuroscience (IoPPN)
Department of Biostatistics & Health Informatics
King's College London
SGDP Centre, IoPPN, PO Box 80 De Crespigny Park, Denmark Hill
London, SE5 8AF
United Kingdom
Phone: 44 02078480924
Email: amos.folarin@kcl.ac.uk

Abstract

Background: With a wide range of use cases in both research and clinical domains, collecting continuous mobile health (mHealth) streaming data from multiple sources in a secure, highly scalable, and extensible platform is of high interest to the open source mHealth community. The European Union Innovative Medicines Initiative Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) program is an exemplary project with the requirements to support the collection of high-resolution data at scale; as such, the Remote Assessment of Disease and Relapse (RADAR)-base platform is designed to meet these needs and additionally facilitate a new generation of mHealth projects in this nascent field.

Objective: Wide-bandwidth networks, smartphone penetrance, and wearable sensors offer new possibilities for collecting near-real-time high-resolution datasets from large numbers of participants. The aim of this study was to build a platform that would cater for large-scale data collection for remote monitoring initiatives. Key criteria are around scalability, extensibility, security, and privacy.

Methods: RADAR-base is developed as a modular application; the backend is built on a backbone of the highly successful Confluent/Apache Kafka framework for streaming data. To facilitate scaling and ease of deployment, we use Docker containers to package the components of the platform. RADAR-base provides 2 main mobile apps for data collection, a Passive App and an Active App. Other third-Party Apps and sensors are easily integrated into the platform. Management user interfaces to support data collection and enrolment are also provided.

Results: General principles of the platform components and design of RADAR-base are presented here, with examples of the types of data currently being collected from devices used in RADAR-CNS projects: Multiple Sclerosis, Epilepsy, and Depression cohorts.

Conclusions: RADAR-base is a fully functional, remote data collection platform built around Confluent/Apache Kafka and provides off-the-shelf components for projects interested in collecting mHealth datasets at scale.

(JMIR Mhealth Uhealth 2019;7(8):e11734) doi:10.2196/11734

KEYWORDS

remote sensing technology; mobile applications; telemedicine; mental health

Introduction

Background

The opportunity in health care for continuous monitoring of patients has steadily grown in parallel with the widespread availability of smartphones, more capacious mobile networks, and development of new wearable sensors that are able to continuously measure a growing set of physiological and phenomenological parameters. Many of these devices are currently in the lifestyle or fitness domain; however, vendors are increasingly developing these devices for medical-grade applications. If these streams of data can be reliably collected, analyzed, and acted on, it opens up the possibility of better understanding disease etiology, diagnosis, prognosis, and detecting relapse in most disease areas.

Existing mobile health (mHealth) platforms include some form of questionnaires, phone (Android, iOS) sensor data collection, wearables integration, backend infrastructure, and management user interface, but few presently include all or are scalable solutions. Moreover, 2 such examples are the Open source AWARE Framework, an Android platform for mobile phone–based context sensing [1], and the Health Insurance Portability and Accountability Act of 1996-compliant Bridge platform, which supports biomedical studies conducted through smartphones [2].

Objectives

The Euro 26 million Innovative Medicines Initiative (IMI) Remote Assessment of Disease and Relapse -Central Nervous System (RADAR-CNS) is a major international academic-industry research program aimed at developing novel methods and infrastructure for monitoring major depressive disorder (MDD), epilepsy (Epi), and multiple sclerosis (MS) using wearable devices and smartphone technology [3]. Beyond supporting the initial goals of the 3 disorder areas in RADAR-CNS, the RADAR-base platform aims to provide a highly extensible platform for mHealth applications [4].

To facilitate adoption by the wider mHealth community, the RADAR-base platform was released under an open source Apache 2 license in January 2018. RADAR-base is composed of backend infrastructure and 2 Android mobile apps: a cross-platform Cordova app for active monitoring of participants (active remote monitoring technology, aRMT) through conscious action (eg, questionnaires, audio questions, timed tests) and a native Android app for passive monitoring via phone and wearable sensors (passive remote monitoring technology, pRMT). RADAR-base also includes capabilities for data aggregation, management of studies, and real-time visualizations.

A key differentiator of RADAR-base platform is that it makes use of Confluent technologies (based around Apache Kafka) to provide an end-to-end solution for remote monitoring use cases (eg, participant management and data analysis) that scales horizontally through the use of Kafka and Confluent ecosystem. Other approaches to centralize information flow and decouple systems exist, in particular Messaging Queues and Enterprise Service Bus/Service Oriented Architecture (SOA) type architectures, which have a number of overlapping and differentiating factors compared with Confluent/Kafka, in particular around routing, scaling, performance, and ecosystem [5,6]. Kafka and the Confluent ecosystem have become the de facto industry standard for high-throughput event streaming data applications. More specifically, it is appropriate for building real-time streaming/transforming data pipelines that reliably move data between systems at scale. The RADAR-base platform has a number of requirements: first, data must flow from sensors or data sources into the platform through mobile devices and second, to transform these data (eg, restructured for Cold Storage, Hot Storage, aggregating data based on time windows and monitor the data coming in. This Kafka based pipeline, along with data schematization makes the platform flexible and open to a variety of devices, types of data, velocity, and throughput.

The RADAR-base platform can be deployed both in local settings, such as a hospital, for local data collection or for ambulatory studies through remote deployment for centralized data collection. The RADAR-base backend has been deployed on various platforms (cloud, bare-metal, etc) as a set of microservices using Docker containers. Both scenarios are used in RADAR-CNS.

The RADAR-base platform is a scalable, secure, open source Internet of Things (IoT) platform for real-time remote sensor data collection in the context of mHealth clinical studies.

Methods

Components

The RADAR-base platform consists of following major categories of components:

1. Data Collection Ecosystem
2. Data sources
3. Data Processing and Visualization
4. Study management and Security

Figures 1 and 2 show the integration and communication between different components of the RADAR-base platform.

Figure 1 shows the Technical Overview of the RADAR-base Platform Stack Platform Capabilities:

1. High throughput, low latency data collection
2. Scalability
3. Generalized device integration for passive data sources
    1. Abstracted and composable integration mechanism for sensor devices

2. Third-Party RESTful (Representational State Transfer) data source integration

4. Configurable data sources at runtime
5. Schema evolution
6. Real-time data processing and analytics
7. Hot and cold storage
8. Data access (Representational State Transfer [REST]-API)
9. Modular, extensible dashboards
10. Electronic Case Report Form (eCRF) integration (REDCap)
11. Remote configuration
12. Cohort Management Portal
13. Security

Figure 2 shows an overview of the RADAR-base platform. Current data sources are as follows: Empatica E4, Pebble 2, Fitbit, Biovotion, Faros, Active aRMT Questionnaire app, and Passive pRMT app.

These functionalities are delivered through the following components:

1. Data ingestion: Recognizing and registering data sources (including smartphones and wearable devices), collecting the data via a direct Bluetooth connection or through a third-party application protocol interface (API), and streaming in near-real-time to the server (green box in Figure 1). Using Apache Kafka, the collected data are streamed to dedicated topics in real-time where the data are schematized using Apache AVRO and the Schema Registry. More details about data source mapping and integration can be found in the Data Sources and Study Management Sections below. Detailed further on the platform's documentation wiki [7].
2. Data storage and management: Consists of 2 centralized storage systems behind an authorized security layer. The cold storage, based on Hadoop Distributed File System (HDFS), that is scalable and fault-tolerant, focused on storing large volumes of raw data, and the hot storage, based on MongoDB, for storing aggregated data to provide a near real-time overview of the raw data, principally for the data dashboards.
3. Data sharing: Visualizing aggregated data in a live dashboard and exporting raw data for further analyses in various formats including AVRO, JSON, and CSV.

Figure 1. Technical overview of the RADAR-base platform stack.

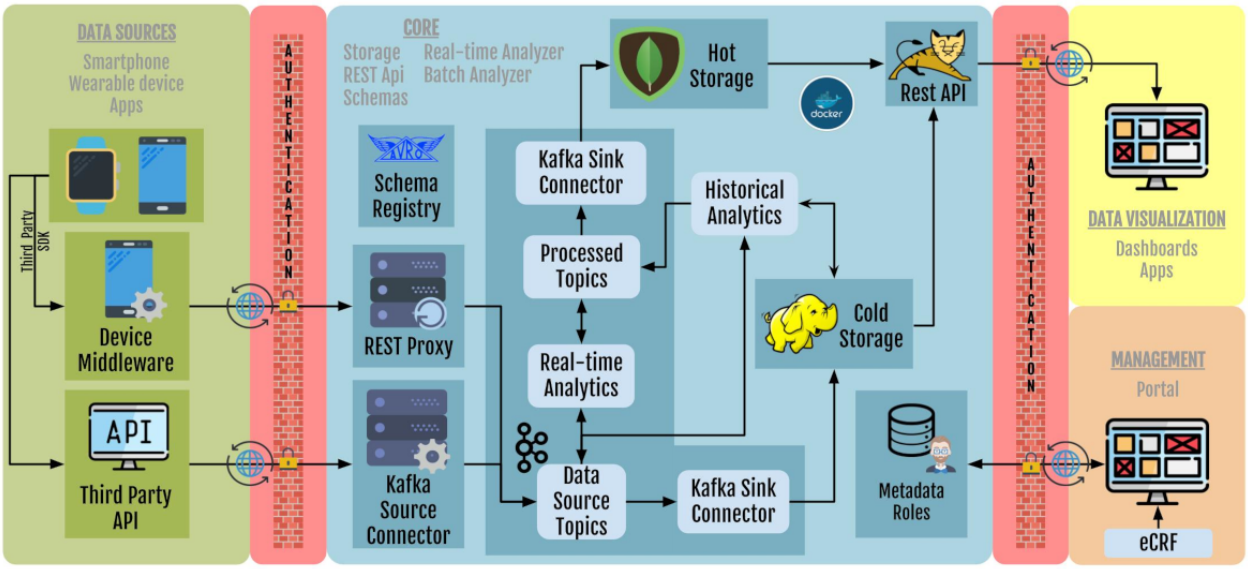
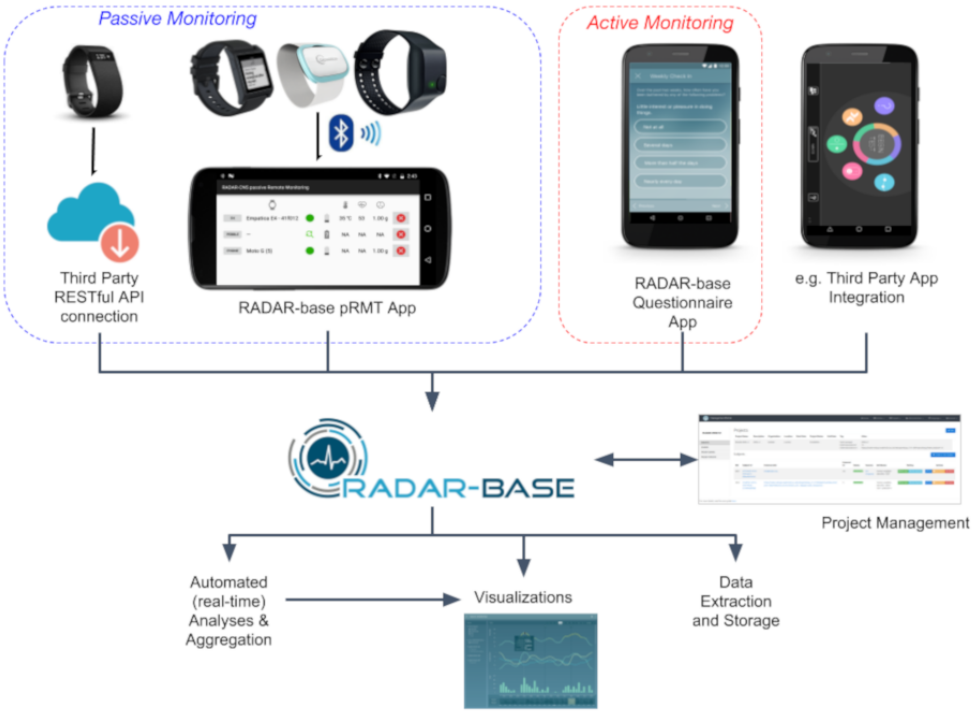


Figure 2. Current data sources: Empatica E4, Pebble 2, Fitbit, Biovotion, Faros, active Remote Monitoring Questionnaire app, and passive Remote Monitoring app.



Data Collection Ecosystem

The entire RADAR-base backend is deployable as a set of microservices based on Docker containers [8].

Representational State Transfer Proxy

As the platform is based on Apache Kafka, we can either send data directly into Kafka via a native Producer or using HTTP and the Confluent REST proxy.

Schema Registry

The data sent into the platform from the data source will be converted into AVRO format before going into Kafka topics. To convert our data to AVRO, the REST proxy needs to know the schema (or format) of the data being sent. These schemas are stored in the Schema Registry to reduce the payload size of each request.

Backend Streams and Monitors

Event-by-event stream processing is built on top of Kafka Streams. It provides an abstract layer to monitor and analyze streams of data and write aggregated/transformed data into Kafka topics. The data are *produced* (into Processed Topics) and *consumed* (from Data Source Topics) in Apache Avro format using the schema stored inside the Schema Registry. This capability provides real-time analytics and is used to trigger real-time interventions (eg, providing real-time data source statistics and sending notifications on source disconnection).

Sink Connectors

The data coming into Kafka is extracted into storage systems such as the HDFS and MongoDB with the help of Kafka Sink Connectors. The HDFS Sink Connector takes raw data coming into the system and deposits it into the HDFS storage in AVRO format (this time with the schema embedded, so the data are self-describing), which can be used for archival storage and historical analysis. The MongoDB sink connector takes aggregated data coming into Kafka from the Streams app's Processed Topics and deposits it into the MongoDB storage.

Data Typing

The AVRO schema is a JSON format specification of the fields and data types, which data values can hold. The schema itself can be embedded in the message or a reference held to the schema stored in a Schema Registry. AVRO is a particularly convenient format for managing schema evolution in RADAR-base, as schema changes can occur frequently and without warning, especially where third-party data sources are concerned.

For illustration purposes, the schema for the phone acceleration is shown in Figure 3.

Figure 3. Schema overview for the phone acceleration.

```
{
  "namespace": "org.radarcns.passive.phone",
  "type": "record",
  "name": "PhoneAcceleration",
  "doc": "Data from 3-axis accelerometer sensor with gravitational constant g as unit.",
  "fields": [
    { "name": "time", "type": "double", "doc": "Device timestamp in UTC (s)." },
    { "name": "timeReceived", "type": "double", "doc": "Device receiver timestamp in UTC (s)." },
    { "name": "x", "type": "float", "doc": "Acceleration in the x-axis (g)." },
    { "name": "y", "type": "float", "doc": "Acceleration in the y-axis (g)." },
    { "name": "z", "type": "float", "doc": "Acceleration in the z-axis (g)." }
  ]
}
```

Data Sources

Data Sources represent a wide variety of systems able to send data into the RADAR-base platform; these include devices containing sensors, mobile phones, questionnaires and digital games/assessments, and Web-APIs data portals.

In RADAR-base, passive data sources are collected via the pRMT app, active data sources via the aRMT app, and third-party data sources via the THINC-it app in RADAR-CNS.

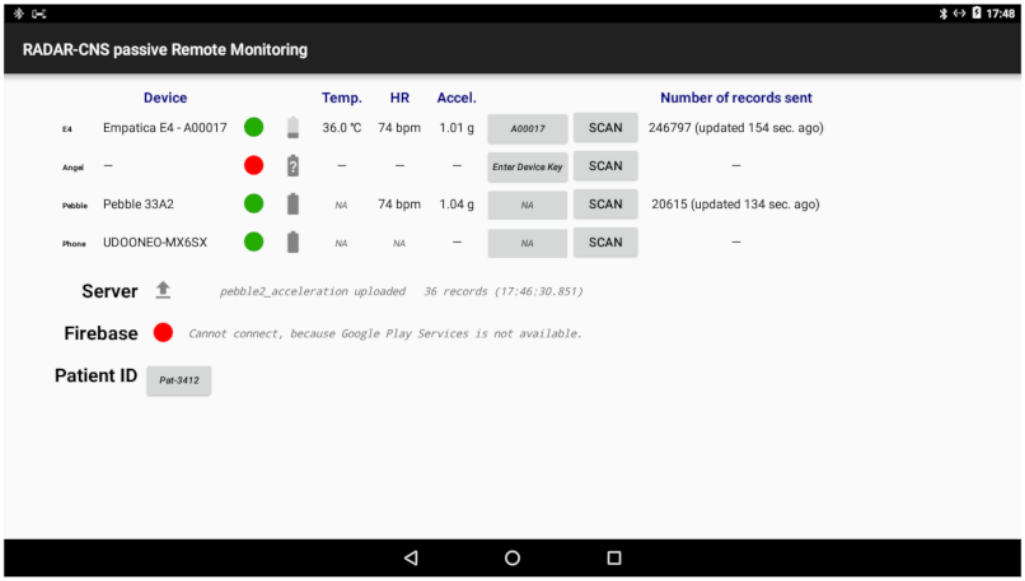
Another type of data source includes middleware connecting a vendor's Web API to the RADAR-base platform. For example, Fitbit does not provide a mobile Software Development Kit (SDK) to stream data to the pRMT app directly; instead, all the data are uploaded to the vendor data warehouse and provided to developers via a Web API. Getting these data into the RADAR system is achieved by implementing a server-side Kafka Source Connector, which continuously queries data from the vendor's Web API and dumps it into Kafka inside the RADAR-base platform; this approach can be used to integrate other Web API/OAuth2 data sources [9].

Passive Remote Monitoring App: Sensor Data Collection

The native passive Android application (pRMT) has been designed to passively collect data from sensors on the user's smartphone as well as to integrate wearable devices that offer SDKs. Its enhanced modularity (via pRMT plugins) allows easy integration of new devices/sensors. It currently supports Empatica E4 Wristband, Pebble 2 Smartwatch, and Biovotion VSM devices.

Figure 4 shows the pRMT app interface after login.

Figure 4. Passive Remote Monitoring app user interface. The Device column lists all the devices that are connected to the app and collect data. Device connection/disconnection is shown by green and red icons, respectively. The 3 columns next to "Device show the different values that are being measured on the connected devices. The last column shows the amount of data (or records) that have been collected.



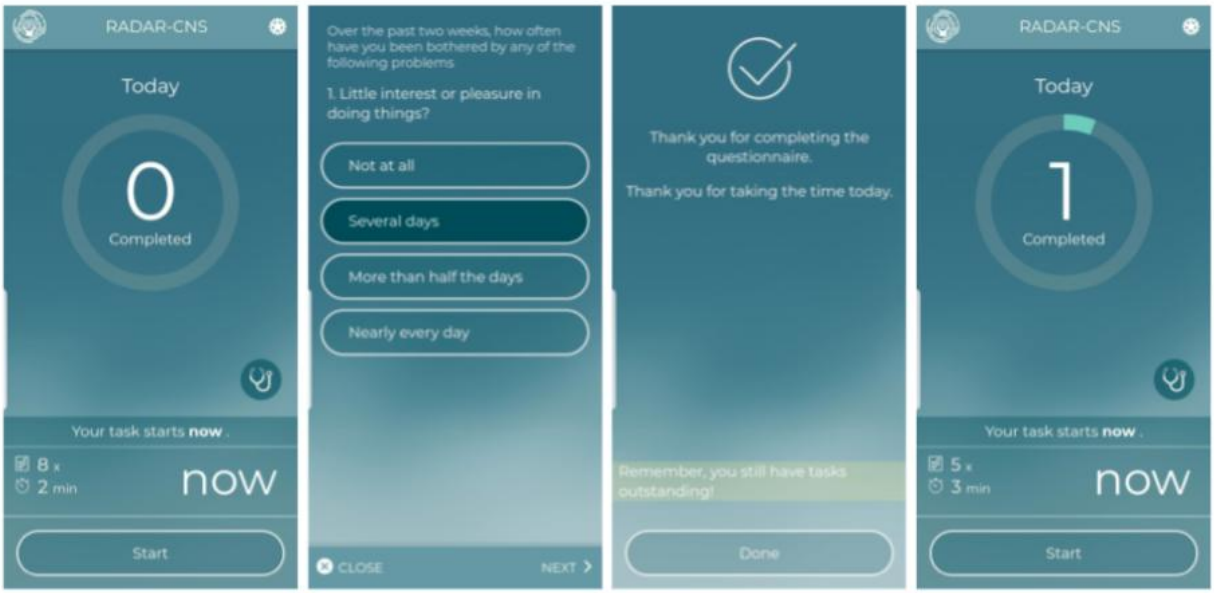
Active Remote Monitoring App: Composable Questionnaire Delivery

The primary goal for the aRMT mobile app is to allow users to submit questionnaires through the user's smartphone at a notified time. The questionnaire definitions and their regimen are defined by simple JSON configuration files, which are Web-served and therefore remotely configurable. New questionnaire configuration files can be easily created either manually or authored as REDCap data dictionaries and parsed

via a simple script. The regimen or protocol configuration file defines the sequence of the questionnaires delivered and the local notifications or Firebase Cloud Messaging push notifications used to alert the user.

The aRMT app was designed as a hybrid Cordova app and usable on both iOS and Android. Furthermore, it includes Cordova plugins to collect active audio responses to questions, allowing active samples of raw speech audio to be collected for analysis. Finally, the aRMT app also serves as a means of providing time markers to data collected in parallel by the pRMT app, such as start and end labels of walking and balance tests used in the MS study in RADAR-CNS. Figure 5 shows a selected interface of the aRMT app.

Figure 5. User interface of the active Remote Monitoring app.



Third-Party App Integration: THINC-it

THINC-it is a third-party mobile app that makes use of 5 quick interactive tests to assess memory, concentration, and attention [10].

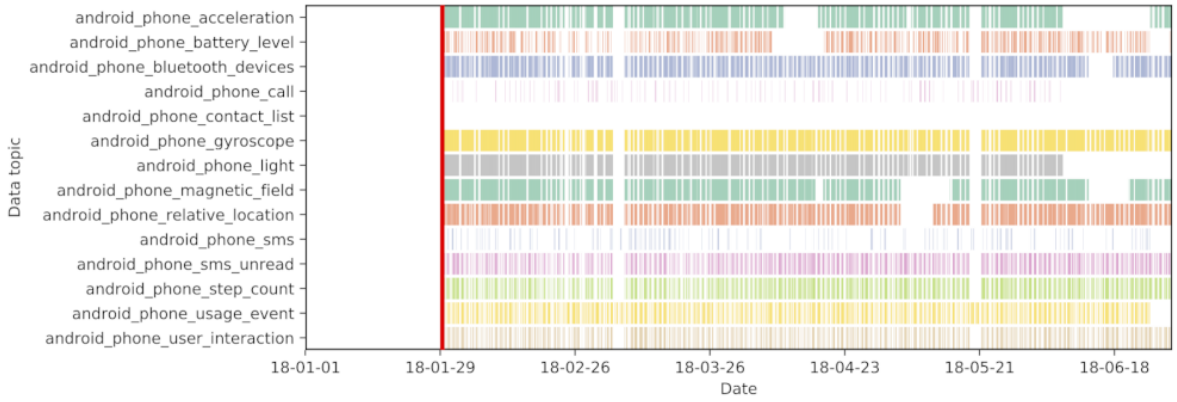
The THINC-it app uses the RADAR-base platform backend infrastructure as part of the RADAR-CNS project. It also provides a reasonable paradigm for other third-party app integration into the RADAR-base platform.

Data Processing and Visualization

Historic Raw Data Processing and Visualization

A common task is the exploration of collected raw data. In addition to the near-real-time visualization through the dashboard, the RADAR platform includes a python package for the processing and visualization of historic data. The package provides standard tooling for exploratory visualization of RADAR-base data (see Figure 6) and the easy implementation of preprocessing pipelines to take data exported from a RADAR-base project and output the processed data, with any accompanying labels, in a format suitable for use in standard machine learning libraries. Dask, a python library for parallel and larger-than-memory computing [11], is primarily used as the backbone. The potential for large longitudinal studies collecting high-frequency data necessitates the ability to distribute computation or to work on subsets. Moreover, the use of existing libraries allows straightforward integration with the rest of the python data science ecosystem.

Figure 6. Contiguity of phone sensor data over 6 months collected through RADAR-base for aparticipant in the major depressive disorder study. The red line corresponds to the enrollment date, whereas a coloured segment on each row corresponds to recorded data at an hourly resolution.



The package can make use of the common structure of RADAR-base data through the defined AVRO *schema* and *specification* for each modality. A schema describes data types, allowing automatic parsing of modalities stored without type. Each data source integrated into RADAR also contains a *specification* listing associated sensors with information on each. That information includes field names and units, allowing visualizations to be labeled automatically. As the specifications give the device and type of sensor for the associated data, those data can be mapped to relevant processing pipelines. For example, the Biovotion VSM1 device provides all its data in 1-second batches. These batches are ordered and so can be upsampled. Using the device specification, a function to upsample the data can be mapped to all the relevant data modalities. Alternatively, an analyst may wish to map a function to sensors of the same type across different devices, which is also made simple by using the specification to identify data from sensors of the same type.

Real-Time Data Processing and Representational State Transfer (RESTful)Application Protocol Interface

The RADAR platform exposes RESTful Services implemented using Jersey 2 and deployed on Grizzly server. Data collected in the platform are processed in real-time by a Kafka Streams application to provide aggregations (mean, max, etc) at various time resolutions (second, minute, hour, day, and week) and stored in MongoDB, which is served through the REST-API. The REST-API provides various API endpoints, which can be used to request the aggregated data in near-real-time and allows various combinations of queries. All the endpoints are secured; therefore, only a valid user or a client registered with the management portal with appropriate permissions/scopes can access these endpoints. All REST endpoints have been documented following OpenAPI specifications using Swagger: a powerful open source framework offering a large ecosystem of tools that help to design, build, document, and consume RESTful-APIs.

The REST-API also exposes some real-time information about the current status (connected or disconnected) of the sources registered for a particular subject along with when a data source was last detected sending data. All this information is used by the Dashboard to provide a real-time visualization of the current state of the studies/projects to project admins.

Real-Time Visualization Via Dashboards

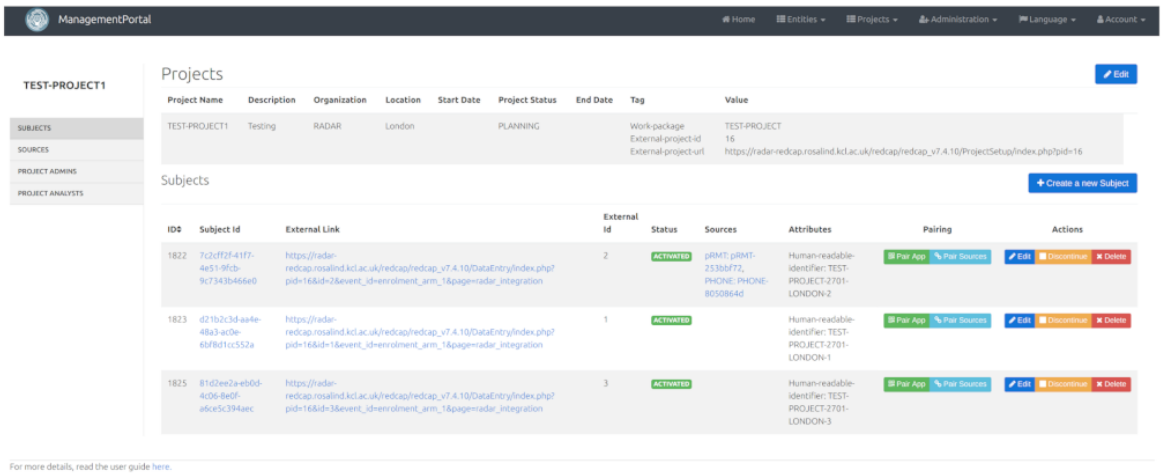
Overview and visualization are provided by a clear, customizable user interface with an emphasis on exploring different aggregation and zoom levels in the data. The RADAR-base dashboards use Angular, RxJS, and D3 to construct views on data from the REST-API. These presently provide management project/study lists, compliance views, and participant-level visualization of longitudinal data. Figure 7 shows an example data view. More detailed figures are given in the Multimedia Appendix 1.

Figure 7. Participant data view (battery and accelerometer streams).

Study Management and Security

The management portal Web application is the main user interface for creating and organizing RADAR projects, enrolling participants, and managing the association of participants with corresponding data sources. Figure 8 shows the management portal interface.

Figure 8. RADAR-base Management Portal.



Authentication, authorization, deidentification, and encryption are compulsory due to the sensitive information collected by the platform and to manage additional unknown risks associated with IoT using large numbers of network edge devices and endpoints. These are implemented for these following elements:

Quick Response Code Authentication

The management portal is used to issue a Quick Response (QR) code or Token for data-source to participant registration. This QR code can be scanned with the embedded QR code scanner in the integrated apps or alternatively a token can be entered directly as text. The decoded QR code provides some valuable information required by clients including sources, user ID, roles, scopes.

Open Authentication 2

To provide authorization and authentication, we utilize the OAuth2 workflow, an industry standard protocol for authorization. In the mobile apps, we use the Refresh Token grant type [12], and for other internal clients, we use the client credentials grant type. For instance, the gateway and the REST-API use the latter, whereas the mobile apps use the former.

For the mobile apps, an access token from the management portal is required for authorization across the platform. Without this token, data sources can neither register nor send data into the platform. The management portal provides a Refresh Token in the form of a QR code associated with a subject. This QR code can be scanned by the mobile apps to obtain a URL to a JSON Web Token, which embeds a Refresh Token and the authorization endpoint as a Bearer Token in an HTTP(S) request to obtain a new Refresh-token and Access token pair. This access token can then be used to access and post data according to the resources, roles, and scopes specified. Once the Access token expires, the most recently obtained Refresh Token is used to obtain a new (Refresh-token+Access token) pair.

The RADAR-base platform provides utilities for clients to easily manage the OAuth2 authorization flow [13]. An online manual is available, which specifies a step-by-step process of integrating new apps along with a list of utilities and libraries for QR code and OAuth [14]. The pRMT Android app uses OpenID Connect for authorization and authentication [15].

Deidentification

As discussed in eCRF Integration and REDCap Integration WebApp Sections, the strongly identifiable information is saved in a separate eCRF system, which is isolated from the RADAR-base platform [16]. Only the nonidentifiable data are saved in the platform. The 2 separated datasets are linked to each other as pseudonymized data only via RADAR-base Universally Unique ID.

Reverse Proxy

An nginx web server is used to proxy traffic into the platform and provide Cross-Origin Resource Sharing. Furthermore, the reverse proxy (nginx web server) can also be configured to act as a mitigation against Distributed Denial-of-Service attacks or using it as an HTTP load balancer.

Gateway

This component controls access to the Confluent Kafka REST Proxy for posting data from clients to Kafka through HTTP POST requests. It performs authentication and authorization, content validation, and decompression if needed. It also verifies if the access token sent in the HTTP POST request is valid and has the required privileges to perform the POST request for the specified resource, role, and scope.

Audit Log

The various components of RADAR-base keep activity logs at levels appropriate for that component. As the management portal keeps track of all study-related information, device assignments, and participant enrollment information, it keeps the most detailed audit logs. Any modification to the management portal database is stored in an audit record. These audit records store the user who made the modification, the time at which the modification was made, as well as the old and new state of the modified entity, allowing the complete history of all study metadata to be tracked or roll back modifications when necessary. Finally, the management portal also logs when, to what application, and for which user access tokens are being granted. It is important to note that this log is only there for the purposes of auditing. Validation of the tokens is not handled

by the management portal. Instead, clients can use the management portal's public key to validate the digital signature embedded in the access token. This way, components in RADAR-base can horizontally scale up, without the need for the management portal to scale up with them just to be able to keep up with validation requests. See Multimedia Appendix 1 for audit log figure.

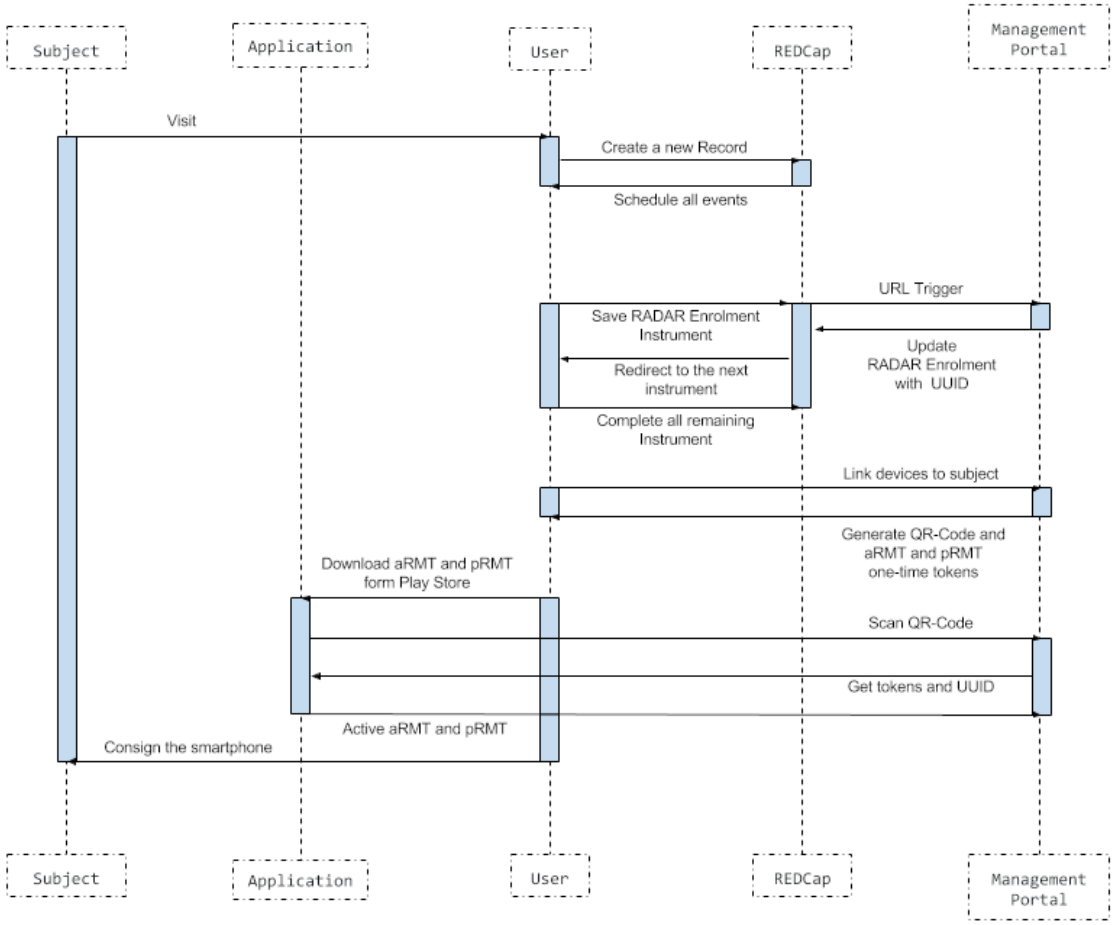
REDCap eCRF Integration

Optional integration of one or more REDCap eCRFs servers is provided with RADAR-base. REDCap is a secure 21 CFR Part 11, FISMA, and HIPAA-compliant Web application for building and managing online surveys and databases [17]. When used in this mode, subject creation is linked automatically between the REDCap and RADAR-base.

User Registration Workflow

A brief workflow of the registration is shown in Figure 9. A subject or recruiter will register a new record in REDcap (optional). Creating this record will trigger the creation of a corresponding subject in the management portal via RESTful-API calls. From the management portal subject, the project admin (recruiter) will be able to register apps via a QR code (aRMT and pRMT). Apps can be downloaded from the playstore, and registration QR codes can be obtained and scanned from the management portal for the particular subject. Once the apps get registered, the relevant subject will start streaming wearables and phone data.

Figure 9. User registration workflow.



Software Availability

The entire RADAR-base platform is freely available at github repository as open source software [18]. The details of the platform can be read at the official RADAR-base website [16]. A detailed quickstart, deployment details, and developer documentation are made available at Confluence Wiki [19]. The Docker images for all the components are available at Docker Hub [8].

Results

Multiple instances of RADAR-base are deployed and in use for real-world studies of Epilepsy, MS, and MDD under the umbrella of RADAR-CNS [20,21].

The catalog of devices currently integrated into the pRMT app include onboard Android smartphone sensors, Empatica E4, Pebble 2 smartwatch, BiovotionEverion, Faros 180, and Fitbit; a list is maintained here [22]. Pluggable capability is provided to integrate new wearable devices offering a native SDK (eg, Empatica E4) or through third-party vendors' REST-API (eg, Fitbit). The aRMT app provides highly extensible aRMT

functionality to the platform, rendering questionnaires from a JSON configuration file, for example, questionnaires using RADAR-CNS include RSES, PHQ8, and ESM. Figure 6 shows the contiguity of data collected from an MDD study participant's phone over a 6-month period. There are daily patterns visible even at this resolution; topics that rely on user interaction are typically not collected during the night, creating a striped pattern in the figure.

Discussion

Remote Assessment of Disease and Relapse-Base Current Deployments

The current deployments of RADAR-base for the different disorders are explained below.

Major Depression Disorder

The RADAR-base platform has been deployed centrally to collect active (questionnaires) and passively generated (wearable Fitbit and smartphone sensor) data remotely for participants recruited to 3 sites of the MDD study. The sites include King's College Hospital, London; Centro de Investigacion Biomedica en Red, Barcelona; and VU University Medical Center, the Netherlands. The objective being to collect regular self-reported symptoms and metrics such as sleep and ambulatory behavior. High-resolution data are being collected over a period of up to 2 years for each participant. More details about the MDD studies and preliminary data analysis are provided in our study [21].

Epilepsy

The Radar-base platform has been successfully tested and deployed in the Clinical Neurophysiology Department, King's College London, and the Epilepsy Center, Medical Center University of Freiburg, in their respective video electroencephalograms monitoring units, and it is currently in active use in London and Freiburg with enrolled participants.

Latest participants have the facility to wear 3 devices (Faros, Biovotion, and E4) concurrently. We have explained the detailed deployment of the platform for Epilepsy studies and initial collected data in our study [20].

Multiple Sclerosis

MS studies using RADAR-base are underway at different partner sites across Europe. Participant recruitment has been started and data are being streamed to central deployment. An important focus here is to collect data from the Faros 180 devices used for several mobility and balance tests in addition to similar ambulatory behavior collected in the depression study.

These 3 studies expose the versatility of the RADAR-base platform and generate data with very different complexity, volume, velocity, and durations.

Technical Challenges

Several technical challenges were addressed, including:

1. High throughput, volume, and velocity of the data.
2. Processing data in real time.
3. Optimizing phone resources to handle data collection and streaming (particularly high-resolution sensors).
4. Privacy concerns particularly around Global Positioning System data used to track location or audio exposing identifiable information. For this, the RADAR-base platform calculates and sends the relative location from a reference point. Similarly, background audio sampling is one-way-mapped to a vector representation of features using an OpenSmile plugin [23]; in this way, raw background audio is not exposed while retaining useful information content for analytics purposes.
5. Identifiable information is kept separate from sensor data to make it pseudonymized.
6. Security (authorization and authentication) is also a major concern for sensitive data collected from participants. Access to data is provided via a secure data transfer protocol.
7. With the huge amount of raw data the platform is built to collect, it is essential to have efficient compression. All the data collected and stored are compressed and encrypted.
8. Maintaining performance and behavior of the data source pRMT and aRMT apps in an ever-changing Android landscape of OS versions, handsets vendors, and form factors.

Comparison With Prior Work

A summary of other platforms comparing salient features with the RADAR-base platform are provided here. The recently developed mental health Nonintrusive Individual Monitoring Architecture platform, a prototype implementation used alongside an investigation of the key features required of a mHealth data collection platform; these include integrating data sources, a focus on privacy, and flexible user permissions [24]. Intel's Context Sensing SDK is a library for Android and Windows with specific context states; it, however, only provides front-end components [25]. The EmotionSense app is developed by the University of Cambridge to sense emotions with implications for psychological therapy and improving well-being; however, it is only focused on depression [26]. Medopad provides solutions for different health care issues with symptom tracking; this is a commercial solution and mainly focuses on phone sensors and active monitoring methods [27]. PHIT allows users to build health apps based on existing infrastructure [28]. ResearchKit, an open source framework for building apps specifically for iOS, makes it easier to enroll participants and conduct studies. However, new wearable device integration requires strong programming skills and it does not include a data management solution [29].

ResearchStack is an SDK and UX framework for building research study apps on Android, with a similar application domain as ResearchKit [30]. Both ResearchKit and ResearchStack provide software libraries, frameworks, and development tools that require extensive programming skills to create apps. A framework to create observational medical studies for mobile devices without extensive programming skills was presented [31]. Further comparison of currently available sensing platforms/apps is provided in Multimedia Appendix 2.

A key differentiator in the RADAR-base platform is the use of the Confluent platform technologies [32] (based around Apache Kafka) as the underlying infrastructure to provide a highly

scalable end-to-end solution for event-driven messaging, which is able to satisfy a wide variety of use cases, for example, high throughput, low latency messaging, real-time data processing, and fault tolerance/robustness. The platform can be deployed as microservices with Docker containers and with minimal effort extended to integrate new sensors and data sources.

Future Works and Conclusions

RADAR-base aims to stimulate the field of mHealth by providing an off-the-shelf platform for general remote data collection at scale. The project has long-term goals to improve participant care with use cases including predicting and pre-empting relapses and improving outcome measures in trials through the use of remote assessment technologies in a wide variety of disorder areas. Beyond RADAR-CNS, RADAR-base is being deployed across a number of other large EU IMI2–funded programs including RADAR-Alzheimer's Disease and is presently deployed for BigData@Heart for remote monitoring in an atrial fibrillation treatment trial (the UK National Institute for Health Research—NIHR—funded RATE-AF NCT02391337).

Acknowledgments

This study has received support from the EU/EFPIA IMI Joint Undertaking 2 (RADAR-CNS grant No 115902) [33]. This communication reflects the views of the RADAR-CNS consortium, and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. Backend infrastructure facilities were provided by King's College London Open Stack Cloud "Rosalind." Rosalind was funded with the following capital equipment grants: GSTT Charity reference: STR130505 and Maudsley Charity reference 980. The authors receive funding support from the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. The authors also acknowledge the support of NIHR University College London Hospitals Biomedical Research Centre.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Better resolution figures.

[PDF File (Adobe PDF File), 1MB - mhealth_v7i8e11734_app1.pdf]

Multimedia Appendix 2

Available sensing platform.

[PDF File (Adobe PDF File), 27KB - mhealth_v7i8e11734_app2.pdf]

References

1. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015 Apr 20;2 [FREE Full text] [doi: 10.3389/fict.2015.00006]
2. The Bridge Platform. URL: https://developer.sagebridge.org/index.html [accessed 2019-03-04] [WebCite Cache ID 76cN14HcQ]
3. RadarCNS. URL: https://www.radar-cns.org/ [accessed 2019-03-04] [WebCite Cache ID 76cNPLSHq]
4. Ranjan Y, Kerz M, Rashid Z, Böttcher S, Dobson RJ, Folarin AA. RADAR-base: A Novel Open Source m-Health Platform. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. New York, New York, USA: ACM Press; 2018 Presented at: UbiComp'18; October 08-12, 2018; Singapore, Singapore p. 223-226. [doi: 10.1145/3267305.3267579]
5. Waehner K. Confluent Internet. Apache Kafka vs Enterprise Service Bus (ESB) – Friends, Enemies or Frenemies? URL: https://www.confluent.io/blog/apache-kafka-vs-enterprise-service-bus-esb-friends-enemies-or-frenemies/ [accessed 2019-03-04] [WebCite Cache ID 76cNtyIVe]
6. Loya SR, Kawamoto K, Chatwin C, Huser V. Service oriented architecture for clinical decision support: a systematic review and future directions. J Med Syst 2014 Dec;38(12):140 [FREE Full text] [doi: 10.1007/s10916-014-0140-z] [Medline: 25325996]
7. RADAR-base. Catalogue of Passive Data Sources URL: https://radar-base.atlassian.net/wiki/spaces/RAD/pages/87359489/Data+Source+Catalog [accessed 2019-03-04] [WebCite Cache ID 76cOB8I6a]
8. GitHub. RADAR-base/RADAR-Docker URL: https://github.com/RADAR-base/RADAR-Docker [accessed 2019-03-04] [WebCite Cache ID 76cOHbcCH]
9. GitHub. RADAR-base/RADAR-REST-fitbit URL: https://github.com/RADAR-base/RADAR-REST-fitbit [accessed 2019-03-04] [WebCite Cache ID 76cOLXTt8]

10. Progress in Mind. THINC-it® cognition screening tool now clinically validated URL: https://progress.im/en/content/thinc-it-cognition-screening-tool-now-clinically-validated [accessed 2019-03-04] [WebCite Cache ID 76cOjLVJi]
11. Dask Development Team. 2016. Dask: Library for dynamic task scheduling URL: http://docs.dask.org/en/latest/cite.html [accessed 2019-03-04] [WebCite Cache ID 76cOrtEfS]
12. OAuth 2.0 Servers. Access Tokens URL: https://www.oauth.com/oauth2-servers/access-tokens/ [accessed 2019-03-04] [WebCite Cache ID 76cPKLFR9]
13. GitHub. RADAR-base/ManagementPortal URL: https://github.com/RADAR-base/ManagementPortal/tree/master/oauth-client-util [accessed 2019-03-04] [WebCite Cache ID 76cPQjeST]
14. OAuth Community Site. RFC 6750: OAuth 2.0 Bearer Token Usage URL: https://oauth.net/2/bearer-tokens/ [accessed 2019-03-04] [WebCite Cache ID 76cPcDSQ3]
15. OpenID. Welcome to OpenID Connect URL: https://openid.net/connect/ [accessed 2019-03-04] [WebCite Cache ID 76cPqRhWj]
16. RADAR-base. URL: https://radar-base.org/ [accessed 2019-03-04] [WebCite Cache ID 76cP7MSc7]
17. REDCap. URL: https://projectredcap.org/ [accessed 2019-03-04] [WebCite Cache ID 76cPxlQtJ]
18. GitHub. RADAR-base URL: https://github.com/RADAR-base [accessed 2019-03-04] [WebCite Cache ID 76cQ1bhT2]
19. Atlassian. RADAR-base - Confluence URL: https://radar-base.atlassian.net/wiki/spaces/RAD/overview [accessed 2019-03-04] [WebCite Cache ID 76cRCZakm]
20. Rashid Z, Stewart C, Böttcher S, Ranjan Y, Dobson R, Folarin AA. RADAR-base: Epilepsy Case Study. In: Proceedings of the 20 ACM International Joint Conference and 20 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp ?. New York, New York, USA: ACM Press; 2018 Presented at: UbiComp'18; October 8-12, 2018; Singapore, Singapore p. 227-230. [doi: 10.1145/3267305.3267578]
21. Stewart C, Rashid Z, Ranjan Y, Sun S, Dobson R, Folarin A. RADAR-base: Major Depressive Disorder and Epilepsy Case Studies. In: Proceedings of the 20 ACM International Joint Conference and 20 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp ?. New York, New York, USA: ACM Press; 2018 Presented at: UbiComp/ISWC'18; October 8-12, 2018; Singapore, Singapore p. 18. [doi: 10.1145/3267305.3267540]
22. RADAR-base. Supported Devices URL: https://radar-base.org/index.php/data-sensors/supported-devices/ [accessed 2019-03-04] [WebCite Cache ID 76cRxIMA8]
23. audiEERING. OpenSMILE URL: https://www.audeering.com/opensmile/ [accessed 2019-03-04] [WebCite Cache ID 76cSeufLk]
24. Aledavood T, Triana HAM, Alakörkkö T, Kaski K, Saramäki J, Isometsä E, et al. Data collection for mental health studies through digital platforms: requirements and design of a prototype. JMIR Res Protoc 2017 Jun 9;6(6):e110 [FREE Full text] [doi: 10.2196/resprot.6919] [Medline: 28600276]
25. Intel Software. Context Sensing SDK URL: https://software.intel.com/en-us/context-sensing-sdk [accessed 2019-03-04] [WebCite Cache ID 76cTK9WXN]
26. University of Cambridge. Mood-tracking app paves way for pocket therapy URL: https://www.cam.ac.uk/research/news/mood-tracking-app-paves-way-for-pocket-therapy [accessed 2019-03-04] [WebCite Cache ID 76cTQ3Icq]
27. Medopad. URL: https://medopad.com/ [accessed 2019-03-04] [WebCite Cache ID 76cU6GZQm]
28. Eckhoff RP, Kizakevich PN, Bakalov V, Zhang Y, Bryant SP, Hobbs MA. A platform to build mobile health apps: the Personal Health Intervention Toolkit (PHIT). JMIR Mhealth Uhealth 2015 Jun;3(2):e46 [FREE Full text] [doi: 10.2196/mhealth.4202] [Medline: 26033047]
29. 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]
30. ResearchStack. URL: http://researchstack.org/ [accessed 2019-03-04] [WebCite Cache ID 76cUYncq7]
31. Zens M, Grotejohann B, Tassoni A, Duttenhoefer F, Südkamp NP, Niemeyer P. Development of a modular research platform to create medical observational studies for mobile devices. JMIR Res Protoc 2017 May 23;6(5):e99 [FREE Full text] [doi: 10.2196/resprot.7705] [Medline: 28536095]
32. Confluent. Confluent Hub URL: https://www.confluent.io/product/connectors/ [accessed 2018-06-04]
33. Innovative Medicines Initiative. URL: https://www.imi.europa.eu/ [accessed 2019-03-04] [WebCite Cache ID 76cWnvcLe]

Abbreviations

API: application protocol interface
aRMT: active remote monitoring technology
CNS: Central Nervous System
HDFS: Hadoop Distributed File System
IMI: Innovative Medicines Initiative
MDD: major depressive disorder
mHealth: mobile health
MS: multiple sclerosis

NIHR: National Institute for Health Research
pRMT: passive remote monitoring technology
RADAR: Remote Assessment of Disease and Relapse
REST: Representational State Transfer
SDK: Software Development Kit
QR: quick response

Edited by G Eysenbach; submitted 29.08.18; peer-reviewed by S Zheng; comments to author 12.10.18; revised version received 28.11.18; accepted 09.12.18; published 01.08.19.

Please cite as:
Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Boettcher S, The Hyve, Dobson R, Folarin A, The RADAR-CNS Consortium
RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices
JMIR Mhealth Uhealth 2019;7(8):e11734
URL: https://mhealth.jmir.org/2019/8/e11734/
doi:10.2196/11734
PMID:31373275

Original Paper

The Validity of Daily Self-Assessed Perceived Stress Measured Using Smartphones in Healthy Individuals: Cohort Study

Helga Þórarinsdóttir[1], MD; Maria Faurholt-Jepsen[1], MD, Dr med; Henrik Ullum[2], MD, PhD; Mads Frost[3], PhD; Jakob E Bardram[4], PhD; Lars Vedel Kessing[1], MD, Dr med

[1]The Copenhagen Affective Disorder Research Centre, Psychiatric Center Copenhagen, Copenhagen, Denmark

[2]Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark

[3]Monsenso ApS, Valby, Denmark

[4]Copenhagen Center for Health Technology, Technical University of Denmark, Lyngby, Denmark

Corresponding Author:
Helga Þórarinsdóttir, MD
The Copenhagen Affective Disorder Research Centre
Psychiatric Center Copenhagen
Rigshospitalet
Copenhagen, 2100
Denmark
Phone: 45 91104943
Fax: 45 38647077
Email: helgath90@gmail.com

Abstract

Background: Smartphones may offer a new and easy tool to assess stress, but the validity has never been investigated.

Objective: This study aimed to investigate (1) the validity of smartphone-based self-assessed stress compared with Cohen Perceived Stress Scale (PSS) and (2) whether smartphone-based self-assessed stress correlates with neuroticism (Eysenck Personality Questionnaire-Neuroticism, EPQ-N), psychosocial functioning (Functioning Assessment Short Test, FAST), and prior stressful life events (Kendler Questionnaire for Stressful Life Events, SLE).

Methods: A cohort of 40 healthy blood donors with no history of personal or first-generation family history of psychiatric illness and who used an Android smartphone were instructed to self-assess their stress level daily (on a scale from 0 to 2; beta values reflect this scale) for 4 months. At baseline, participants were assessed with the FAST rater-blinded and filled out the EPQ, the PSS, and the SLE. The PSS assessment was repeated after 4 months.

Results: In linear mixed-effect regression and linear regression models, there were statistically significant positive correlations between self-assessed stress and the PSS (beta=.0167; 95% CI 0.0070-0.0026; *P*=.001), the EPQ-N (beta=.0174; 95% CI 0.0023-0.0325; *P*=.02), and the FAST (beta=.0329; 95% CI 0.0036-0.0622; *P*=.03). No correlation was found between smartphone-based self-assessed stress and the SLE.

Conclusions: Daily smartphone-based self-assessed stress seems to be a valid measure of perceived stress. Our study contains a modest sample of 40 healthy participants and adds knowledge to a new but growing field of research. Smartphone-based self-assessed stress is a promising tool for measuring stress in real time in future studies of stress and stress-related behavior.

(JMIR Mhealth Uhealth 2019;7(8):e13418) doi:10.2196/13418

KEYWORDS

emotional stress; smartphone; ecological momentary assessment; mobile phone; self-report; healthy individuals

Introduction

Background

Stress is a common experience that occurs when an individual perceives that the environmental demands exceed his or her adaptive capacity [1]. At a European level, stress has been defined as "a state, which is accompanied by physical, psychological or social complaints or dysfunctions and which results from individuals feeling unable to bridge a gap with the requirement or expectations placed on them [2]." Stress can be categorized as *distress*, which is the unpleasant type of stress,

or *eustress*, which is the good kind of stress, the type that motivates one to deal with whatever is causing the stress [3]. Stress can also be categorized as acute or chronic. Acute stress is short-lived, often relating to a specific stimulus or an event. It is accompanied by physical symptoms such as quickening heartbeat, muscular tensions, shortness of breath, and sweating. Chronic stress, on the other hand, is a long-term reaction to the pressures of daily life [4]. Over time, people may get used to the physical symptoms of chronic stress, but overexposure of the body to stress hormones can have long-term health effects. Stress affects the body in various ways: it can suppress the immune system, impact memory, and disturb digestion [5]. Chronic stress has been associated with cardiovascular disease [6], breast cancer [7], and psychiatric disorders [8]. High stress levels have also been found to be associated with higher all-cause mortality in men [9].

Stress is an individual experience, and even though the physical symptoms of stress (such as increased heart rate) can be measured objectively [10], most measures of stress focus on the individual's perception of stress, that is, subjective stress. Individuals appraise situations and responses to stress differently. Measures of self-assessed stress vary from simple yes or no questions ("do you feel stressed?") to a more complex grading of stress, for example, different point Likert-scales, to specific questions about stressful events (for a review, see [11]).

A more definite measure of subjective stress can be determined by using special instruments, such as stress assessment scales. Stress assessment scales consist of a list of questions relating to stressful events and experiences. By having individuals respond to specific questions, it is more likely that they are assessing the same kind of stress, that is, the kind of stress that has been defined by the assessment scale.

Cohen Perceived Stress Scale (PSS) is widely used for measuring individual perception of stress. It is commonly implemented in a 10-question form and measures the way respondents have found their lives unpredictable, uncontrollable, and overwhelming in the previous 14 days. The PSS has a good internal reliability (Cronbach alpha=.78-.91) and is correlated with various self-report and behavioral criteria, such as stressful life events and depressive symptoms. [12,13].

Some predictors of subjective stress have been identified in individual studies. Thus, female gender, low self-esteem, and neuroticism have consistently been associated with higher levels of subjective stress [14-17]. Having experienced stressful life events is also a predictor of higher levels of subjective stress [18].

In the last decade, with the release of smartphones and tablets, the use of mobile health (mHealth) has been steadily growing. mHealth is the practice of using mobile devices in medicine and public health [19]. One method of collecting health measures, such as subjective stress, on a mobile device is ecological momentary assessment. Ecological momentary assessment is a collection of methods used to collect "assessments of the subjects' current or recent states, sampled repeatedly over time, in their natural environment" [20]. Smartphones are a convenient and nonintrusive tool to measure subjective stress, as most people carry their phones with them throughout the day and are used to interacting with it in many locations, in many situations, and at all times [21]. Subjective stress measured on smartphones could, therefore, potentially reflect a person's real-time stress level. Being able to self-assess one's own stress levels in real time can help individuals to be aware their own stress levels. Awareness of one's own stress level is the first step toward coping with it [22]. Nevertheless, as previously reported in a systematic review by the authors, the validity of smartphone-based self-assessment of stress has never been systematically investigated [11].

Objectives

The objective of this study was to investigate the validity of smartphone-based self-assessed stress evaluated daily. More specifically, the aims were to evaluate (1) the validity of smartphone-based self-assessed stress compared with the PSS and (2) whether smartphone-based self-assessed stress correlates with neuroticism and psychosocial functioning and whether prior stressful life events predict smartphone-based self-assessed stress.

Methods

Design and Settings

This study was conducted at The Copenhagen Affective Disorder Research Centre, Psychiatric Centre Copenhagen, Denmark, using a prospective design. To increase participation, the study only consisted of 1 physical visit at baseline, whereas follow-up was conducted via mail. Participants were recruited by approaching blood donors in the waiting room at the Blood Bank at Rigshospitalet, Copenhagen, at random occasions from November 2015 to August 2016. Inclusion criteria were as follows: individuals older than 18 years, no history of psychiatric illnesses, no first-generation history of psychiatric illnesses, and should have an Android smartphone as their regular smartphone. Exclusion criteria were pregnancy and lack of Danish language skills.

A study protocol was written in August 2015 and can be acquired by contacting the author. No changes were made to the study design during follow-up.

This cohort of healthy individuals was recruited as a control group for ongoing case-control studies, investigating differences between patients with bipolar disorder, healthy individuals at risk of bipolar disorder, and healthy individuals [23].

The Monsenso app used to evaluate stress was installed on the participants' Android smartphones at baseline, and all participants were encouraged to carry their smartphones with them during the day and to use their smartphone as usual during the 4-month study period.

Measures

The Monsenso app is a smartphone app previously developed and investigated in a number of studies by the authors (eg, [23-25]). The app allowed participants to enter self-assessment data and includes automatically collected sensor data, such as measures of smartphone use, physical activity, and voice features. Participants were asked to self-assess parameters that are of importance in bipolar illness, such as mood, sleep,

cognitive impairment, and stress. Both types of data could be historically visualized on the screen, allowing the participants to see their own data. Self-assessments were daily, and the app came with a preset alarm at 8 pm to remind participants to enter their data (Figure 1). Furthermore, analyses of the automatically collected sensor data and other self-assessment measures will be described in future papers.

Baseline Assessments

Clinical Assessments

The participants were included in the study from February to August 2016 and participated for a 4-month long study period. Absence of any psychiatric diagnoses according to the International Classification of Diseases, Tenth Revision was confirmed using *Schedules for Clinical assessment in Neuropsychiatry* (SCAN) interviews [26]. All participants were assessed at baseline using the *Hamilton Depression Rating Scale-17 item* (HDRS-17) [27], the *Young Mania Rating Scale* (YMRS) [28], and the *Functioning Assessment Short Test* (FAST) [29]. Sociodemographic data on the participants were also collected at baseline.

Questionnaires

Participants filled out the following questionnaires at baseline: the *Eysenck Personality Questionnaire* (EPQ) [30], the *Kendler Questionnaire for Stressful Life Events* (SLE) [31], and the PSS [13].

Follow-Up Assessments

At the end of the 4-month study period, participants received the PSS questionnaire [13] by mail, filled it out, and sent it back to the researchers. Participants could then uninstall the Monsenso app from their smartphones.

Smartphone-Based Self-Assessed Stress

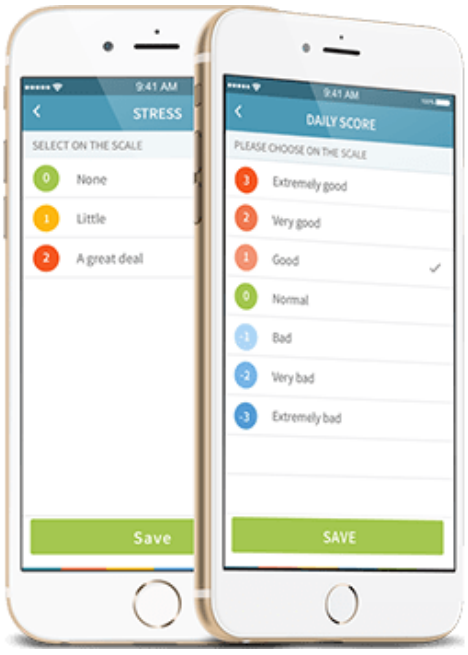
The Monsenso app prompted participants daily to self-evaluate stress. Stress was evaluated on a 3-point Likert scale with the 3 possible answers being 0=no stress, 1=little stress, and 2=much stress (Figure 1). Participants were encouraged to self-evaluate stress at the end of their day during the follow-up period.

Questionnaire-Based Measures of Stress

Participants filled out the PSS at both baseline and follow-up. As the study period was longer than 14 days (as captured by the PSS), the questionnaire was repeated to account for variation over time and to increase the statistical power of the study. The PSS is a self-evaluating questionnaire comprising 10 items on the appraisal of situations as stressful in the last 14 days.

The questionnaire asks the participants to evaluate how often they have felt their lives to be unpredictable, uncontrollable, or overwhelming in the last 14 days. Participants responded on a 5-point scale ranging from 0 (never) to 4 (very often). A total of 4 items were worded in a positive direction and were therefore reverse scored. Total scores are from a scale of 0 to 40. A survey of healthy individuals in 2009 reported a mean (SD) PSS score of 15.52 (7.44) for men and 16.14 (7.56) for women [12].

Figure 1. Examples of the self-assessment screenshot from the Monsenso app.



Additional Measures

Participants also filled out the EPQ and the SLE questionnaires at baseline. As the EPQ and the SLE regard long-time measures of personality and life events, respectively, they were not repeated at follow-up.

The EPQ-Neuroticism (EPQ-N) refers to the neuroticism score in the EPQ. The EPQ is a questionnaire with 100 yes or no questions. Different questions make up the total scores for the different personality traits: neuroticism, extroversion, and psychoticism. The EPQ has been shown to be strongly replicable across 34 countries [32], and the neuroticism score has good internal reliability (mean 0.83) [33]. Individuals who score high on neuroticism are more likely to be emotionally unstable than the average person and experience feelings such as anxiety, worry, guilt, and loneliness [34].

The SLE is a questionnaire about stressful life events [35,36]. It is divided into 2 parts: the first part is about stressful life events throughout an individual's lifetime, and the second part is about stressful life events in the past 12 months. A total of 2 total scores are calculated from the 2 parts. Inter-rater reliability values have been shown to be good to excellent ranging from 0.82 to 0.93 [35,37].

At baseline, the FAST was used to assess functional impairment. The FAST is an interviewer-administrated instrument that comprises 24 questions divided into 6 specific areas of functioning **[29].** These are autonomy, occupational functioning, cognitive functioning, financial issues, interpersonal relationships, and leisure time. Total scores are on a scale from 0 to 72, and higher scores indicate more functional impairments. Evaluations using the FAST scale were conducted rater-blinded by an interviewer without access to smartphone data.

Statistical Methods

First, descriptive analyses were produced (percentages, means, and SDs), and afterward the a priori defined statistical analyses were computed using linear regression models and linear mixed-effect regression models. A 2-level linear mixed-effect regression model, which accommodates both variation of the variables of interest within participants (intraindividual variation) and between participants (interindividual variation), was employed. The model included a fixed effect of visit number (baseline and follow-up) and a participant-specific random effect, allowing for individual intercept and slope for each participant. For all analyses, we first considered an unadjusted analysis and second, an analysis adjusted for sex and age as predefined possible confounding covariables.

Data were assembled into 3 different datasets, as the various measures and questionnaires addressed different time periods. In the first dataset, an average of self-assessed stress over the first and last 14 days of the study, was used as the PSS addresses the previous 14 days. In the second dataset, an average of all self-assessed stress data over the 4-month study period was used as the SLE, and the EPQ-N do not address a specific time period. Finally, in the third dataset, an average of self-assessed stress over the first 7 days of the study was used as the FAST at baseline addresses the previous 7 days. Even though questionnaires at baseline refer to the previous 7 or 14 days, we have decided to use the first 7 or 14 days of self-assessed measures, as we did not have any measures before baseline.

The data collected were entered in Microsoft Excel sheets, and the statistical software program Stata version 13.1 (StataCorp) was used for the statistical analyses. The statistical significance limit was set at $P<.05$.

Ethical Approval

The study was approved by the regional ethics committee in the capital region of Denmark (H-7-2014-007) and the Danish Data Protection Agency. All potential participants received written and oral information before informed consent was obtained, and participants were informed that they could withdraw from the study at any time during the study. The study complied with the Helsinki declaration [38].

Results

Participants' Flow, Background Characteristics, and Questionnaires

A total of 255 individuals were approached at the Blood Bank during the 9-month recruitment period. Over half of those (129) were ineligible to participate, and another 49 were not interested in participating in the study or thought it would be too time consuming. Of the remaining 77 individuals, 46 were included in the study. Of these, 6 were not included in the final cohort because of a lack of smartphone data. Additional 6 participants did not return the follow-up questionnaire (see flow diagram, Figure 2).

Thus, the final cohort in this study consisted of 40 healthy blood donors. The mean age was 35.24 (SD 12.79) years; 55% (22/40) of them were women, and 65% (26/40) of them were in a relationship. Information on background and sociodemographic characteristics of participants is shown in Table 1.

The mean score of self-assessed stress on smartphones was 0.12 (SD 0.34) measured on a scale from 0 to 2. Scores from questionnaires at baseline and follow-up are shown in Table 2. Participants had an average of 81.82 (SD 38.83) self-assessment days. There was no difference in the PSS and smartphone-based self-assessed stress between sexes ($P>.82$). There was no association between the age of the participants and the PSS ($P>.54$), but there was a statistically significant positive association (beta=.002; 95% CI 0.001-0.003; $P<.001$) between smartphone-based self-assessed stress and the age of the participants, namely, that for every 10-year increase in age, there was an increase in smartphone-based self-assessed stress by 0.02 on a scale from 0 to 2.

The Validity of Smartphone-Based Self-Assessed Stress Compared With Perceived Stress Scale

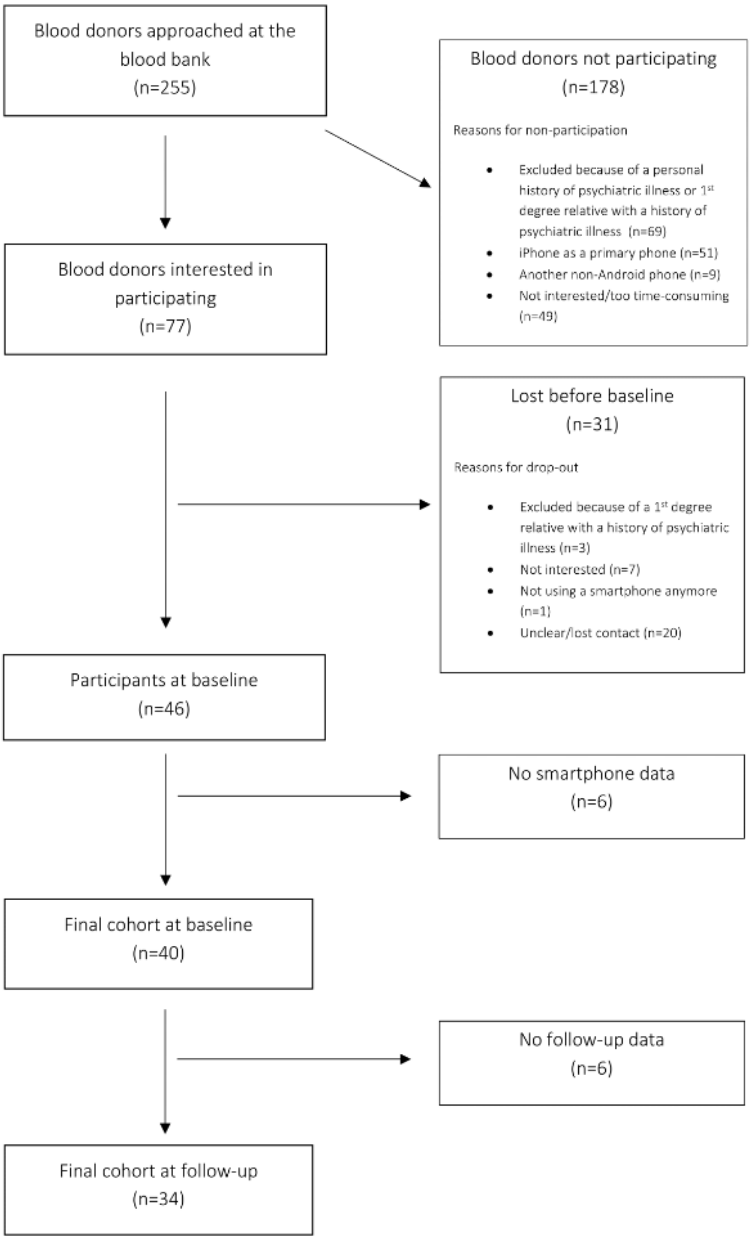
Table 3 presents the results of linear mixed-effect regression model for the self-assessed stress using smartphones and the sum scores on the PSS.

As can be seen, a statistically significant positive correlation was found between smartphone-based self-assessed stress and the PSS in both the unadjusted model and the model adjusted for age and sex (unadjusted model beta=.0167; 95% CI 0.0070-0.0026; $P=.001$), indicating that for every 10-point increase on the PSS, the smartphone-based self-assessed stress increased 0.167 on a scale from 0 to 2. Overall, there was little to no difference between the unadjusted and the adjusted models.

Association Between Smartphone-Based Self-Assessed Stress and Neuroticism

Table 3 presents the results of linear mixed-effect regression model for the self-assessed stress using smartphones and the sum scores on the EPQ-N.

Figure 2. Flow diagram of participants recruited from Rigshospitalets blood bank from November 2015 to August 2016.



A statistically significant positive correlation was found between smartphone-based self-assessed stress and the neuroticism score on the EPQ in both the unadjusted model and the model adjusted for age and sex (unadjusted model beta=.0174; 95% CI 0.0023-0.0325; *P*=.02), indicating that for every score that increased 10 points on the EPQ-N, the smartphone-based self-assessed stress was 0.174 higher on a scale from 0 to 2.

Association Between Smartphone-Based Self-Assessed Stress and Functioning Assessment Short Test

Table 3 presents the results of linear mixed-effect regression model for the self-assessed stress using smartphones and the sum scores on the FAST.

A statistically significant positive correlation was found between smartphone-based self-assessed stress and the FAST in both the unadjusted model and the model adjusted for age and sex (unadjusted model beta=.0329; 95% CI 0.0036-0.0622; *P*=.03), indicating that for every score that increased 10 points on the FAST, the smartphone-based self-assessed stress was 0.329 higher on a scale from 0 to 2.

Association Between Smartphone-Based Self-Assessed Stress and Prior Stressful Life Events

Analysis of the correlation between smartphone-based self-assessed stress and prior SLE, measured with Kendler SLE Questionnaire, yielded no statistically significant results. Neither the SLE in the previous year before baseline (unadjusted model beta=–.0055; 95% CI –0.0329 to 0.0219; *P*=.69) nor the SLE over lifetime (unadjusted model beta=.0064; 95% CI –0.0373 to 0.0502; *P*=.77) correlated with smartphone-based self-assessed stress.

Table 1. Background and sociodemographic characteristics of study participants (N=40).

Characteristics	Statistics
Age (years), mean (SD)	35.24 (12.79)
Female gender, n (%)	22 (55)
Occupation, n (%)	
Employed	23 (57.5)
Unemployed	2 (5)
Student	15 (37.5)
Education (years), mean (SD)	14.77 (2.09)
Sick days, median (IQR[a])	2 (1-4.5)
Civil status, n (%)	
Alone	22 (55)
Cohabiting	18 (45)
Marital status, n (%)	
Never married	32 (80)
Married	7 (17.5)
Divorced	1 (2.5)
Civil partner, n (%)	
In a relationship	26 (65)
Single	14 (35)
Smoking, n (%)	
Smoker	8 (20)
Former smoker	6 (15)
Never smoked	26 (65)
Alcohol units per week, median (IQR)	
Total	5 (2-7.5)
Female gender	4 (1-6)
Male gender	6 (3-12)
Former alcohol abuse, n (%)	1 (2.5)
Use of cannabis, n (%)	
Never	36 (90)
<1 monthly	4 (10)
Height (centimeters), median (IQR)	174. 68 (9.69)
Weight (kilograms), median (IQR)	78.22 (13.91)

[a]IQR: interquartile range.

Table 2. Total scores on different questionnaires and smartphone-based self-assessed stress.

Questionnaires and measures (score range; N)	Scores
EPQ-N[a] (0-14; N=38), median (IQR[b])	4.5 (2-7)
PSS[c]	
Baseline (0-20; N=37), median (IQR)	6 (4-9)
Follow-up (0-21; N=34), median (IQR)	5.5 (3-9)
SLE[d], last 12 months (0-10; N=38), median (IQR)	2 (1-4)
SLE, lifetime (0-5; N=38), median (IQR)	1 (0-2)
FAST[e] (0-8; N=40), median (IQR)	1 (0-3)
Smartphone-based self-assessed stress (0-2; N=40), mean (SD)	0.12 (0.34)

[a]EPQ-N: Eysenck Personality Questionnaire—Neuroticism.

[b]IQR: interquartile range.

[c]PSS: Perceived Stress Scale.

[d]SLE: Kendler Questionnaire for Stressful Life Events.

[e]FAST: Functioning Assessment Short Test.

Table 3. Correlations between smartphone-based self-assessed stress and the Perceived Stress Scale, the Eysenck Personality Questionnaire-Neuroticism, the Functioning Assessment Short Test, and the Kendler Stressful Life Events questionnaire (N=39).

Stress (scale 0-2)	Unadjusted		Adjusted[a]	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
PSS[b] (n=39)	0.0167 (0.0070 to 0.0026)	.001	0.0162 (0.0064 to 0.0259)	.001
EPQ-N[c] (n=38)	0.0174 (0.0023 to 0.0325)	.02	0.0167 (0.0011 to 0.0323)	.04
FAST[d] (n=38)	0.0329 (0.0036 to 0.0622)	.03	0.0307 (0.0012 to 0.6018)	.04
SLE[e], last year (n=36)	–0.0055 (–0.0329 to 0.0219)	.69	–0.0191 (–0.0520 to 0.0138)	.25
SLE, lifetime (n=36)	0.0064 (–0.0373 to 0.0502)	.77	–0.0020 (–0.0483 to 0.0444)	.93

[a]Adjusted for age and sex.

[b]PSS: Perceived Stress Scale.

[c]EPQ-N: Eysenck Personality Questionnaire-Neuroticism.

[d]FAST: Functioning Assessment Short Test.

[e]SLE: Kendler Questionnaire for Stressful Life Events

Additional Analyses

Finally, additional analyses of the correlation between the PSS and the EPQ-N, the FAST, and the SLE, respectively, were made to assess the internal validity between the PSS and smartphone-based self-assessed stress. A statistically significant positive correlation was found between the PSS and the EPQ-N (unadjusted model beta=.8663, 95% CI 0.6362-1.0965, *P*=.001; adjusted model beta=.8770, 95% CI 0.6413-1.1128, *P*=.001), indicating that for every score that increased 1 on the EPQ-N, the PSS score was 0.87 higher on a scale from 0 to 40. A statistically significant positive correlation was also found between the PSS and the FAST (unadjusted model beta=1.0965, 95% CI 0.4229-1.7702, *P*=.001; adjusted model beta=1.095, 95% CI 0.4208-1.7699, *P*=.001), indicating that for every score that increased 1 on the FAST, the PSS score was 1.09 higher on a scale from 0 to 40.

As found in the models using smartphone-based self-assessed stress, no statistically significant correlations were found between the PSS and the SLE. Neither the SLE in the previous year (unadjusted model *P*=.06; adjusted model *P*=.08), nor the SLE over lifetime (unadjusted model *P*=.95; adjusted model *P*=.77), correlated with the PSS scores.

Discussion

Principal Findings

This study followed 40 healthy blood donors for 4 months with daily self-assessment of stress using their smartphone. We found statistically significant positive correlations between smartphone-based self-assessed stress, and the PSS, the EPQ-N, and the FAST, respectively. Smartphone-based self-assessed stress did not correlate with prior stressful life events, neither in the previous year nor over a lifetime. Thus, smartphone-based self-assessed stress was validly evaluated as compared with the PSS. Furthermore, increased smartphone-based self-assessed stress was associated with increased neuroticism and decreased functioning.

To the best of our knowledge, this is the first study to explicitly investigate the validity of smartphone-based self-assessed stress. The findings from this study indicate that smartphone-based self-assessed stress is a valid measure of subjective stress on its own. As previously reported in a systematic review by the authors on smartphone-based self-assessment of stress in healthy adult individuals, 2 other previous studies have investigated the correlation between smartphone-based self-assessed stress and the PSS [11]. A study by Wang et al [39] on college students reported a statistically significant positive correlation between smartphone-based self-assessed stress and the PSS [39]. However, the objective of the study was to investigate associations to automatic objective sensor data from smartphones. Thus, Wang et al [39] did not investigate stress as the primary objective of the study, and all data were collected either on the smartphone or online, with no interviewer-administrated measures, thus increasing the risk of chance findings. Adams et al reported a nonsignificant correlation between the 2 measures of stress in a small sample (n=7) of graduate students and postdoctoral researchers [40].

The finding, that increased smartphone-based self-assessed stress was associated with increased neuroticism, adds to the validity of self-assessed stress using smartphones. Neuroticism is a personality trait and can be defined as a temperamental trait of emotionality; a tendency to arouse quickly when stimulated and to inhibit slowly when aroused [41,42]. Neuroticism is generally associated with a higher level of subjective stress and a tendency to inefficiently cope with stress [14,43-45]. People with high neuroticism scores are generally more at risk for developing psychiatric disorders, such as mood and anxiety disorders, sometimes called neurotic or stress-related disorders [46]. Awareness of one's stress level could potentially be important in people who have high neuroticism scores as it could help them to cope with stress better.

Furthermore, the finding, that increased smartphone-based self-assessed stress was associated with decreased functioning, also adds to the validity of self-assessed stress using smartphones. We used the FAST in our study, a rater-blinded measure for psychosocial functioning, and this study is the first one in this field of research to have used such a measure. Psychosocial function is an important measure, as an individual's function is essential for being able to take care of oneself and one's family. Very high levels of stress, such as those seen in posttraumatic stress disorder, lead to psychosocial functional impairment, including social and occupational impairment [47]. The relationship between stress and psychosocial function has primarily been investigated in patient populations, but studies including healthy controls find the same relationship, namely that higher stress levels are associated with impairments of psychosocial functioning, such as social functioning [48]. Stress levels have also been found to predict level of disability later in life [49].

In this study, we did not find a significant correlation between smartphone-based self-assessed stress and prior stressful life events. The SLE does not distinguish between dependent and independent stressful life events, and thus, the total score comprised all kinds of stressful life events. Independent stressful life events are those that are not influenced by the individual (eg, death of a relative), whereas dependent stressful life events are those that are in some way influenced by the individual (eg, a fight with a loved one) [50]. Dependent stressful life events are the ones that are associated with stress and depressive symptoms [51]. It is also possible that our participants, who have low average scores on the PSS, the EPQ-N, and the FAST, are better at coping with stressful life events than the general population.

Additional analyses of our data showed that higher scores on the PSS were associated with higher scores on the EPQ-N and the FAST, but no relationship was found between the PSS and the SLE. The same pattern was found in the primary analyses of the study: higher scores on smartphone-based self-assessed stress were associated with higher scores on the EPQ-N and the FAST, but no relationship was found between smartphone-based self-assessed stress and the SLE. This suggests that even though our 2 measures of stress are different in form, they are in fact measuring the same phenomenon, subjective stress.

Advantages

This study was the first study to investigate the validity of smartphone-based self-assessed stress. It was also 1 of the first studies in a relatively new field of research that did not focus primarily on the technical side of the smartphone system. Another strength of the study is that it uses ecological momentary assessment to collect measures of self-assessed stress on a daily basis and therefore minimizes recall bias [20]. A further strength of the study is that participating individuals were not aware of the aims and focus on stress in this report, hereby decreasing the risk of false-positive associations between self-reported measures (smartphone-based self-assessed stress vs the PSS and the EPQ-N, respectively), as individuals were recruited as control individuals for reported and ongoing studies [23]. The Monsenso app has been used in previous studies and has shown to have high usability [52,53]. Participants received no economic compensation for participating in the study and used their own smartphones and were, therefore, already familiar with the devices and were used to interacting with them. We used different interviewer-administrated and validated measures at baseline to ensure that our participants were healthy (no SCAN diagnosis of mental illness, no depressive or manic symptoms according to the HDRS-17 and YMRS) and had a rater-blinded measure of psychosocial function (FAST). Both unadjusted analyses and analyses adjusted for sex and age were presented.

Limitations

There are also some limitations to this study that should be mentioned. Our participants were recruited from the Blood Bank at Rigshospitalet and are likely to represent a *super healthy* population group [54]. Similarly, bias may have been introduced both in the selection of active participation and in the loss of follow-up or incomplete data. Baseline data on the PSS did not differ between participants lost to follow-up and participants with complete data (P>.72). Participants had an average of 81.82 (SD 38.83) days with self-assessment, which amounts to 68.2% (81.82/120) adherence. An adherence of close to 100% would be optimal but is difficult to achieve. By using smartphones, we were able to get an accurate measure of adherence as

participants were not able to self-assess retrospectively. Only individuals with an Android-based smartphone as their main smartphone were recruited to the study. This was because of technical reasons but could have introduced bias to the study. This affected the sample size of the study as well. A recent study investigated Android and iPhone users and found that they differed only a little in personality, but research in this area is scarce, and knowledge is limited [55]. It would be optimal to include participants using all kinds of smartphones in future studies.

Perspectives and Implications

As smartphone ownership has grown over the past decade, digital phenotyping has become a new and promising research field. Smartphones generate a high amount of data that can be collected in real time [56] and, thus, have the potentiality to reflect an individual's current state.

Stress is an increasing public health problem, and chronic exposure to stress is a risk factor for developing mental and physical diseases [6-8]. Awareness of one's own stress level is important as it is the first step toward coping with it [22], and smartphones are an unobtrusive and easily accessible tool for this. Self-assessment of daily stress using smartphones may be a step toward stress awareness.

Having a valid measure of subjective stress on a smartphone makes it possible to investigate further stress and stress-related behavior. Understanding more details regarding stress is a possible step toward decreasing it, and as stress is an individual experience, it is important that individuals are aware of their own stress level and stressors. It is possible that self-monitoring of subjective stress may help decrease stress levels, but it has never been investigated. The alternative, that self-monitoring of subjective stress increases stress levels, is a possibility as well. This should be investigated in further studies.

It is important that future studies using smartphone-based self-assessed stress also use previously validated measures of stress, such as the PSS, to confirm the validity of their smartphone-based self-assessed stress.

Conclusions

This study investigated the validity of smartphone-based self-assessed stress in relation to scores on the PSS, as well as the association between smartphone-based self-assessed stress and neuroticism, psychosocial functioning, and prior stressful life events, respectively. Smartphone-based self-assessed stress was a valid measure of subjective stress and correlated with the EPQ-N and the FAST. Measuring subjective stress on a smartphone represents a new and promising way to measure perceived stress in real time. As a valid measure of stress, smartphone-based self-assessed stress can be used in future studies of stress and stress-related behavior and may be a step toward stress awareness.

Acknowledgments

The authors would like to thank the participants for participating in the study and the Blood Bank at Rigshospitalet, Copenhagen, Denmark, for their hospitality. The study was funded by the Lundbeck foundation and the AP Møllller foundation for Icelandic students in Denmark. The funders had no role in the study design, data collection, analyses, and preparation of the manuscript.

Authors' Contributions

LVK, MFJ, and HÞ conceived the study and authored the protocol. MF and JEB were the designers of the Monsenso Solution and handled all technical matters. HÞ recruited the participants and undertook the clinical examinations. HÞ, MFJ, and LVK performed the statistical analysis, and HÞ wrote the first draft of the manuscript. All authors contributed to and have approved the final version of the manuscript.

Conflicts of Interest

MF and JEB are cofounders and shareholders in Monsenso ApS. LVK has been a consultant for Sunovion and Lundbeck in the last 3 years.

References

1. Cohen S, Gordon LU, Kessler RC. Measuring Stress: A Guide for Health and Social Scientists. Oxford: Oxford University Press; 1997.
2. European Foundation for the Improvement of Living and Working Conditions. 2010. Work-Related Stress URL: https://www.eurofound.europa.eu/sites/default/files/ef_files/docs/ewco/tn1004059s/tn1004059s.pdf [accessed 2019-01-16]
3. Selye H. The Stress of Life. Revised Edition. New York, US: McGraw-Hill; 1976.
4. American Psychological Association. 2018. Stress: The Different Kinds of Stress URL: https://www.apa.org/helpcenter/stress-kinds.aspx [accessed 2018-10-04]
5. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. Excli J 2017;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]
6. Steptoe A, Kivimäki M. Stress and cardiovascular disease. Nat Rev Cardiol 2012 Apr 3;9(6):360-370. [doi: 10.1038/nrcardio.2012.45] [Medline: 22473079]
7. Kruk J. Self-reported psychological stress and the risk of breast cancer: a case-control study. Stress 2012 Mar;15(2):162-171. [doi: 10.3109/10253890.2011.606340] [Medline: 21875303]

8. Plieger T, Melchers M, Montag C, Meermann R, Reuter M. Life stress as potential risk factor for depression and burnout. Burn Res 2015 Mar;2(1):19-24 [FREE Full text] [doi: 10.1016/j.burn.2015.03.001]
9. Nielsen NR, Kristensen TS, Schnohr P, Grønbaek M. Perceived stress and cause-specific mortality among men and women: results from a prospective cohort study. Am J Epidemiol 2008 Sep 1;168(5):481-91; discussion 492. [doi: 10.1093/aje/kwn157] [Medline: 18611955]
10. Föhr T, Tolvanen A, Myllymäki T, Järvelä-Reijonen E, Rantala S, Korpela R, et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. J Occup Med Toxicol 2015;10:39 [FREE Full text] [doi: 10.1186/s12995-015-0081-6] [Medline: 26504485]
11. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41 [FREE Full text] [doi: 10.2196/jmir.6397] [Medline: 28193600]
12. Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012 Apr 16;42(6):1320-1334 [FREE Full text] [doi: 10.1111/j.1559-1816.2012.00900.x]
13. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
14. Gramstad TO, Gjestad R, Haver B. Personality traits predict job stress, depression and anxiety among junior physicians. BMC Med Educ 2013 Nov 9;13:150 [FREE Full text] [doi: 10.1186/1472-6920-13-150] [Medline: 24207064]
15. Juth V, Smyth JM, Santuzzi AM. How do you feel? Self-esteem predicts affect, stress, social interaction, and symptom severity during daily life in patients with chronic illness. J Health Psychol 2008 Oct;13(7):884-894 [FREE Full text] [doi: 10.1177/1359105308095062] [Medline: 18809639]
16. Saleh D, Camart N, Romo L. Predictors of stress in college students. Front Psychol 2017;8:19 [FREE Full text] [doi: 10.3389/fpsyg.2017.00019] [Medline: 28179889]
17. Wolf L, Stidham AW, Ross R. Predictors of stress and coping strategies of US accelerated vs generic Baccalaureate nursing students: an embedded mixed methods study. Nurse Educ Today 2015 Jan;35(1):201-205. [doi: 10.1016/j.nedt.2014.07.005] [Medline: 25127928]
18. Feizi A, Aliyari R, Roohafza H. Association of perceived stress with stressful life events, lifestyle and sociodemographic factors: a large-scale community-based study using logistic quantile regression. Comput Math Methods Med 2012;2012:151865 [FREE Full text] [doi: 10.1155/2012/151865] [Medline: 23091560]
19. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
20. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
21. Raento M, Oulasvirta A, Eagle N. Smartphones: an emerging tool for social scientists. Sociol Methods Res 2009 Feb 1;37(3):426-454. [doi: 10.1177/0049124108330005]
22. Shapiro SL, Schwartz GE, Bonner G. Effects of mindfulness-based stress reduction on medical and premedical students. J Behav Med 1998 Dec;21(6):581-599. [doi: 10.1023/A:1018700829825] [Medline: 9891256]
23. Faurholt-Jepsen M, Busk J, Þórarinsdóttir H, Frost M, Bardram JE, Vinberg M, et al. Objective smartphone data as a potential diagnostic marker of bipolar disorder. Aust N Z J Psychiatry 2019 Feb;53(2):119-128. [doi: 10.1177/0004867418808900] [Medline: 30387368]
24. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 2627-2636. [doi: 10.1145/2470654.2481364]
25. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
26. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for clinical assessment in neuropsychiatry. Arch Gen Psychiatry 1990 Jun;47(6):589-593. [doi: 10.1001/archpsyc.1990.01810180089012] [Medline: 2190539]
27. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967 Dec;6(4):278-296. [doi: 10.1111/j.2044-8260.1967.tb00530.x] [Medline: 6080235]
28. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
29. Rosa AR, Sánchez-Moreno J, Martínez-Aran A, Salamero M, Torrent C, Reinares M, et al. Validity and reliability of the functioning assessment short test (FAST) in bipolar disorder. Clin Pract Epidemiol Ment Health 2007 Jun 7;3:5 [FREE Full text] [doi: 10.1186/1745-0179-3-5] [Medline: 17555558]
30. Eysenck HJ. Manual of the Eysenck Personality Questionnaire (Junior and Adult). London, UK: Hodder and Stoughton; 1975.

31. Kendler KS, Karkowski LM, Prescott CA. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. J Nerv Ment Dis 1998 Nov;186(11):661-669. [doi: 10.1097/00005053-199811000-00001] [Medline: 9824167]
32. Barrett PT, Petrides KV, Eysenck SB, Eysenck HJ. The Eysenck Personality Questionnaire: an examination of the factorial similarity of P, E, N, and L across 34 countries. Pers Individ Dif 1998 Nov;25(5):805-819. [doi: 10.1016/S0191-8869(98)00026-9]
33. Caruso JC, Witkiewitz K, Belcourt-Dittloff A, Gottlieb JD. Reliability of scores from the Eysenck Personality Questionnaire: a reliability generalization study. Educ Psychol Meas 2016 Jul 2;61(4):675-689 [FREE Full text] [doi: 10.1177/00131640121971437]
34. Tackett JL, Lahey BB. Neuroticism. In: Widiger TA, editor. The Oxford Handbook of the Five Factor Model. Oxford, UK: Oxford University Press; 2017.
35. Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. Am J Psychiatry 1995 Jun;152(6):833-842. [doi: 10.1176/ajp.152.6.833] [Medline: 7755111]
36. Vinberg M, Mortensen E, Kyvik K, Kessing L. Personality traits in unaffected twins discordant for affective disorder. Acta Psychiatr Scand 2007 Jun;115(6):442-450. [doi: 10.1111/j.1600-0447.2006.00909.x] [Medline: 17498155]
37. Kendler KS, Gardner CO. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. Arch Gen Psychiatry 2010 Nov;67(11):1120-1127 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.136] [Medline: 21041613]
38. World Medical Association. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
39. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp'14; September 13-17, 2014; Seattle, Washington p. 3-14. [doi: 10.1145/2632048.2632054]
40. Adams P, Rabbi M, Rahman T, Matthews M, Voida A, Gay G, et al. Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: PervasiveHealth'14; May 20-23, 2014; Oldenburg, Germany p. 72-79. [doi: 10.4108/icst.pervasivehealth.2014.254959]
41. Eysenck M. Personality and Individual Differences: A Natural Science Approach. New York, NY: Plenum Press; 1985.
42. McCrae RR, John OP. An introduction to the five-factor model and its applications. J Pers 1992 Jun;60(2):175-215. [doi: 10.1111/j.1467-6494.1992.tb00970.x] [Medline: 1635039]
43. Abbasi IS. The role of neuroticism in the maintenance of chronic baseline stress perception and negative affect. Span J Psychol 2016 Mar 4;19:E9. [doi: 10.1017/sjp.2016.7] [Medline: 26940336]
44. Dawson BF, Thompson NJ. The effect of personality on occupational stress in veterinary surgeons. J Vet Med Educ 2017;44(1):72-83. [doi: 10.3138/jvme.0116-020R] [Medline: 28206844]
45. Schneider TR. The role of neuroticism on psychological and physiological stress responses. J Exp Soc Psychol 2004 Nov;40(6):795-804. [doi: 10.1016/j.jesp.2004.04.005]
46. Kotov R, Gamez W, Schmidt F, Watson D. Linking 'big' personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol Bull 2010 Sep;136(5):768-821. [doi: 10.1037/a0020327] [Medline: 20804236]
47. Rodriguez P, Holowka DW, Marx BP. Assessment of posttraumatic stress disorder-related functional impairment: a review. J Rehabil Res Dev 2012;49(5):649-665 [FREE Full text] [doi: 10.1682/JRRD.2011.09.0162] [Medline: 23015577]
48. Grove TB, Tso IF, Chun J, Mueller SA, Taylor SF, Ellingrod VL, et al. Negative affect predicts social functioning across schizophrenia and bipolar disorder: findings from an integrated data analysis. Psychiatry Res 2016 Sep 30;243:198-206 [FREE Full text] [doi: 10.1016/j.psychres.2016.06.031] [Medline: 27416540]
49. Kulmala J, von Bonsdorff MB, Stenholm S, Törmäkangas T, von Bonsdorff ME, Nygård CH, et al. Perceived stress symptoms in midlife predict disability in old age: a 28-year prospective cohort study. J Gerontol A Biol Sci Med Sci 2013 Aug;68(8):984-991. [doi: 10.1093/gerona/gls339] [Medline: 23371968]
50. Kendler KS, Karkowski LM, Prescott CA. The assessment of dependence in the study of stressful life events: validation using a twin design. Psychol Med 1999 Nov;29(6):1455-1460. [doi: 10.1017/S0033291798008198] [Medline: 10616952]
51. Harkness KL, Monroe SM, Simons AD, Thase M. The generation of life events in recurrent and non-recurrent depression. Psychol Med 1999 Jan;29(1):135-144. [doi: 10.1017/S0033291798007752] [Medline: 10077302]
52. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. Bipolar Disord 2015 Nov;17(7):715-728. [doi: 10.1111/bdi.12332] [Medline: 26395972]
53. Frost M, Doryab A, Faurholt-Jepsen M, Kessing LV, Bardram JE. Supporting Disease Insight Through Data Analysis: Refinements of the Monarca Self-Assessment System. In: Proceedings of the 2013 ACM International Joint Conference

on Pervasive and Ubiquitous Computing. 2013 Presented at: UbiComp'13; September 8-12, 2013; Zurich, Switzerland p. 133-142. [doi: 10.1145/2493432.2493507]
54. Schwartz S, Susser E. The use of well controls: an unhealthy practice in psychiatric research. Psychol Med 2011 Jun;41(6):1127-1131. [doi: 10.1017/S0033291710001595] [Medline: 20810003]
55. Götz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. PLoS One 2017;12(5):e0176921 [FREE Full text] [doi: 10.1371/journal.pone.0176921] [Medline: 28467473]
56. Torous J, Staples P, Onnela JP. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]

Abbreviations

EPQ-N: Eysenck Personality Questionnaire-Neuroticism
FAST: Functioning Assessment Short Test
HDRS-17: Hamilton Depression Rating Scale-17 item
mHealth: mobile health
PSS: Perceived Stress Scale
SCAN: Clinical assessment in Neuropsychiatry
SLE: Kendler Questionnaire for Stressful Life Events
YMRS: Young Mania Rating Scale

Edited by G Eysenbach; submitted 16.01.19; peer-reviewed by D Saleh, M Rasic, J Cahill; comments to author 10.04.19; revised version received 17.05.19; accepted 17.05.19; published 19.08.19.

Please cite as:
Þórarinsdóttir H, Faurholt-Jepsen M, Ullum H, Frost M, Bardram JE, Kessing LV
The Validity of Daily Self-Assessed Perceived Stress Measured Using Smartphones in Healthy Individuals: Cohort Study
JMIR Mhealth Uhealth 2019;7(8):e13418
URL: http://mhealth.jmir.org/2019/8/e13418/
doi:10.2196/13418
PMID:31429413

Original Paper

Intraindividual Variability Measurement of Fine Manual Motor Skills in Children Using an Electronic Pegboard: Cohort Study

Diego Rivera[1], PhD; Antonio García[1], MSc; Jose Eugenio Ortega[2], PhD; Bernardo Alarcos[1], PhD; Kevin van der Meulen[3], PhD; Juan R Velasco[1], PhD; Cristina del Barrio[3], PhD

[1]Departamento de Automática, Escuela Politécnica Superior, Universidad de Alcalá, Alcalá de Henares, Spain

[2]Departamento de Psicología Biológica y de la Salud, Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain

[3]Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain

Corresponding Author:
Diego Rivera, PhD
Departamento de Automática
Escuela Politécnica Superior
Universidad de Alcalá
Campus Universitario
Ctra Madrid-Barcelona, Km 33,600
Alcalá de Henares, 28871
Spain
Phone: 34 918856644
Email: diego.rivera@uah.es

Abstract

Background: Pegboard tests are a powerful technique used by health and education professionals to evaluate manual dexterity and fine motor speed, both in children and adults. Using traditional pegboards in tests, the total time that, for example, a 4-year-old child needs for inserting pegs in a pegboard, with the left or right hand, can be measured. However, these measurements only allow for studying the variability among individuals, whereas no data can be obtained on the intraindividual variability in inserting and removing these pegs with one and the other hand.

Objective: The aim of this research was to study the intraindividual variabilities in fine manual motor skills of 2- to 3-year-old children during playing activities, using a custom designed electronic pegboard.

Methods: We have carried out a pilot study with 39 children, aged between 25 and 41 months. The children were observed while performing a task involving removing 10 pegs from 10 holes on one side and inserting them in 10 holes on the other side of a custom-designed sensor-based electronic pegboard, which has been built to be able to measure the times between peg insertions and removals.

Results: A sensor-based electronic pegboard was successfully developed, enabling the collection of single movement time data. In the piloting, a lower intraindividual variability was found in children with lower placement and removal times, confirming Adolph et al's hypothesis.

Conclusions: The developed pegboard allows for studying intraindividual variability using automated wirelessly transmitted data provided by its sensors. This novel technique has been useful in studying and validating the hypothesis that children with lower movement times present lower intraindividual variability. New research is necessary to confirm these findings. Research with larger sample sizes and age ranges that include additional testing of children's motor development level is currently in preparation.

(JMIR Mhealth Uhealth 2019;7(8):e12434) doi:10.2196/12434

KEYWORDS

child development; psychology, developmental; play and playthings; motor skills; smartphone

Introduction

Overview

In the research on child development, especially in the field of manual dexterity and fine motor skills, professionals use scales and tests to help them in the evaluation of the child's progress and in the detection of possible developmental delays [1,2]. These scales often involve the performance of specific activities using particular objects; among them are pegboards. A pegboard is a wooden or other material platform with holes, usually displayed along 2 parallel rows. A set of cylindrical pegs can be inserted into the holes. Depending on the scale, different measures are obtained with pegboards. For example, the time a child needs to place a particular number of pegs into the holes, withdrawing them from a container at the side of the pegboard is measured by means of a stopwatch (9-Hole Pegboard Dexterity Test). In another scale, the number of pegs that are placed in a given time lapse is counted (30 seconds in the Purdue Pegboard Test) [3].

Background

Pegboards have a long tradition in research on manual dexterity, so much that some of them have not changed substantially since 1948, as is the case of the Purdue pegboard [4]. Over time, new pegboards have appeared with their corresponding normative scores [5], such as the Functional Dexterity Test, which has been designed specifically to evaluate the functional level of persons with impairments of the dominant or nondominant hand, or the 9-Hole Peg Test, with existing norms for both children and adults [6]. The latter pegboard has been included in the test battery *Toolbox Assessment of Neurological and Behavioral Function*, recommended for its high reliability, easy application, and low cost by the National Institutes of Health [7]. These various pegboards have been tested with children [8,9] and adults [10], and can be used to evaluate manual dexterity of persons with Parkinson disease [11], Down syndrome [12], Asperger syndrome or autism [13,14], or primary school children with writing difficulties [15].

When using these pegboards, measurement accuracy depends on the skills of the professional. In addition, it is not possible to measure the time of inserting each single peg using a manual stopwatch, and consequently only the total time per trial can be obtained, for example, the time of a trial that comprises inserting 10 pegs in a board with the right hand.

An essential quality of motor development is its variability [16], both among various subjects (interindividual variability), as well as within the same person at different test moments (intraindividual variability). Although traditional pegboards provide information of interindividual variability [17,18], they cannot provide immediate data on intraindividual variability. Trials would need to be repeated several times to test the variability among performances of the same individual, which is difficult to do with children, especially very young children.

A growing number of authors emphasize the importance of intraindividual variability in motor development [19-21] and relate it to a lack of motor control in the process of acquiring new skills [16]. This hypothesis has been confirmed in research that includes measuring the reaction times to stimuli presented to children with attention-deficit/hyperactivity disorder [22,23], as well as in various studies on drawing and writing, with both children with typical development and children with developmental coordination disorder [24].

However, there are few tools available for the measurement of this intraindividual variability nowadays. On the other hand, using sensor-embedded devices in electronic health environments is becoming a popular research topic as the aggregation and analysis of sensor information is a very promising feature for developing new prevention and management techniques for diverse health issues [25,26], and also there is much literature on the design of sensor-based tools for this purpose [27]. Specifically, some efforts have already been made in the design of devices to help professionals in the study of children's developmental issues such as autism spectrum disorder [28,29].

Objectives

Following this approach, we have designed a sensor-based electronic pegboard, which enables the measurement of the time taken for every single peg removal and placement during the activity. This board, as opposed to traditional pegboards, allows the accurate measurement of intraindividual manual skills by automatically determining the absence or presence of pegs each time a peg is moved. The board is based on the proposals of the *Desarrollo de juguetes inteligentes para atención temprana a niños con trastornos del desarrollo en el entorno educativo y en el hogar digital* (EDUCERE) project [30,31], which aimed to develop an ecosystem of connected toys and tools to aid professionals in the development assessment tasks. The main architecture and modules developed in the project are described in Rivera et al [32] and are the base of the specific tool proposed here.

Using this electronic pegboard, we have designed and conducted a cohort study with 2- to 3-year-old children for the evaluation of intraindividual variability when using it for removing and placing pegs. The study and its results are described and discussed in this paper.

The rest of the paper is organized as follows: the *Methods* section describes the study carried out, its initial hypothesis, the recruitment process, and a brief description of the designed tool. The results obtained in the cohort study are shown in the *Results* section, and finally, in the *Discussion* section, we discuss the results and the possible future steps to take in this research. Finally, we add some technological design details of the sensor-based pegboard in Multimedia Appendix 1.

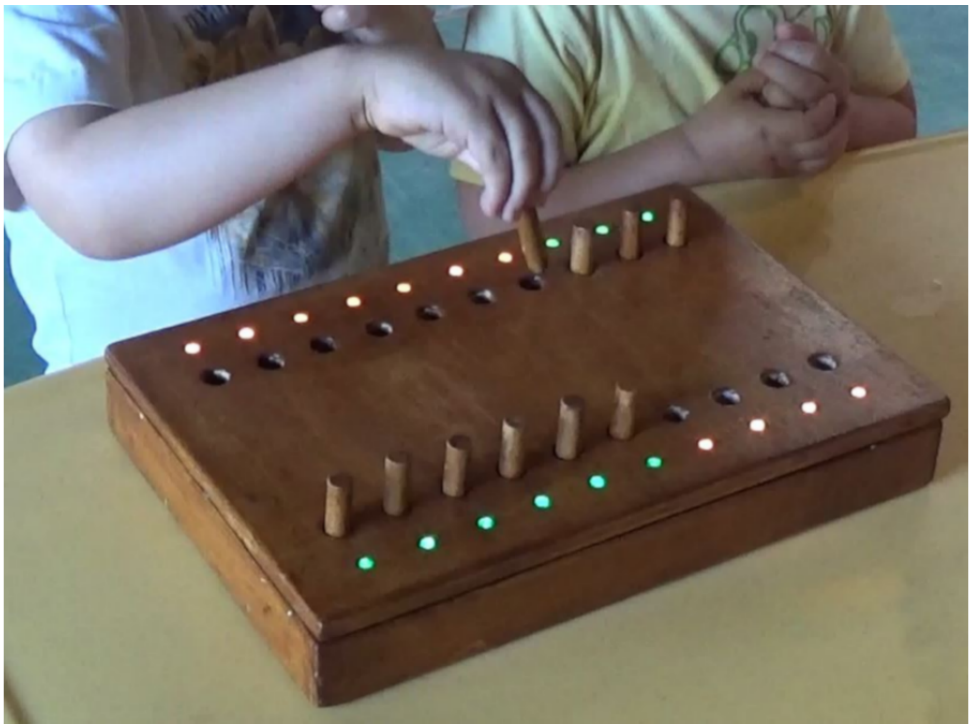
Methods

According to the hypothesis of Adolph et al [16], motor dexterity development will be accompanied by a decrease of intraindividual variability. Those children who are faster with the pegs will show less intraindividual variability, and this result will be reflected in medium- or high-level correlations between placement and removal times of each hand and the SDs.

To prove this hypothesis, we have designed a study which could determine the variability of movement times during a playing activity comprising the removal and placement of pegs in a pegboard. A total of 39 children were interviewed, 21 boys and 18 girls. The children were aged between 25 and 41 months (mean=34.03, SD 3.96) and attended a day care center for 0- to 3-year-olds, located at the university campus on the northern side of Madrid.

Once the school agreed to participate in the study, a letter was sent to the families of the 3 groups of children in the upper grade (2- to 3-year-olds), with a summary of the project and a reference to a more detailed document at their disposal in the school to obtain informed parental consent from each family (see [33] for a detailed description on ethical cautions). The project had been previously approved by the university Commission of Ethics in Science. The children were interviewed in the school setting, in a quiet room by one of the researchers, while another researcher videotaped the child-pegboard interaction and a third one supervised the registration of the activity information. Every child was invited to play with the board and move 5 of the 10 pegs, finding out how the lights would change. After this familiarization activity, the test started, and each child was invited to move the 10 pegs from their original position in row A (the row positioned facing the professional) to the empty holes in row B (the row positioned facing the child; first trial). Once finished, the child was asked to perform the same task with the other hand (second trial). A snapshot of the videotaped activity can be seen in Figure 1.

Figure 1. Snapshot of the prototype sensor-based pegboard being used by a child.



Given that the traditional pegboards do not allow to measure the intraindividual variability during its use by a single child, we have designed and developed a prototype of a sensor-based electronic pegboard. The physical design of the pegboard proposed in this paper has been based on the current designs used for dexterity tests to maintain the reliability of the tests as much as possible. The design of the board is based on a series of requirements derived from (1) the experts' needs related to the assessment tests (eg, the mobility-driven design of the device or its size), (2) the children's interactions with this type of tools (ie, the necessity of a user-friendly simple interface and interaction without any special learning requirement for its use), and (3) the technological limitations and its costs. These requirements have been compiled from different sources such as the literature on pegboard tests, experts' knowledge of the matter, and the documentation on the available sensor and communications technologies in these environments. Considering these requirements, the proposed electronic pegboard design has been developed as a modular system comprising the pegboard itself, a data collector module, and a user interface accessed through mobile or desktop devices (an example of this interface is shown in Figure 2).

The modules are independent and communicate between each other using a wireless communications system. The full components overview and its communications can be seen in Figure 3. Further details on the modules' design can be found in Multimedia Appendix 1.

To determine the reliability of the pegboard, an additional study was carried out. In this study, the total trial times of the activity of 17 children with the pegs were recorded manually by an expert in manual dexterity using a stopwatch and was compared with the total times obtained electronically. The children were aged between 30 and 41 months and completed 2 trials, one with each hand. Hereafter, the measurements were analyzed, comparing the manually registered with those using the electronic pegboard, finding a high Pearson correlation between the manually and the electronically measured times of the first (r=0.998; P<.001) and the second trials (r=0.997; P<.001).

Figure 2. Example of results of a test as shown in the Web interface from a tablet.

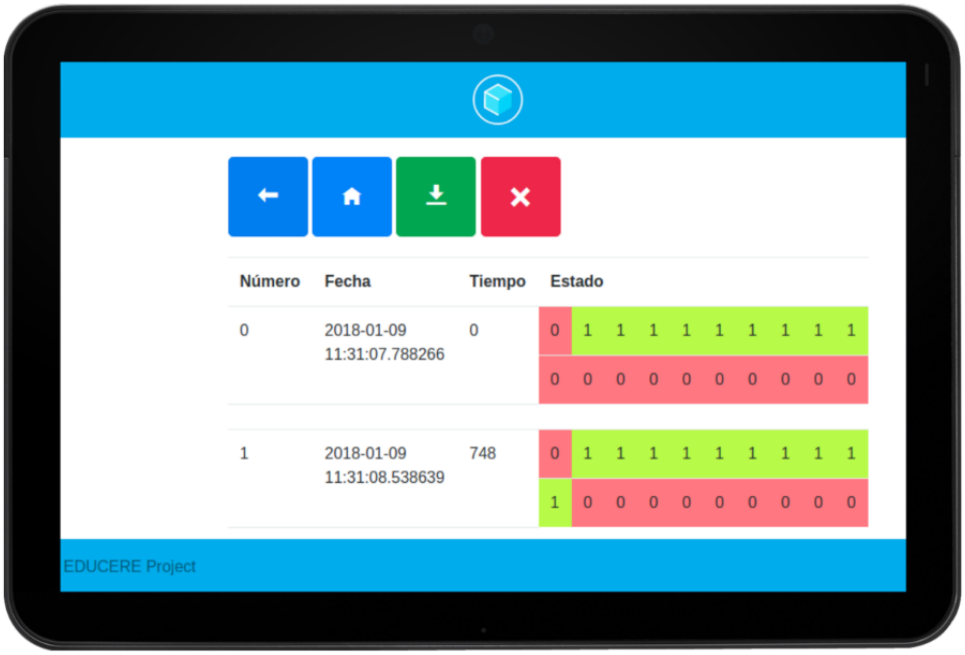
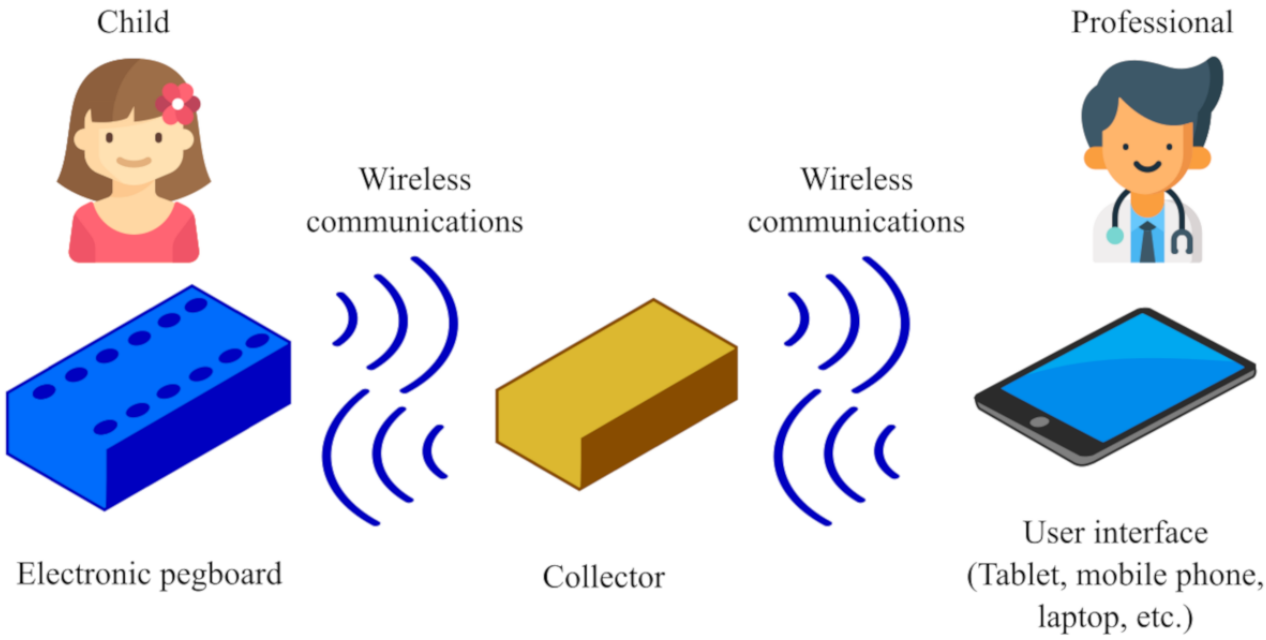


Figure 3. Electronic pegboard system components overview.



Results

The electronic pegboard has enabled the study of intraindividual variability as the system allows for recording 10 placement times (for each peg) and 9 removal times (for pegs 2 to 10, excluding the first one) in a single trial, instead of 1 single time measurement for the total amount of pegs. In addition, it is possible to obtain the individual profiles of task performance for each child. In the following paragraphs, results concerning the intraindividual variability and individual profiles are reported.

Specifically, we have recorded the mean and intraindividual values (IIV) for each child performing the activity. The values have been classified depending on the hand used for the movement and the type of movement (placement or removal of pegs) in 4 groups:

- right hand placement (RP)
- left hand placement (LP)
- right hand removal (RR)
- left hand removal (LR)

The IIV value for each child is calculated as the SD of the times measured during the activity for each group. These values along with the mean time taken for each movement can contribute to

determine if there is a relationship between the intraindividual variability, measured as the time taken to move the pegs, and the dexterity (the mean speed of placement or removal for each hand).

A correlation analysis among age, the 4 mean measures of dexterity (mean RP, mean LP, mean RR, and mean LR), and the corresponding IIV values was performed. Results (see Table 1) show various negative correlations with age (eg, older children are faster at placing pegs with the right hand). Moreover, positive correlations between mean values and IIV for each hand and movement were found, from r=0.46 (P=.003) to r=0.72 (P<.001).

Table 1. Pearson correlations between age and placement and removal measures.

Measure[a]	Age	Mean[b] RP[c]	IIV[d] RP	Mean LP[e]	IIV LP	Mean RR[f]	IIV RR	Mean LR[g]	IIV LR
Age	—[h]	—	—	—	—	—	—	—	—
Mean RP	−0.33[i]	—	—	—	—	—	—	—	—
IIV RP	−0.17	0.66[j]	—	—	—	—	—	—	—
Mean LP	−0.31	0.44[j]	−0.01	—	—	—	—	—	—
IIV LP	−0.38[i]	0.35[i]	0.20	0.46[j]	—	—	—	—	—
Mean RR	−0.18	0.57[j]	0.19	0.24	0.13	—	—	—	—
IIV RR	−0.19	0.29	−0.01	−0.05	−0.15	0.72[j]	—	—	—
Mean LR	−0.31	0.47[j]	0.14	0.36[i]	0.27	0.52[j]	0.36[i]	—	—
IIV LR	−0.40[i]	0.35[i]	0.15	0.13	0.21	0.25	0.40[i]	0.72[j]	—

[a]All tests are 2-tailed.

[b]Mean: mean value for each movement in an activity.

[c]RP: right hand placement.

[d]IIV: intraindividual variability (SD for the movements in an activity).

[e]LP: left hand placement.

[f]RR: right hand removal.

[g]LR: left hand removal.

[h]Empty cells are meant to avoid repeated correlation values (the table is diagonally symmetric).

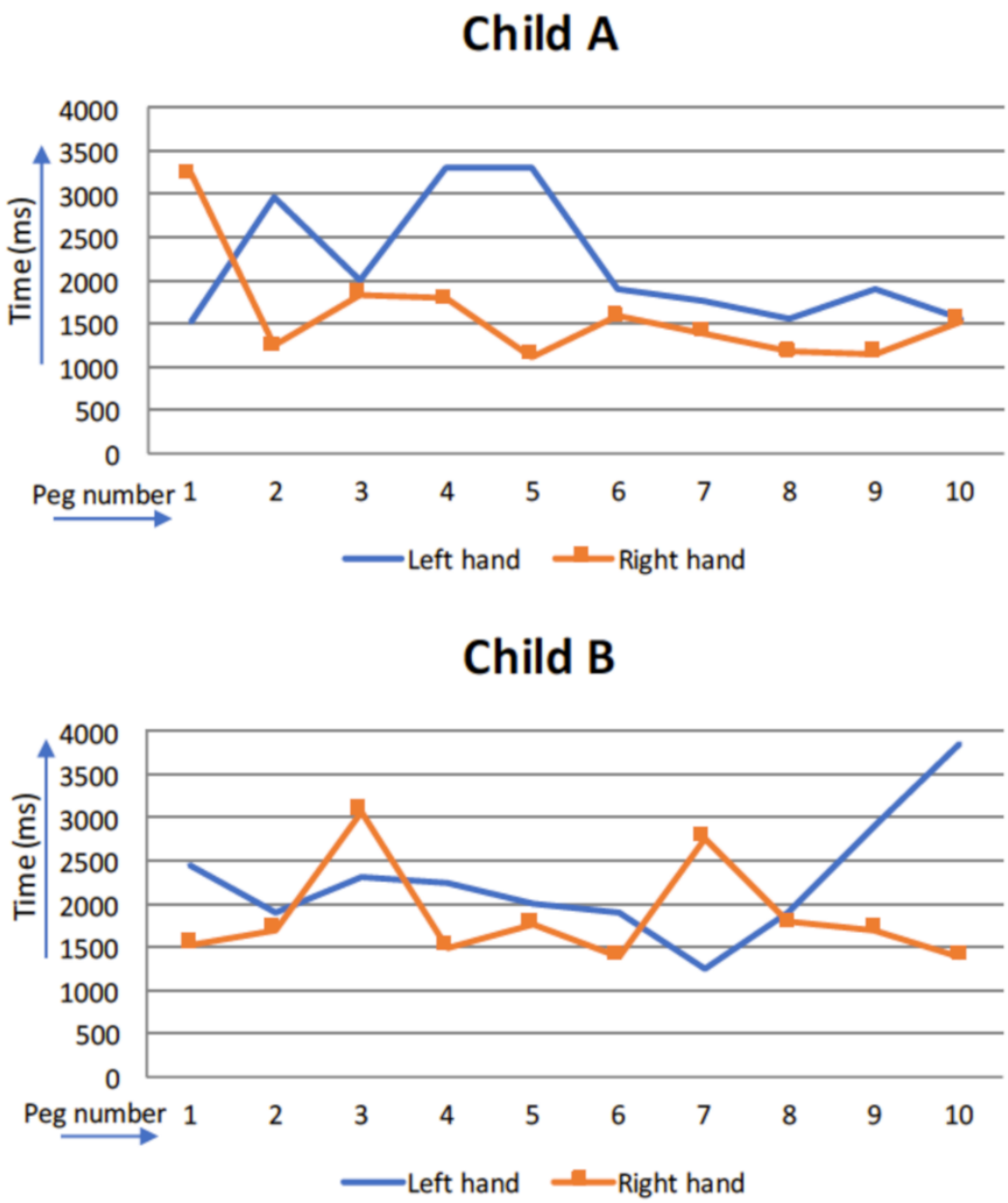
[i]P<.05.

[j]P<.01.

The increased measurement possibilities with the electronic pegboard makes it possible to obtain individual profiles of the children showing the degree of variability in their performances [15]. As an example, Figure 4 shows placement times of the 10 pegs with both the right and the left hands of 2 children, both 30 months old. The mean time of Child A's activity with the left hand is 2174.40 millisecond and the IIV value is 728.30 millisecond; with the right hand, mean=1418.78 millisecond and IIV=272.31 millisecond. Child B's mean time for task performance with the left hand is 2089.78 millisecond and IIV=463.20 millisecond; with the right hand, mean=1725.00 millisecond and IIV=416.04 millisecond.

For instance, by comparing these results for the 2 children, the differences in their profiles can be observed. Child A shows a lower mean and variability of the right hand compared with the left hand, which possibly reflects the further development of a right manual preference. However, Child B displays more comparable means and very similar variabilities of both hands, and therefore no clear manual preference.

Figure 4. Placement times in ms with left and right hands in two children (Child A and Child B). ms: milliseconds.



Discussion

The pilot study carried out with children has proven to be useful to study the intraindividual variability in children. On the basis of Adolph et al's [16] hypothesis, a lower intraindividual variability was expected in individuals with lower placement and removal times as their motor control increases. Our results show this relation in 3-year-old children, thereby confirming the initial hypothesis.

Nevertheless, some limitations can be observed in this preliminary study. A first limitation, as a consequence of its exploratory nature, is the low number of participating children. Although significant results were found, this low number of participants did now allow applying Bonferroni corrections, which is necessary when the number of correlations is high, as the results showed. Second, the age ranges of the participants were reduced, which might have caused lower correlations because of the increased homogeneity of the studied population. A third limitation lies in the absence of additional testing of the motor development level and manual dexterity of the children, as well as their manual preference, even though this may not be apparent yet. These issues are being addressed in subsequent studies.

From a technological point of view, we have presented in this paper a novel sensor-based pegboard design, which shows interesting features for the motor development in children. The

study carried out can be seen also as a validation test of the tool design, as the 2- and 3-year-old children have been able to perform all the requested activities without any added difficulty derived from its design.

In future research, we will work on the integration of the pegboard within an ecosystem of similar sensor-based wireless devices that contribute to centralized data storage and analysis from different activities and perspectives, and therefore, improve the development assessment task. This would facilitate early interventions preventing children from more severe difficulties in manual activities, including handwriting. Moreover, this instrument seems promising for more precise diagnoses of eventual motor difficulties not only in children with atypical development, but in adults with Parkinson disease as well [11-15,34].

Acknowledgments

The authors would like to thank the Universidad de Alcalá that supported the results of this paper through the *Sistemas de monitorización de actividad del desarrollo infantil y la vida cotidiana mediante juguetes y otros objetos del entorno* project (UAH-GP2018-5). This work has been funded by the Spanish National Plan for Scientific and Technical Research and Innovation 2013-2016 (EDUCERE, TIN2013-47803-C2-R [MINECO/BIE/FEDER, UE]) and 2017-2020 (*Detección precoz de trastornos de Desarrollo mediante el uso de juguetes y objetos cotidianos*, TRASGO, RTI2018-101962-B-I00 [MCIU/AEI/FEDER, UE]) and through the Spanish Ministry of Economy and Competitiveness PhD students funding program (BES-2014-067912). The vector hand drawing used in Multimedia Appendix 1 was created by Ddraw (Freepik) and the users icons used in Figure 3 were created by Freepik and all of them are licensed by Creative Commons BY 3.0. The development of the pegboard program was carried out using Arduino language and NRF24 libraries. The collector and mobile app where developed using Python and the Django Framework. The authors wish to thank the children, staff, and families of the initial education school participating in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensor-based pegboard design.
[PDF File (Adobe PDF File), 4MB - mhealth_v7i8e12434_app1.pdf]

References

1. Fewell RR, Folio MR. Peabody Developmental Motor Scales. Second Edition. Austin, TX: Pro-Ed; 2000.
2. Henderson SE, Sugden DA, Barnett SA. Movement Assessment Battery for Children. Second Edition. London, UK: Pearson Education; 2007.
3. Koyama T, Domen K, Yokoe M, Sakoda S, Kandori A. Psychometrics of dominant right hand during the 9-hole PEG test: differences between PEG placement and removal. PM R 2011 Jan;3(1):40-44. [doi: 10.1016/j.pmrj.2010.09.004] [Medline: 21257132]
4. Tiffin J, Asher EJ. The Purdue pegboard; norms and studies of reliability and validity. J Appl Psychol 1948 Jun;32(3):234-247. [doi: 10.1037/h0061266] [Medline: 18867059]
5. Aaron DH, Jansen CW. Development of the functional dexterity test (FDT): construction, validity, reliability, and normative data. J Hand Ther 2003;16(1):12-21. [doi: 10.1016/S0894-1130(03)80019-4] [Medline: 12611441]
6. Mathiowetz V, Weber K, Kashman N, Volland G. Adult norms for the nine hole peg test of finger dexterity. Occup Ther J Res 2016 Aug 24;5(1):24-38. [doi: 10.1177/153944928500500102]
7. Reuben DB, Magasi S, McCreath HE, Bohannon RW, Wang YC, Bubela DJ, et al. Motor assessment using the NIH toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S65-S75 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872e01] [Medline: 23479547]
8. Poole JL, Burtner PA, Torres TA, McMullen CK, Markham A, Marcum ML, et al. Measuring dexterity in children using the Nine-Hole Peg Test. J Hand Ther 2005;18(3):348-351. [doi: 10.1197/j.jht.2005.04.003] [Medline: 16059856]
9. Smith YA, Hong E, Presson C. Normative and validation studies of the Nine-Hole Peg Test with children. Percept Mot Skills 2000 Jun;90(3 Pt 1):823-843. [doi: 10.2466/pms.2000.90.3.823] [Medline: 10883762]
10. Wang YC, Bohannon RW, Kapellusch J, Garg A, Gershon RC. Dexterity as measured with the 9-Hole Peg Test (9-HPT) across the age span. J Hand Ther 2015;28(1):53-9; quiz 60. [doi: 10.1016/j.jht.2014.09.002] [Medline: 25449717]
11. Earhart GM, Cavanaugh JT, Ellis T, Ford MP, Foreman KB, Dibble L. The 9-hole PEG test of upper extremity function: average values, test-retest reliability, and factors contributing to performance in people with Parkinson disease. J Neurol Phys Ther 2011 Dec;35(4):157-163. [doi: 10.1097/NPT.0b013e318235da08] [Medline: 22020457]
12. Chen CC, Ringenbach SD, Albert AR. Assisted cycling exercise improves fine manual dexterity in persons with Down's syndrome. J Appl Res Intellect Disabil 2014 May;27(3):264-272. [doi: 10.1111/jar.12061] [Medline: 23775674]

13. Smith IM. Motor functioning in Asperger syndrome. In: Klin A, Volkmar FR, Sparrow SS, editors. Asperger Syndrome. New York, USA: Guilford Publications; 2000:97-124.
14. Barbeau EB, Meilleur AS, Zeffiro TA, Mottron L. Comparing motor skills in autism spectrum individuals with and without speech delay. Autism Res 2015 Dec;8(6):682-693 [FREE Full text] [doi: 10.1002/aur.1483] [Medline: 25820662]
15. Ortega JE, del Barrio C, Nunez-Nagy S, Malagón J. Evaluation of Manual Dexterity in Early Education. In: Proceedings of the 17th European Conference on Developmental Psychology. 2015 Presented at: ECDP'15; September 8-12, 2015; Braga, Portugal p. 63-67.
16. Adolph KE, Cole WG, Vereijken B. Intraindividual variability in the development of motor skills in childhood. In: Diehl M, Hooker K, Sliwinski MJ, editors. Handbook of Intraindividual Variability Across the Life Span. New York, USA: Routledge; 2015:59-83.
17. Annett M. Left, Right, Hand and Brain: Right Shift Theory. London, UK: Psychology Press; 1985.
18. Kakebeeke TH, Knaier E, Chaouch A, Caflisch J, Rousson V, Largo RH, et al. Neuromotor development in children. Part 4: new norms from 3 to 18 years. Dev Med Child Neurol 2018 Aug;60(8):810-819 [FREE Full text] [doi: 10.1111/dmcn.13793] [Medline: 29732550]
19. van Geert P, van Dijk M. Focus on variability: new tools to study intra-individual variability in developmental data. Infant Behav Dev 2002 Jan;25(4):340-374. [doi: 10.1016/S0163-6383(02)00140-6]
20. Rueckriegel SM, Blankenburg F, Burghardt R, Ehrlich S, Henze G, Mergl R, et al. Influence of age and movement complexity on kinematic hand movement parameters in childhood and adolescence. Int J Dev Neurosci 2008 Nov;26(7):655-663. [doi: 10.1016/j.ijdevneu.2008.07.015] [Medline: 18725285]
21. Palmis S, Danna JL, Velay J, Longcamp M. Motor control of handwriting in the developing brain: a review. Cogn Neuropsychol 2017;34(3-4):187-204. [doi: 10.1080/02643294.2017.1367654] [Medline: 28891745]
22. Rosch KS, Dirlikov B, Mostofsky SH. Increased intrasubject variability in boys with ADHD across tests of motor and cognitive control. J Abnorm Child Psychol 2013 Apr;41(3):485-495 [FREE Full text] [doi: 10.1007/s10802-012-9690-z] [Medline: 23135288]
23. Geurts HM, Grasman RP, Verté S, Oosterlaan J, Roeyers H, van Kammen SM, et al. Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. Neuropsychologia 2008 Nov;46(13):3030-3041. [doi: 10.1016/j.neuropsychologia.2008.06.013] [Medline: 18619477]
24. Bo J, Bastian AJ, Kagerer FA, Contreras-Vidal JL, Clark JE. Temporal variability in continuous versus discontinuous drawing for children with developmental coordination disorder. Neurosci Lett 2008 Feb 6;431(3):215-220 [FREE Full text] [doi: 10.1016/j.neulet.2007.11.040] [Medline: 18162326]
25. Hassanalieragh M, Page A, Soyata T, Sharma G, Aktas M, Mateos G, et al. Health Monitoring and Management Using Internet-of-Things (IoT) Sensing with Cloud-Based Processing: Opportunities and Challenges. In: Proceedings of the International Conference on Services Computing. 2015 Presented at: SCC'15; June 27-July 2, 2015; New York, USA p. 285-292. [doi: 10.1109/SCC.2015.47]
26. Domingo MC. An overview of the Internet of Things for people with disabilities. J Netw Comput Appl 2012 Mar;35(2):584-596. [doi: 10.1016/j.jnca.2011.10.015]
27. Marschollek M, Gietzelt M, Schulze M, Kohlmann M, Song B, Wolf KH. Wearable sensors in healthcare and sensor-enhanced health information systems: all our tomorrows? Healthc Inform Res 2012 Jun;18(2):97-104 [FREE Full text] [doi: 10.4258/hir.2012.18.2.97] [Medline: 22844645]
28. Taffoni F, Focaroli V, Formica D, Gugliemelli E, Keller F, Iverson JM. Sensor-Based Technology in the Study of Motor Skills in Infants at Risk for ASD. In: Proceedings of the 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics. 2012 Presented at: BioRob'12; June 24-27, 2012; Rome, Italy p. 1879-1883 URL: http://europepmc.org/abstract/MED/25055839 [doi: 10.1109/BioRob.2012.6290922]
29. Focaroli V, Taffoni F, Parsons SM, Keller F, Iverson JM. Performance of motor sequences in children at heightened vs low risk for ASD: a longitudinal study from 18 to 36 months of age. Front Psychol 2016;7:724 [FREE Full text] [doi: 10.3389/fpsyg.2016.00724] [Medline: 27242630]
30. Martin-Ruiz ML, Valero MA, Lindén M, Nunez S, García AG. Foundations of a smart toy development for the early detection of motoric impairments at childhood. Int J Pediatr Res 2015 Dec 31;1(2):1-5. [doi: 10.23937/2469-5769/1510011]
31. García MA, Ruiz ML, Rivera D, Vadillo L, Valero MA. A smart toy to enhance the decision-making process at children's psychomotor delay screenings: a pilot study. J Med Internet Res 2017 May 19;19(5):e171 [FREE Full text] [doi: 10.2196/jmir.7533] [Medline: 28526666]
32. Rivera D, García A, Alarcos B, Velasco JR, Ortega JE, Martínez-Yelmo I. Smart toys designed for detecting developmental delays. Sensors (Basel) 2016 Nov 20;16(11):pii: E1953 [FREE Full text] [doi: 10.3390/s16111953] [Medline: 27879626]
33. Martín-Ruíz ML, Fernández-Aller C, Portillo E, Malagón J, del Barrio C. Developing a system for processing health data of children using digitalized toys: ethical and privacy concerns for the Internet of Things paradigm. Sci Eng Ethics 2018 Aug;24(4):1057-1076. [doi: 10.1007/s11948-017-9951-x] [Medline: 28815460]
34. LeBarton ES, Landa RJ. Infant motor skill predicts later expressive language and autism spectrum disorder diagnosis. Infant Behav Dev 2019 Feb;54:37-47. [doi: 10.1016/j.infbeh.2018.11.003] [Medline: 30557704]

Abbreviations

EDUCERE: Desarrollo de juguetes inteligentes para atención temprana a niños con trastornos del desarrollo en el entorno educativo y en el hogar digital
IIV: intraindividual values
LP: left hand placement
LR: left hand removal
RP: right hand placement
RR: right hand removal

Edited by G Eysenbach; submitted 06.10.18; peer-reviewed by WC Su, B Worm, A Paglialonga; comments to author 04.02.19; revised version received 07.06.19; accepted 19.07.19; published 28.08.19.

Please cite as:
Rivera D, García A, Ortega JE, Alarcos B, van der Meulen K, Velasco JR, del Barrio C
Intraindividual Variability Measurement of Fine Manual Motor Skills in Children Using an Electronic Pegboard: Cohort Study
JMIR Mhealth Uhealth 2019;7(8):e12434
URL: http://mhealth.jmir.org/2019/8/e12434/
doi:10.2196/12434
PMID:31464193

Original Paper

Proximity-Based Emergency Response Communities for Patients With Allergies Who Are at Risk of Anaphylaxis: Clustering Analysis and Scenario-Based Survey Study

Michal Gaziel Yablowitz[1], PhD; Sabine Dölle[2], PhD; David G Schwartz[1], PhD; Margitta Worm[2], MD, PhD

[1]The Graduate School of Business Administration, Bar-Ilan University, Ramat Gan, Israel

[2]Allergy-Centre-Charite, Department of Dermatology and Allergy, Charite-Universitatsmedizin Berlin, Berlin, Germany

Corresponding Author:
Margitta Worm, MD, PhD
Allergy-Centre-Charite
Department of Dermatology and Allergy
Charite-Universitatsmedizin Berlin
Charitéplatz 1, 10117 Berlin
Berlin,
Germany
Phone: 49 30 450 518092
Email: margitta.worm@charite.de

Abstract

Background: Anaphylaxis is a potentially fatal allergic reaction. However, many patients at risk of anaphylaxis who should permanently carry a life-saving epinephrine auto injector (EAI) do not carry one at the moment of allergen exposure. The proximity-based emergency response communities (ERC) strategy suggests speeding EAI delivery by alerting patient-peers carrying EAI to respond and give their EAI to a nearby patient in need.

Objectives: This study had two objectives: (1) to analyze 10,000 anaphylactic events from the European Anaphylaxis Registry (EAR) by elicitor and location in order to determine typical anaphylactic scenarios and (2) to identify patients' behavioral and spatial factors influencing their response to ERC emergency requests through a scenario-based survey.

Methods: Data were collected and analyzed in two phases: (1) clustering 10,000 EAR records by elicitor and incident location and (2) conducting a two-center scenario-based survey of adults and parents of minors with severe allergy who were prescribed EAI, in Israel and Germany. Each group received a four-part survey that examined the effect of two behavioral constructs—shared identity and diffusion of responsibility—and two spatial factors—emergency time and emergency location—in addition to sociodemographic data. We performed descriptive, linear correlation, analysis of variance, and *t* tests to identify patients' decision factors in responding to ERC alerts.

Results: A total of 53.1% of EAR cases were triggered by food at patients' home, and 46.9% of them were triggered by venom at parks. Further, 126 Israeli and 121 German participants completed the survey and met the inclusion criteria. Of the Israeli participants, 80% were parents of minor patients with a risk of anaphylaxis due to food allergy; their mean age was 32 years, and 67% were women. In addition, 20% were adult patients with a mean age of 21 years, and 48% were female. Among the German patients, 121 were adults, with an average age of 47 years, and 63% were women. In addition, 21% were allergic to food, 75% were allergic to venom, and 2% had drug allergies. The overall willingness to respond to ERC events was high. Shared identity and the willingness to respond were positively correlated (r=0.51, P<.001) in the parent group. Parents had a stronger sense of shared identity than adult patients (t_{243}= –9.077, P<.001). The bystander effect decreased the willingness of all patients, except the parent group, to respond ($F_{1,269}$=28.27, P<.001). An interaction between location and time of emergency ($F_{1,473}$=77.304, P<.001) revealed lower levels of willingness to respond in strange locations during nighttime.

Conclusions: An ERC allergy app has the potential to improve outcomes in case of anaphylactic events, but this is dependent on patient-peers' willingness to respond. Through a two-stage process, our study identified the behavioral and spatial factors that could influence the willingness to respond, providing a basis for future research of proximity-based mental health communities.

(JMIR Mhealth Uhealth 2019;7(8):e13414) doi:10.2196/13414

KEYWORDS

consumer health informatics; anaphylaxis; emergency responders; social networking; telemedicine

Introduction

Background

Anaphylaxis is a serious, potentially fatal, systemic allergic reaction with a rapid onset. Symptoms of anaphylaxis range from skin and mucosal tissue changes such as urticaria and angioedema to life-threatening respiratory and cardiovascular conditions [1]. Anaphylaxis affects the lives of 49 million individuals in the United States alone and international guidelines consider it a medical emergency that requires rapid intervention [2-4]. To prevent potentially fatal anaphylaxis, patients with severe allergy are advised to permanently carry one or more doses of epinephrine in auto injector form (EAI). With the onset of symptoms, the immediate intramuscular injection of epinephrine from the personal EAI is used as lifesaving treatment until the arrival of emergency medical services (EMS) for continuing care. However, a significant proportion of adult patients and caregivers of minor patients with EAI prescription do not always carry their EAI and thereby expose themselves to life threats, which may not be quickly answered by EMS [5-7].

Research suggests that nearby patient-peers who maintain long-term EAI prescriptions can potentially deliver their personal EAI to the patient in need through a proximity-based emergency response communities (ERC) app [8,9]. As described in Figure 1, ERC apps dispatch nearby registered patients with allergy to help a patient in immediate need of an EAI, in certain configurations following the approval of EMS.

Based on the proximity of patient-peers, the ERC intervention is dependent, in part, on factors such as population density, prescription density, and EAI availability [10] as well as the willingness of patients to join ERCs [11]. The intervention is also dependent on the willingness of patients to respond to a stranger's request on short notice and in unfamiliar circumstances.

Prior studies that examined the efficacy of proximity-based mobile health (mHealth) interventions in improving emergency outcomes point to the insufficient participation rates as a major unsolved challenge [12-15]. Studies discussing helping behavior between strangers suggest the dual effect of diffused responsibility and shared identity as a potential explanation [16-18].

Indeed, a central factor that can impact the willingness to respond and help among a crowd of strangers is diffused responsibility, also known as the bystander effect [18-20]. According to the bystander effect, when a person has noticed an emergency event, the mere knowledge of other witnessing bystanders can decrease willingness to respond [21]. In an ERC scenario, the bystander effect may manifest when an emergency request is transmitted to a patient-peer, given that they know there are many more users of the app. This knowledge can decrease the willingness to respond at the single user level, leading to no community response. Nevertheless, the shared identity between patient-peers may serve to counteract the bystander effect, leading to mutual aid [16,22-26]. Indeed, studies about group identity among peer chronic patients show their strong willingness to help each other during crises [26].

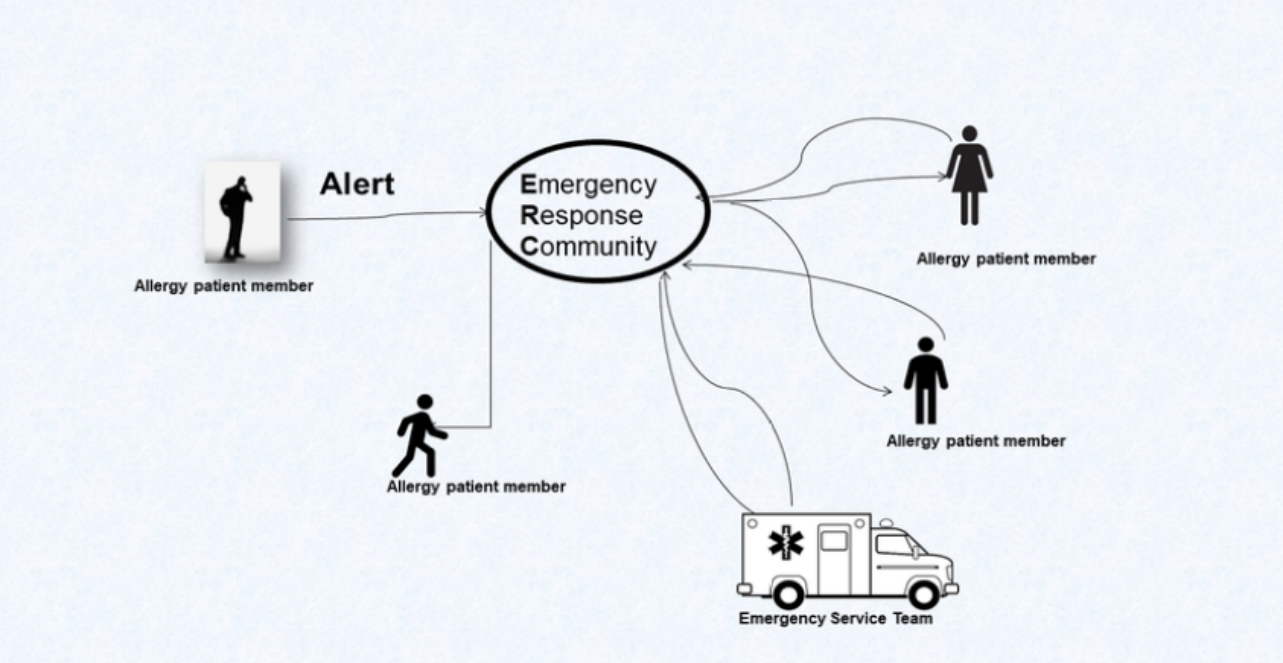
Spatial and temporal factors such as the time of emergency event can limit this helping motivation. Studies show that the willingness to help a distressed stranger varies during daytime and nighttime. Research found that women's fear of attackers in public areas during the night decreases their willingness to respond [27]. On the other hand, men are less likely to respond at night due to a lower motivation for volunteering activity [28]. Hence, the time of emergency occurrence can affect the willingness to respond. Researchers pointed to the level of familiarity with event location that may affect the willingness to travel from the responder's location to the victim's spot, as it demands an increase in the cognitive effort needed for traveling in a potentially unfamiliar environment [29]. In ERC events, the willingness of ERC members to respond can decrease if the incident location is unfamiliar and demands a concentrated effort.

Furthermore, the willingness to respond might be limited by the high cost of EAI in some communities [30]. To examine the influence of these behavioral and spatial decision factors on the response decision to ERC alerts among adult patients and parents of minor patients with EAI prescription, we conducted a scenario-based survey across two centers: (1) The Yahel Israeli Foundation for Food Allergy in Israel (Yahel), where 73% of the members are parents of minor patients with food allergy and 27% are adult patients, and (2) the outpatient clinic at the Department of Dermatology, Venereology and Allergology, Charité - Universitätsmedizin Berlin in Germany (Charité), where more than 90% of the patients are adults with allergies. Studies point to high levels of volunteering activity in Israel and Germany, suggesting that their willingness to help a distressed patient-peer may not be affected despite cultural differences [31,32]. In addition, both centers are characterized by the low cost of EAI due to a national health insurance program allowing elimination of the EAI cost from the list of behavioral barriers to respond to ERC alerts in these communities.

Finally, data analysis of the European Anaphylaxis Registry (EAR) of anaphylaxis events [33,34], collected by 90 tertiary allergy centers in 10 European countries, will allow us to develop scenarios of typical acute anaphylactic events that will reflect the emergency events of an ERC allergy app.

The improved knowledge of the behavioral and spatial decision factors would be useful in designing interventions that support patients with allergy to respond to and potentially improve emergency outcomes.

Figure 1. Proximity-based emergency response community during an anaphylactic event.



Objectives

The main goal of this study is to identify behavioral and spatial factors influencing patients at risk of an acute anaphylactic event in responding to emergency events and providing their personal EAI to a nearby patient-peer in need through a scenario-based survey. To determine typical acute anaphylactic scenarios, we analyzed the EAR dataset with 10,000 records of anaphylactic events.

Methods

Recruitment

Between May 2017 and June 2018, we conducted a scenario-based survey that was distributed in The Yahel Israeli Foundation for Food Allergy in Israel, whose members are parents of minor patients with food allergy and adult patients, and the outpatient clinic at the Department of Dermatology, Venereology and Allergology, Charité - Universitätsmedizin Berlin in Germany, which mainly treats adult patients.

Inclusion criteria for both centers were as follows: current diagnosis of allergy, considered to be at continuing risk for anaphylactic reaction, and current EAI prescription.

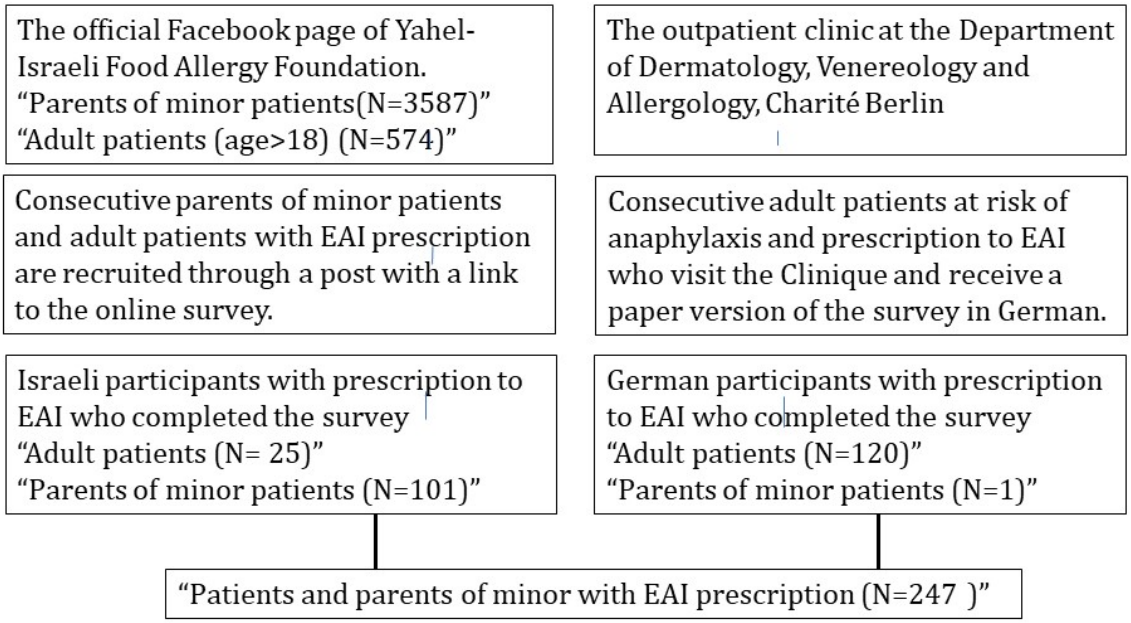
Exclusion criteria for both centers were as follows: at low risk for anaphylactic reaction as per an assessment and no current EAI prescription.

A total of 126 Israeli participants, members of the official Facebook page of Yahel, were recruited consecutively through a post. The post invited group members at risk of anaphylactic events and having an EAI prescription to participate in a survey. Members who wished to participate were directed through a link to answer an external online four-part survey.

Although recruitment through Facebook may lead to a selection bias in which the most motivated group members will respond, the distribution of a detailed four-part survey among group members with a majority of parents of minor patients with food allergies during their summer vacation months leads to a low motivation to respond [35]. To increase responsiveness, personal messages were sent to the group members every 3 days and the post was repeated periodically over a period of 2 months. Eventually, only 3% of Yahel Facebook members were recruited.

A total of 121 German patients at risk of an anaphylactic event and with an EAI prescription, who visited the Charité outpatient clinic, were recruited in a consecutive manner during the same period. Figure 2 described the selection process of the dual sampling methodology of Israeli and German participants.

Figure 2. Recruitment process of study participants. EAI: epinephrine auto injector.



Questionnaires and Scenario Development

The survey was built to examine the effect of shared identity, bystander effect [19,21,36], time of emergency [27,37], and location familiarity [38,39] on willingness to respond to anaphylaxis events through an ERC app. The questionnaire contained four sections:

The first section covered the participants' anonymous personal data and information on allergy status (Multimedia Appendix 1).

In the second section, participants were given an explanation of the ERC app for allergy patients and its potential lifesaving activity during acute anaphylactic events. Participants were assured that they will not bear any legal liability for participating in emergency events that are protected by Good Samaritan laws [22,40]. Subsequently, they were asked to project themselves as ERC members who use the app. Their initial level of shared identity was measured through an in-group identification tool [41] (Multimedia Appendix 2).

The third section of the survey presented usage scenarios [42,43] describing various anaphylaxis events when no EAI was available and help was summoned through the ERC app (Multimedia Appendix 3).

The scenarios were developed through the following two-stage process.

Stage 1: Categorizing European Anaphylaxis Registry Events According to Location and Triggers

The EAR records details of anaphylaxis events collected by 90 tertiary allergy centers in 10 European countries. It was created in 2007 and as of 2019, it contains nearly 14,000 records [34].

To determine scenarios that represent typical triggers and locations of acute anaphylactic events, we clustered 10,000 cases from the EAR dataset using a two-step cluster algorithm applied through SPSS [computer software] (Version: 25.0. Armonk, NY: IBM Corp). Clustering results were used to analyze the data set by two variables: (1) allergy elicitor, which included the following seven values—food, venom, latex, drugs, exercise, stress, cold and "I don't know"—and (2) incident location, which included the following 10 values—place of work, medical practice, garden or park, urban public place, restaurant or takeaway food, friend's home, dentist, home, school or kindergarten, public transportation, and "I don't know."

Stage 2: Developing Scenarios of Anaphylactic Events

Cluster analysis results formed the basis of the scenarios' storylines in which location familiarity, time of emergency, and bystander effect were activated to measure their impact on willingness to respond through the levels noted in Table 1.

Table 1. Decision factors' research matrix.

Bystander effect	Daytime		Nighttime	
	Unfamiliar location	Familiar location	Unfamiliar location	Familiar location
Single responder	1	2	5	7
Multiple responders	3	4	6	8

The 2 × 2 × 2 research matrix of Table 1 resulted in final eight scenarios with the following combinations:

1. Daytime, unfamiliar location, a single responder
2. Daytime, familiar location, a single responder
3. Daytime, unfamiliar location, multiple responders
4. Daytime, familiar location, multiple responders
5. Nighttime, unfamiliar location, a single responder
6. Nighttime, familiar location, a single responder
7. Nighttime, unfamiliar location, multiple responders
8. Nighttime, familiar location, multiple responders

To improve participants' ability to project themselves into scenarios, the locations of incidents and the main character's personal details were attuned to each study region. Following presentation of the scenario, participants were asked to rate their willingness to respond to ERC emergency requests considering the scenario circumstances. Finally, to validate our selection of scenarios' location, we asked participants to rate their familiarity levels with each scenario location through a location familiarity questionnaire (Multimedia Appendix 4).

Data Collection

Each group was divided into two subgroups. Each Israeli subgroup received a different online version of a survey through a closed Facebook page of Yahel foundation. Each of these versions contained two scenarios. Each German subgroup received a similar paper version of the survey with two scenarios each, distributed at the Charité. None of the groups was made aware of the other survey version. All participants were given an explanation of the ERC app for allergy patients and its role in calling responders during anaphylactic events.

Measures

Developed in English, Hebrew, and German, the survey was built to examine the effect of shared identity, bystander effect, time of emergency, and location familiarity on willingness to respond to anaphylactic events through an ERC app using 18 items measured on a 10-point Likert scale. All questions were piloted for comprehensibility and content validity on an independent sample of 40 patients randomly selected and not included in the study sample. Minor adjustments were made to the first draft after the pilot study. The scenarios were validated for content by a group of 25 allergy patients not included in the study sample. Their feedback indicated that the scenarios were reasonable and that the participants were able to put themselves in the hypothetical position of the ERC responding member (Multimedia Appendix 5). Participants' anonymous personal data and information on allergy status were collected through a five-item demographic questionnaire. To measure shared identity, participants were asked to rate their level of agreement or disagreement with an eight item in-group identification tool. To measure willingness to respond, participants were asked to rate their level of agreement or disagreement through a three-item agreement or disagreement tool, accompanied by a series of survey statements on willingness to respond. Location familiarity levels were measured using a three-item geospatial familiarity questionnaire [36]. All survey instruments are presented in the Multimedia Appendices.

Survey Statistical Analysis

We used simple descriptive statistics to analyze participants' sociodemographic and clinical data. To assess the strength of the association between shared identity and the willingness to respond, we obtained Pearson correlations.

We performed subgroup analyses through an independent sample *t* test to determine whether the levels of shared identity and willingness to respond differed when the sample was restricted to sociodemographic subgroups. Level of familiarity with the locations of emergency scenarios was compared using independent sample *t* test on the whole-sample level.

To test for the main effects of bystander effect, time of emergency, and location familiarity, we preformed two-way and three-way analyses of variance and *t* tests on each stratified sample: the full sample, parents of minor patients with food allergy, adult patients with food allergy, and patients with venom allergy.

Results

Participant Characteristics

A total of 247 questionnaires were delivered to 146 adults with severe allergy having an EAI prescription and 101 parents of children with severe allergic having an EAI prescription.

The majority of the 126 Israeli participants (80%) were parents of children with the risk of an anaphylactic episode due to food allergy, who had an EAI prescription. The mean age of this group was 32 years, and 67% were women (Table 2). The second subgroup of the Israeli patients were adult anaphylaxis patients with a long-term EAI prescription; their average age was 21 years, and 48% of them were women.

In Germany, all 121 patients with EAI prescription were adults. Their average age was 47 years, and 64% were women. Only one participant was a parent of a patient. In addition, 21% suffered from food allergy, 75% had venom allergy, 2% had drug allergies, and 2% had other allergies.

Comparing Shared Identity and Willingness to Respond According to Sociodemographic Data

Investigating the effect of shared identity on the willingness to respond for the whole sample required examination of the possibility to bind the sample regardless of the patients' origin.

Table 3 compares the mean scores for shared identity and willingness to respond between Israeli and German adult patients with food allergy.

Table 2. Baseline characteristics of the study participants (N=247).

Characteristics	Israel (n=126)	Germany (n=121)
Gender, n (%)		
Male	43 (35)	43 (36)
Female	83 (65)	78 (64)
Age: mean, median	30, 31	47, 49
Participant type, n (%)		
Adult patient	25 (20)	120 (99)
Parent of minor patient	101 (80)	1 (1)
Elicitor, n (%)		
Food	126 (100)	26 (21)
Venom	—[a]	91 (75)
Drug	—	2 (2)
Other	—	2 (2)

[a]Not available.

Table 3. Scores for shared identity and willingness to respond among Israeli and German adult patients with food allergy.

Adult patients with food allergy	Shared identity		Willingness to respond	
	n (%)	Mean (SD)	n (%)	Mean (SD)
Israeli patients	25 (20)	6.85 (2.08)	25 (20)	7.15 (1.56)
German patients	26 (21)	6.05 (1.53)	24 (20)	6.93 (2.13)

Although a statistical test was not applicable due to the limited sample size of the two groups, results indicate small differences between the scores of shared identity and willingness to respond between German and Israeli adult patients with food allergy. As such, we reported unified results for the entire sample.

Assessment of the strength of association between shared identity and the willingness to respond revealed a significant moderate positive correlation between these two constructs for the full sample (r=0.31, *P*<.001), a significant strong correlation for the parents of minor patients with food allergy (r=0.505, *P*=.03), a significant small correlation for adults with food allergy (r=0.289, *P*<.001), and a smaller correlation for patients with venom allergy (r=0.189, *P*<.001). Table 4 shows subject characteristics according to their level of shared identity and willingness to respond. An independent sample *t* test across all baseline characteristics revealed significant differences in the level of shared identity and the willingness to respond between the type of participant and the allergy elicitor. These results indicate that parents of young patients have significantly stronger shared identity (t_{243}= –9.077, *P*<.001) and significantly higher levels of willingness to respond than adult patients. Patients with food allergy have a significantly stronger shared identity (t_{238}=8.9, *P*<.001) and significantly higher levels of willingness to respond than patients with venom allergy.

Table 4. Results of *t* testing for shared identity and willingness to respond according to participants' characteristics.

Variable	Shared identity			Willingness to respond		
	n (%)	Mean (SD)	*P* value	n (%)	Mean (SD)	*P* value
Type of participants						
Adult patient	143 (57)	5.28 (2.20)	<.001	143 (57)	7.7 (2.18)	<.001
Minor patients' parent	102 (41)	7.56 (1.51)	<.001	102 (41)	8.7 (1.53)	<.001
Gender						
Male	86 (35)	6.10 (2.24)	.36	86 (35)	8.2 (0.98)	.74
Female	159 (65)	6.30 (2.25)	.36	159 (64)	8.15 (2.01)	.74
Elicitor						
Food	152 (61)	7.08 (1.92)	<.001	152 (62)	8.5 (1.69)	<.001
Venom	89 (36)	4.78 (2.03)	<.001	89 (36)	7.5 (2.24)	<.001
Drug	2 (0.8)	5.43 (2.09)	.15	2 (0.8)	6.12 (3.79)	.41
Other	2 (0.8)	7.37 (1.01)	.15	2 (0.8)	8 (1.77)	.41

Detecting Spatial and Behavioral Decision Factors Through Emergency Response Communities Scenarios

To detect the influence of spatial and behavioral decision factors on the willingness to respond, we first developed eight ERC survey scenarios according to the results of the EAR cluster analysis.

Data were clustered into two high-quality clusters.

The first cluster included 46.9% of anaphylaxis cases. In this cluster, 100% of incidents were triggered by venom; 50% of the incidents occurred in gardens and parks, 25% occurred in public places or at the place of work, and 25% of the cases did not report the location of the incident.

The second cluster included 53.1% of cases. In this cluster, 48% of incidents were triggered by food, 30% were triggered by drugs, 8% were triggered by venom, and 14% did not report the trigger. In addition, 50% of the incidents occurred at home; 15% occurred at medical practices and hospitals; 7% occurred in restaurants and hotels; and 28% were split between school, friend's home, urban public places, gardens and parks, place of work, and missing data.

Next, we compared the response decisions of the eight survey groups across three stratified populations.

Table 5 compares the average mean levels of willingness to respond across eight scenarios and four stratified survey populations.

Table 5. Mean levels of willingness to respond, for each assigned scenario across survey populations.

Scenarios	Full sample		Patients with food allergy				Patients with venom allergy (all adults, no parents)	
			Parents of minor patients		Adult patients			
	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)
NUM[a]	55 (22)	5.95 (1.46)	—[b]	—	11 (4)	6.18 (1.34)	40 (16)	6.0 (40)
NUS[c]	55 (22)	9.2 (1.22)	1 (0.4)	10 (1)	10 (4)	9.8 (0.34)	40 (16)	9.0 (40)
DUS[d]	63 (25)	8.5 (1.67)	53 (21)	8.6 (1.57)	10 (4)	8.0 (1.90)	—	—
DUM[e]	63 (25)	8.6 (1.65)	48 (19)	8.5 (1.69)	15 (6)	8.8 (1.52)	—	—
DFS[f]	61 (25)	8.6 (2.08)	1 (0.4)	5.0 (1)	13 (5)	8.8 (2.1)	48 (19)	8.6 (2.07)
DFM[g]	58 (23)	6.1 (1.85)	—	—	14 (5)	5.8 (0.69)	44 (18)	6.22 (2.09)
NFS[h]	63 (26)	9.1 (1.42)	48 (19)	9.1 (1.47)	15 (6)	9.33 (1.24)	—	—
NFM[i]	63 (26)	8.8 (1.33)	53 (21)	8.8 (1.30)	10 (4)	8.7 (1.56)	—	—

[a]NUM: Nighttime, unfamiliar location, multiple responders.

[b]Not available.

[c]NUS: Nighttime, unfamiliar location, a single responder.

[d]DUS: Daytime, unfamiliar location, a single responder.

[e]DUM: Daytime, unfamiliar location, multiple responders.

[f]DFS: Daytime, familiar location, a single responder.

[g]DFM: Daytime, familiar location, multiple responders.

[h]NFS: Nighttime, familiar location, a single responder.

[i]NFM: Nighttime, familiar location, multiple responders.

To test the impact of the bystander effect, time of emergency, and the type of location and its interactions on the willingness to respond, we performed variance analysis on four stratified samples (Table 6). As described in Table 4, significant differences in the scores of shared identity and the willingness to respond were detected between adult patients and parents of minor patients with food and venom allergy. Table 6 presents the main effects of each treatment factor for the full sample, parents of minors with food allergy, adults with food allergy, and adults with venom allergy: A significant main effect was observed for the bystander effect ($F_{1,473}$=108.20, P<.001) for the full sample. The post hoc t test analysis showed that the bystander effect significantly decreased the willingness to respond to emergency alerts among allergy patients (t_{479}=8.47, P<.001).

The bystander effect was also observed as a significant factor for adults with food allergy ($F_{1,90}$=28.33, P<.001) and adults with venom allergy ($F_{1,168}$=100.435, P<.001), but not for the parents of minor patients with food allergy (P=.48).

To test for the impact of emergency location on the willingness to respond, we first examined patients' familiarity with scenarios' locations. The results of a t test for independent samples (t_{476}=–6.9, P<.001) validated that patients distinguished between unfamiliar and familiar locations and decided how to respond. A two-way significant interaction was revealed between emergency location and time for the full sample ($F_{1,473}$=77.304, P<.001) and for adults with food allergy ($F_{1,90}$=13.44, P<.001).

A post hoc independent sample t test for the full sample (t_{234}=–3.9, P<.001), and patients with food allergy (t_{132}=–2.110, P=.04) showed that the ERC night alerts received in strange locations significantly decreased participants' willingness to respond compared to ERC alerts received in strange locations during the day.

In addition, t test results for the full sample (t_{243}=6.43, P<.001) and adults with food allergy (t_{141}=3.39, P<.001) indicate that night alerts received in familiar locations significantly increase the willingness to respond compared to the alerts received during the day in strange locations.

A significant three-way interaction was revealed between the bystander effect, location familiarity, and time of emergency for the full sample ($F_{1,473}$=88.19, P<.001) and for adults with food allergy ($F_{1,90}$=69.5 , P<.001). The result of a post hoc t test for the full sample (t_{116}=–11.10, P<.001) and adults with food allergy (t_{107}=12.63, P<.001) indicated that the knowledge of other potential responders in range will significantly decrease the patient-peer willingness to respond to ERC night alerts received in strange locations.

Table 6. Main effects and interactions of the three independent variables found by three-way analysis of variance test.

Treatment factors	Full sample		Patients with food allergy				Patients with venom allergy (all adults, no parents)	
			Parents of minor patients		Adult patients			
	F (df)	*P* value	*F* (df)	*P* value	*F* (df)	*P* value	*F* (df)	*P* value
Bystander effect	108.202 (1, 473)	<.001	0.857 (1, 198)	0.36	28.33 (1, 90)	<.001	100.435 (1, 168)	<.001
Time of day	4.513 (1, 473)	.03	6.457 (1, 198)	0.01	4.43 (1, 90)	.04	—[a]	—
Type of location	0.547 (1, 473)	.46	4.311 (1, 198)	0.04	0.024 (1, 90)	.88	—	—
Bystander effect* Time of Day	4.177 (1,473)	.04	—	—	2.88 (1,90)	.09	—	—
Time of Day*Type of location	77.3 (1,473)	<.001	1.63 (1, 198)	0.2	13.44 (1,90)	<.001	—	—
Type of location* Bystander effect	0.282 (1,473)	.60	—	—	0.412 (1, 90)	.52	—	—
Type of Location* Bystander effect* Time of day	88.91 (1,473)	<.001	—	—	69.5 (1,90)	<.001	—	—

[a]Not available.

Discussion

Principal Findings

This two-center study set out to identify the behavioral and spatial decision factors influencing food and venom allergy in patients and parents of minor patients to participate in the emergency response app, in which these patients can provide their personal EAI to a patient-peer in need.

The results of this study with 247 Israeli and German participants show the following: The overall score of willingness to respond to ERC alerts was high, especially among parents of minor patients with food allergy. The overall score of shared identity among patients with food allergy was high and varied across different types of participant. The association between shared identity and the willingness to respond was strong among parents of patients with allergy and moderate among the full sample. Bystander effect significantly decreased the willingness to respond in the full sample, patients with food allergy, and patients with venom allergy. It did not decrease the willingness to respond among parents with minors having food allergy. The combination of emergency time and its location significantly affected the willingness to respond among the full sample and adults with allergy. A three-way interaction between these two spatial constructs and the bystander effect significantly decreased the willingness to respond among the full sample and adult patients with allergy.

These findings lead to the following observations. First, the high scores of willingness to respond suggest that an ERC allergy app has the potential to improve outcomes in the case of anaphylactic events. Nevertheless, the significant differences in the sense of shared identity among adults with food allergy, parents of minors with food allergy, and patients with venom allergy indicate that not all patients at risk of anaphylaxis are the same and their response decisions will differ.

Second, the low score for shared identity found among the full sample of patients with venom allergy compared to the full sample of patients with food allergy is not surprising. With a strong lobbying activity for food labeling and allergen-free public spaces, patients with allergy and their families have a stronger common ground than those who are allergic to venom [44,45]. The high score of shared identity found among parents of minors with food allergy compared to adult patients with food allergy can be attributed to the former's strong parental identity as caregivers of chronically ill children who share a common medical condition [46,47].

Third, the significant strong correlation between shared identity and the willingness to respond found among parents with food allergy along with their high score of shared identity and willingness to respond confirm the results of previous studies about the association between active bystander intervention and high sense of shared identity, especially among patient-peers [22,26,48-51].

The bystander effect, which was identified as a significant barrier to response [17,21,52] among the full sample, including adults with food allergy and patients with venom allergy, did not impact the willingness to respond among parents of minor patients with food allergy. These results show that patients will not always respond to ERC alerts, irrespective of the circumstances, and are selective in their response decision. Hence, there is a low possibility that participants' response decisions were biased due to a social desirability effect. The use of an anonymous online instrument for half of the survey population further reduces the chance of social desirability bias [53,54].

These results also suggest that an effective ERC allergy app should combine the members' proximity and strong collective identity. Thus, designing such proximity-based apps for local social community grouping of severe allergy patients can increase response rates to ERC alerts; in other words, being coincidentally in proximity is not as effective as feeling you belong to a community group. This conclusion matches what we know about the strong ability of peer patients' local communities to benefit from health care delivery and improve health behavior among its members due to strong social cohesion and local networks [55,56].

Fourth, spatial decision factors decrease the response rate only when combined. The interaction between the time of emergency and its location, which was significant for the full sample and

adults with food allergy showed that response to anaphylactic events would be limited in a strange location during the night. In accordance, a three-way interaction between ERC night alert, strange location, and the bystander effect will significantly decrease the response scores.

Indeed, studies reported low rates of prosocial behavior during nighttime due to fear, especially when the emergency occurs in unfamiliar locations [27,37]. Nevertheless, the results of post hoc *t* tests for the interaction between time and emergency location identified participants' favorability to respond to night alerts over day alerts in unfamiliar locations. This further corresponds with studies showing that bystander intervention occurs when it helps the victim and does not harm the helper [57,58]. Hence, the fear of responding to night alerts in a strange location was much greater for the responder than the "potential profit" of benefiting a patient-peer in need. It follows that improving members' sense of personal security during an ERC event may contribute to higher response scores. This need can be answered through real-time communication platforms that support constant contact between ERC responders and the EMS center in the same manner that the EMS center communicates with its own first responders [59-61].

Limitations

This study has several limitations. The study samples consisted of Israeli and German participants with EAI prescriptions due to a risk of anaphylaxis. Although this represents two distinct cultures and geographies, the generalizability of findings to allergy patients from other geographic areas and cultures is still limited and further studies are needed to this effect.

The sample population included Israeli parents of children with allergy, Israeli adult patients, and German adult patients. Future research including German parents of children with allergy can complete the picture about the barriers and facilitators of participation in ERC events among Israeli and German patients at risk of anaphylaxis.

Conclusions

An ERC allergy app has the potential to improve outcomes in the case of anaphylactic events, but this depends on patient-peer willingness to respond. The results of this study showed that adults and parents of young patients at risk of anaphylaxis were willing to travel and give their personal EAI to a patient in need when dispatched by an ERC app. The strong positive correlation between shared identity and the willingness to respond for the parent subgroup suggest that an effective ERC allergy app should prefer a local online social community of patients with severe allergy who can respond to ERC alerts, thanks to their proximity and a strong social cohesion, rather than taking national or generally branded approaches. The significant impact of the bystander effect on the willingness to respond on the subgroups and significant lower scores for shared identity reinforce this conclusion.

The decision factors of time and location of emergency event significantly decreased the response score only when interacting with each other. Future examinations of design strategies that would increase ERC members' sense of personal security may overcome these barriers.

Finally, as proximity-based mHealth interventions take an innovative role in delivering emergency care, identifying its members' decision factors for participation through patient-centered strategies becomes crucial. This study identified the behavioral and spatial decision factors of severe allergy patients, providing a basis for future research of participation behavior among members of proximity-based mHealth interventions.

Acknowledgments

We thank Shlomit Reder, head of The Yahel Israeli Foundation for Food Allergy, for the assistance in data collection.

Authors' Contributions

MY and DS conceived the presented idea. MY developed the theory, designed the survey, collected the Israel data, and performed data analyses. MY conducted the cluster analysis on data provided by SD and MW. SD translated the survey tool, collected the Germany data, and reviewed the manuscript. MW and DS were involved in planning, supervised the work, and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participants' sociodemographic data.
[PDF File (Adobe PDF File), 117KB - mhealth_v7i8e13414_app1.pdf]

Multimedia Appendix 2
Shared identity survey.
[PDF File (Adobe PDF File), 173KB - mhealth_v7i8e13414_app2.pdf]

Multimedia Appendix 3

Anaphylaxis scenarios (four scenarios followed by survey items).
[PDF File (Adobe PDF File), 98KB - mhealth_v7i8e13414_app3.pdf]

Multimedia Appendix 4
Location familiarity survey.
[PDF File (Adobe PDF File), 97KB - mhealth_v7i8e13414_app4.pdf]

Multimedia Appendix 5
Validation results of survey scenarios and questionnaire.
[PDF File (Adobe PDF File), 388KB - mhealth_v7i8e13414_app5.pdf]

References

1. Muraro A, Roberts G, Worm M, Bilò MB, Brockow K, Fernández Rivas M, EAACI Food AllergyAnaphylaxis Guidelines Group. Anaphylaxis: guidelines from the European Academy of Allergy and Clinical Immunology. Allergy 2014;69(8):1026-1045. [doi: 10.1111/all.12437] [Medline: 24909803]
2. Anaphylaxis. 2018. The American Academy of Allergy, Asthma & Immunology URL: https://www.aaaai.org/conditions-and-treatments/allergies/anaphylaxis [accessed 2019-08-05]
3. Campbell RL, Luke A, Weaver AL, St Sauver JL, Bergstralh EJ, Li JT, et al. Prescriptions for self-injectable epinephrine and follow-up referral in emergency department patients presenting with anaphylaxis. Annals of Allergy, Asthma & Immunology 2008 Dec;101(6):631-636. [doi: 10.1016/S1081-1206(10)60227-X]
4. Wood RA, Camargo CA, Lieberman P, Sampson HA, Schwartz LB, Zitt M, et al. Anaphylaxis in America: The prevalence and characteristics of anaphylaxis in the United States. Journal of Allergy and Clinical Immunology 2014 Feb;133(2):461-467. [doi: 10.1016/j.jaci.2013.08.016]
5. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Prev Chronic Dis 2013 Apr 25;10:E66 [FREE Full text] [doi: 10.5888/pcd10.120239] [Medline: 23618546]
6. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
7. Grabenhenrich LB, Dölle S, Ruëff F, Renaudin J, Scherer K, Pföhler C, et al. Epinephrine in Severe Allergic Reactions: The European Anaphylaxis Register. The Journal of Allergy and Clinical Immunology: In Practice 2018 Nov;6(6):1898-1906.e1. [doi: 10.1016/j.jaip.2018.02.026]
8. Schwartz DG, Bellou A, Garcia-Castrillo L, Muraro A, Papadopoulos N. Exploring mHealth Participation for Emergency Response Communities. AJIS 2017 Mar 08;21. [doi: 10.3127/ajis.v21i0.1378]
9. Gaziel-Yablowitz M, Schwartz DG. A Review and Assessment Framework for Mobile-Based Emergency Intervention Apps. ACM Comput Surv 2018 Apr 14;51(1):1-32. [doi: 10.1145/3145846]
10. Khalemsky M, Schwartz DG. Emergency Response Community Effectiveness: A simulation modeler for comparing Emergency Medical Services with smartphone-based Samaritan response. Decision Support Systems 2017 Oct;102:57-68. [doi: 10.1016/j.dss.2017.07.003]
11. Khalemsky M, Schwartz, DG, Silberg T, Khalemsky A, Jaffe E, et al. Willingness to Join a Smartphone-based Emergency Response Community: Evidence from a Field Study. JMIR Mhealth Uhealth 2019:E. [doi: 10.2196/13892]
12. Ringh M. Karolinska Institutet. 2014. Out of hospital cardiac arrest: with focus on bystander CPR and public access defibrillation URL: http://publications.ki.se/xmlui/handle/10616/41963 [accessed 2019-08-05]
13. Brooks S, Worthington H, Gonedalles T, Bobrow B, Morrison L. Implementation of the PulsePoint smartphone application for crowd-sourcing bystander resuscitation. Crit Care 2014;18(Suppl 1):P484. [doi: 10.1186/cc13674]
14. Brooks SC, Simmons G, Worthington H, Bobrow BJ, Morrison LJ. The PulsePoint Respond mobile device application to crowdsource basic life support for patients with out-of-hospital cardiac arrest: Challenges for optimal implementation. Resuscitation 2016 Jan;98:20-26. [doi: 10.1016/j.resuscitation.2015.09.392] [Medline: 26475397]
15. Ringh M, Rosenqvist M, Hollenberg J, Jonsson M, Fredman D, Nordberg P, et al. Mobile-Phone Dispatch of Laypersons for CPR in Out-of-Hospital Cardiac Arrest. N Engl J Med 2015 Jun 11;372(24):2316-2325. [doi: 10.1056/NEJMoa1406038]
16. Allyn Piliavin J. Altruism and Helping: The Evolution of a Field: The 2008 Cooley-Mead Presentation. Soc Psychol Q 2009 Sep;72(3):209-225. [doi: 10.1177/019027250907200305]
17. Darley J, Latane B. Bystander intervention in emergencies: Diffusion of responsibility. Journal of Personality and Social Psychology 1968;8(4, Pt.1):377-383. [doi: 10.1037/h0025589]
18. Latane B, Darley JM. Group inhibition of bystander intervention in emergencies. Journal of Personality and Social Psychology 1968;10(3):215-221. [doi: 10.1037/h0026570]
19. Bickman L. The effect of another bystanders ability to help on bystander intervention in an emergency. J Exp Soc Psychol 1971;7(3):367-379.

20. Fischer P, Greitemeyer T, Pollozek F, Frey D. The unresponsive bystander: are bystanders more responsive in dangerous emergencies? Eur J Soc Psychol 2006 Mar;36(2):267-278. [doi: 10.1002/ejsp.297]
21. Latane B, Darley JM. The Unresponsive Bystander: Why Doesn't He Help?. Upper Saddle River, New Jersey: Prentice Hall; 1970.
22. Piliavin IM, Rodin J. Good samaritanism: an underground phenomenon? J Pers Soc Psychol 1969 Dec;13(4):289-299. [Medline: 5359227]
23. Batson C, Ahmad N. Using Empathy to Improve Intergroup Attitudes and Relations. Soc Issues Policy Rev 2009;3(1):141-177. [doi: 10.1111/j.1751-2409.2009.01013.x]
24. Reicher S. The psychology of crowd dynamics. In: Blackwell Handb Soc Psychol Group Process. Malden, MA: Blackwell Publishers Ltd; 2001:182-208.
25. Stotland E. Exploratory investigations of empathy. Adv Exp Soc Psychol 1969;4:271-314.
26. Shostak S, Fox NS. Forgetting and remembering epilepsy: collective memory and the experience of illness. Sociol Health Illn 2012 Mar;34(3):362-378 [FREE Full text] [doi: 10.1111/j.1467-9566.2011.01373.x] [Medline: 21707662]
27. Valentine G. The Geography of Women's Fear. Area 1989;21(4):385-390.
28. Harrison DA. Volunteer motivation and attendance decisions: Competitive theory testing in multiple samples from a homeless shelter. Journal of Applied Psychology 1995;80(3):371-385. [doi: 10.1037/0021-9010.80.3.371]
29. Lovelace KL, Hegarty M, Montello DR. Elements of Good Route Directions in FamiliarUnfamiliar Environments. In: Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 1999:65-82.
30. Pepper AN, Westermann-Clark E, Lockey RF. The High Cost of Epinephrine Autoinjectors and Possible Alternatives. J Allergy Clin Immunol Pract 2017;5(3):665-668.e1 [FREE Full text] [doi: 10.1016/j.jaip.2016.12.018] [Medline: 28215605]
31. Anheier HK, Salamon LM. Volunteering in cross-national perspective: Initial comparisons. Law and Contemporary Problems 1999;62:43.
32. Hameister N, Kausmann C, Müller D. Der Deutsches Zentrum fuer Altersfragen: Forchungsdatenzentrum. 2019. The German Survey on Volunteering - Der Deutsche Freiwilligensurvey 2014 URL: https://www.dza.de/fileadmin/dza/pdf/fdz/FWS2014_User_Manual.pdf [accessed 2019-08-05]
33. Grabenhenrich LB, Dölle S, Moneret-Vautrin A, Köhli A, Lange L, Spindler T, et al. Anaphylaxis in children and adolescents: The European Anaphylaxis Registry. Journal of Allergy and Clinical Immunology 2016 Apr;137(4):1128-1137.e1. [doi: 10.1016/j.jaci.2015.11.015]
34. Worm M, Timmermans F, Moneret-Vautrin A, Muraro A, Malmheden Yman II, Lövik M, et al. Towards a European registry of severe allergic reactions: current status of national registries and future needs. Allergy 2010 Jun 01;65(6):671-680. [doi: 10.1111/j.1398-9995.2010.02332.x] [Medline: 20148805]
35. Choi SH, Mitchell J, Lipkus I. Lessons Learned From an Online Study with Dual-smoker Couples. Am J Health Behav 2017 Jan;41(1):61-66. [Medline: 27935791]
36. Kendra C. Verywellmind. What is the Bystander Effect? Why Bystanders Fail to Intervene URL: http://psychology.about.com/od/socialpsychology/a/bystandereffect.htm [accessed 2019-08-05]
37. Brownlow A. A geography of men's fear. Geoforum 2005 Sep;36(5):581-592. [doi: 10.1016/j.geoforum.2004.11.005]
38. Gale N, Golledge RG, Halperin WC, Couclelis H. Exploring Spatial Familiarity. The Professional Geographer 2010 Mar 15;42(3):299-313. [doi: 10.1111/j.0033-0124.1990.00299.x]
39. Kuipers B. Modeling Spatial Knowledge. Cogn Sci 1978;2(2):129-153. [doi: 10.1207/s15516709cog0202_3]
40. Pardun JT. Good Samaritan Laws: A Global Perspective. Int Comp Law J 1997;20:A.
41. Leach CW, van Zomeren M, Zebel S, Vliek MLW, Pennekamp SF, Doosje B, et al. Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. J Pers Soc Psychol 2008 Jul;95(1):144-165. [doi: 10.1037/0022-3514.95.1.144] [Medline: 18605857]
42. Bødker S. Scenarios in user-centred design—setting the stage for reflection and action. Interacting with Computers 2000 Sep;13(1):61-75. [doi: 10.1016/S0953-5438(00)00024-2]
43. D'Arcy J, Hovav A, Galletta D. User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach. Information Systems Research 2009 Mar;20(1):79-98. [doi: 10.1287/isre.1070.0160]
44. Centers for Disease Control and Prevention. 2013. Food Allergies in Schools URL: http://www.cdc.gov/healthyyouth/foodallergies/pdf/13_243135_A_Food_Allergy_Web_508.pdf [accessed 2015-09-29]
45. DeSantiago-Cardenas L, Rivkina V, Whyte SA, Harvey-Gintoft BC, Bunning BJ, Gupta RS. Emergency epinephrine use for food allergy reactions in Chicago Public Schools. Am J Prev Med 2015 Feb;48(2):170-173. [doi: 10.1016/j.amepre.2014.09.005] [Medline: 25442236]
46. Cousino MK, Hazen RA. Parenting Stress Among Caregivers of Children With Chronic Illness: A Systematic Review. Journal of Pediatric Psychology 2013 Jul 10;38(8):809-828. [doi: 10.1093/jpepsy/jst049]
47. Hassall R, Rose J, McDonald J. Parenting stress in mothers of children with an intellectual disability: the effects of parental cognitions in relation to child characteristics and family support. J Intellect Disabil Res 2005 Jun;49(6):405-418. [doi: 10.1111/j.1365-2788.2005.00673.x]

48. Dovidio J, Piliavin J, Gaertner S, Schroeder D, Clark IR. The arousal: Cost-reward modelthe process of intervention: A review of the evidence. In: Review of personality and social psychology. Thousand Oaks, CA: Sage Publications; 1991:86-118.
49. Piliavin JA, Dovidio JE, Clark RF. Emergency Intervention. Orlando, FL: Academic Press; 1981.
50. Caldwell PE, Cutter HS. Alcoholics Anonymous affiliation during early recovery. J Subst Abuse Treat 1998;15(3):221-228. [Medline: 9633034]
51. Crape BL, Latkin CA, Laris AS, Knowlton AR. The effects of sponsorship in 12-step treatment of injection drug users. Drug Alcohol Depend 2002 Feb 01;65(3):291-301. [Medline: 11841900]
52. Bennett S, Banyard V, Garnhart L. To act or not to act, that is the question? Barriers and facilitators of bystander intervention. J Interpers Violence 2014;29(3):476-496. [Medline: 25638837]
53. Joinson A. Social desirability, anonymity, and internet-based questionnaires. Behavior Research Methods, Instruments, & Computers 1999 Sep;31(3):433-438. [doi: 10.3758/BF03200723]
54. Phillips DL, Clancy KJ. Some Effects of "Social Desirability" in Survey Studies. American Journal of Sociology 1972 Mar;77(5):921-940. [doi: 10.1086/225231]
55. O'Mara-Eves A, Brunton G, Oliver S, Kavanagh J, Jamal F, Thomas J. The effectiveness of community engagement in public health interventions for disadvantaged groups: a meta-analysis. BMC Public Health 2015 Feb 12;15(1):129 [FREE Full text] [doi: 10.1186/s12889-015-1352-y] [Medline: 25885588]
56. Rini C, Symes Y, Campo RA, Wu LM, Austin J. I Keep my Problems to Myself: Negative Social Network Orientation, Social Resources, and Health-Related Quality of Life in Cancer Survivors. Ann Behav Med 2016 Jun 22;50(3):385-396 [FREE Full text] [doi: 10.1007/s12160-015-9765-5] [Medline: 26693932]
57. Batson CD, Powell AA. Altruism and Prosocial Behavior. In: Handbook of Psychology. Hoboken, NJ: John Wiley & Sons, Inc; 2003.
58. Batson CD, Powel AA. Altruism and Prosocial Behavior. In: Handbook of psychology: Personality and social psychology. Oxford: Oxford University Press; 1998.
59. Pozner CN, Zane R, Nelson SJ, Levine M. International EMS systems: The United States: past, present, and future. Resuscitation 2004 Mar;60(3):239-244. [doi: 10.1016/j.resuscitation.2003.11.004] [Medline: 15050754]
60. Davis DP, Garberson LA, Andrusiek DL, Hostler D, Daya M, Pirrallo R, et al. A Descriptive Analysis of Emergency Medical Service Systems Participating in the Resuscitation Outcomes Consortium (ROC) Network. Prehospital Emergency Care 2009 Jul 02;11(4):369-382. [doi: 10.1080/10903120701537147]
61. Metelmann C, Metelmann B, Wendt M, Meissner K. LiveCity: The Impact of Video Communication on Emergency Medicine. IJEGR 2014;10(3):47-65. [doi: 10.4018/ijegr.2014070104]

Abbreviations

EAI: epinephrine auto injector
EAR: European Anaphylaxis Registry
EMS: emergency services
ERC: emergency response communities
mHealth: mobile health
SI: shared identity
WR: willingness to respond

Edited by G Eysenbach; submitted 16.01.19; peer-reviewed by J del Hoyo Francisco, D Carvalho; comments to author 21.02.19; revised version received 24.03.19; accepted 18.06.19; published 22.08.19.

Please cite as:
Gaziel Yablowitz M, Dölle S, Schwartz DG, Worm M
Proximity-Based Emergency Response Communities for Patients With Allergies Who Are at Risk of Anaphylaxis: Clustering Analysis and Scenario-Based Survey Study
JMIR Mhealth Uhealth 2019;7(8):e13414
URL: http://mhealth.jmir.org/2019/8/e13414/
doi:10.2196/13414
PMID:31441432

Original Paper

Assessing Mobile Phone Digital Literacy and Engagement in User-Centered Design in a Diverse, Safety-Net Population: Mixed Methods Study

Sarah S Nouri[1], MD, MPH; Patricia Avila-Garcia[1,2], BA; Anupama Gunshekar Cemballi[1,3], MA; Urmimala Sarkar[1,3], MD, MPH; Adrian Aguilera[2,4], PhD; Courtney Rees Lyles[1,3], PhD

[1]Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Psychiatry, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[3]Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[4]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author:
Sarah S Nouri, MD, MPH
Division of General Internal Medicine
Department of Medicine
University of California, San Francisco
1545 Divisadero Street
Box 0320
San Francisco, CA, 94143
United States
Phone: 1 2149126555
Email: sarah.nouri@ucsf.edu

Abstract

Background: Health care systems are rapidly deploying digital tools for disease management; however, few studies have evaluated their usability by vulnerable populations. To understand the barriers to app usage among vulnerable populations, we employed user-centered design (UCD) methods in the development of a new text messaging app.

Objective: The study aimed to describe variations in patients' engagement in the app design process, focusing on limited health literacy (LHL), limited English proficiency (LEP), and limited digital literacy (LDL).

Methods: We conducted 20 in-depth semistructured interviews with primary care patients at a public health care system, used open-ended discussions and card sorting tasks to seek input about mobile phones and text messaging, and used open coding to categorize the patterns of mobile phone usage and to evaluate engagement in the card sorting process. We examined qualitative differences in engagement by examining the extensiveness of participant feedback on existing and novel text messaging content and calculated the proportion of patients providing extensive feedback on existing and novel content, overall and by health literacy, English proficiency, and digital literacy.

Results: The average age of the 20 participants was 59 (SD 8) years; 13 (65%) were female, 18 (90%) were nonwhite, 16 (80%) had LHL, and 13 (65%) had LEP. All had depression, and 14 (70%) had diabetes. Most participants had smartphones (18/20, 90%) and regularly used text messaging (15/20, 75%), but 14 (70%) of them reported having difficulty texting because of inability to type, physical disability, and low literacy. We identified 10 participants as specifically having LDL; 7 of these participants had LEP, and all 10 had LHL. Half of the participants required a modification of the card sorting activity owing to not understanding it or not being able to read the cards in the allotted time. The proportion of participants who gave extensive feedback on existing content was lower in participants with limited versus adequate English proficiency (4/13, 30% vs 5/7, 71%), limited versus adequate health literacy (7/16, 44% vs 3/4, 75%), and limited versus adequate digital literacy (4/10, 40% vs 6/10, 60%); none of these differences were statistically significant. When examining the proportion of patients who gave extensive feedback for novel messaging content, those with LHL were less engaged than those with adequate health literacy (8/16, 50% vs 4/4, 100%); there were no statistical differences by any subgroup.

Conclusions: Despite widespread mobile phone use, digital literacy barriers are common among vulnerable populations. Engagement in the card sorting activity varied among participants and appeared to be lower among those with LHL, LEP, and

LDL. Researchers employing traditional UCD methods should routinely measure these communication domains among their end-user samples. Future work is needed to replicate our findings in larger samples, but augmentation of card sorting with direct observation and audiovisual cues may be more productive in eliciting feedback for those with communication barriers.

(JMIR Mhealth Uhealth 2019;7(8):e14250) doi:10.2196/14250

KEYWORDS

health information technology; mHealth; user-centered design; health literacy; digital literacy; limited English proficiency

Introduction

Background

Despite the potential promise of health-related mobile apps in improving the health of individuals with chronic diseases [1-4], few studies evaluate their usability and feasibility, and even fewer do so among vulnerable populations [5-7]. Existing data suggest that populations with limited health literacy (LHL), limited digital literacy (LDL), and limited English proficiency (LEP) are less likely to use health information technology (HIT), including mobile apps [6,8-11]. Given that smartphone ownership rates are similar in low income, minority populations compared with the general population of the United States, there are likely other barriers contributing to this decreased use of health apps [12]. To better understand these barriers and to reduce the disparities in the use of HIT for the management of chronic diseases, it is critical to include the vulnerable populations in the process of app development [13,14].

Even more specifically, there is a need to advance the field in terms of appropriate design science methods for working with vulnerable populations, including racial or ethnic minorities and those with low socioeconomic status and educational attainment [7]. There is similarly a need to understand the use and effectiveness of these methods in patients with mental health disorders that are associated with lower activation and engagement, such as depression [15]. Standard user-centered design (UCD) methods, such as prototyping and card sorting, often use rapid-fire sessions with end users to generate and prioritize a large amount of potential content for a digital health program or intervention [16-18]. Although the goal of these methods is to understand the experiences, beliefs, and preferences of end users, they could represent a cognitively demanding approach for the participants—especially in terms of hypothetical discussions about future health behaviors and sifting through large volumes of potential content.

Objectives

We sought to understand which design methods worked well within a larger study employing UCD methods to develop a text messaging app aimed at increasing physical activity among patients with comorbid diabetes and depression. In this paper, we evaluate data from 2 sets of semistructured interviews conducted during the early phases of the app development and describe the patterns of mobile phone usage and variations in engagement with design methods by health literacy, English proficiency, and digital literacy of patients recruited from a public health care system.

Methods

Research Setting and Sampling Procedure

Data for this study comprise 20 transcripts of semistructured interviews collected during 2016 to 2018 over 2 phases. All patients were recruited from primary care clinics from the public health care system for the city and county of San Francisco. A total of 10 transcripts are from patients who had completed a previous trial evaluating automated text messages as an adjunct to cognitive behavioral therapy for depression (MoodText trial, NCT01083628) [19], and 10 are from a separate group of patients participating in early design sessions to develop a text messaging app to increase physical activity in patients with diabetes and depression (Diabetes and Depression Text Messaging Intervention [DIAMANTE] trial, NCT03490253). Inclusion criteria were as follows: age ≥18 years, English- or Spanish-speaking, ownership of any type of mobile phone, and a diagnosis of depression. Exclusion criteria were active suicidal ideation with a plan and active severe psychosis.

Data Collection Procedure

Short questionnaires were administered to the participants during recruitment to assess sociodemographic factors (including age, gender, race and ethnicity, education level, income, employment status, and English proficiency), health status, and health literacy. Semistructured, in-depth interviews were conducted with all 20 participants in either English or Spanish. All interviews were 1.5 hours in duration, and the interviews in Spanish were conducted by study staff who were native speakers. Interview guides for all participants included questions about mobile phone usage, physical activity, and feedback on sample text messages and text messaging interventions (interview guides and card sorting instructions are available by request). Participants also completed a closed card sorting activity. Card sorting is a method used to explore how people group concepts, and this has previously been used in the development of text messaging interventions [20]. Participants were given a set of note cards with sample text messages written at a sixth-grade reading level in either English or Spanish, depending on their preferred language, and were asked to sort the cards into 3 piles depending on whether they liked, disliked, or felt neutral about the messages on the cards. They were then asked to explain why they liked or disliked these messages. If they did not provide a reason for liking or disliking the messages, they were probed once more by the interviewer. If the participants did not understand the card sorting activity or had difficulty reading the cards, card sorting was modified such that the interviewer read cards aloud and elicited feedback on each sample text message one at a time. On completion of the card sorting activity, participants were asked if they had any

additional ideas for potential text message content or structure of the text messaging intervention.

Qualitative Analysis

The qualitative analysis was done with open coding of all transcripts using inductive and deductive coding techniques (Dedoose version 8.1.8). One author (SN) created the initial codebook based on the interview questions. A second author (PA) read a subset of transcripts and coded using the original codebook. SN, PA, and CL met frequently to review the codebook, discuss the emerging themes, and resolve disagreements. For this study, we specifically coded for (1) patterns of mobile phone usage and (2) engagement in the design methods. This involved categorizing both the thematic ideas that emerged from the interviews, as well as the extensiveness of participation in (1) the card sorting activity to provide feedback on the existing text messaging content and (2) the semistructured interview to provide novel suggestions for new text messaging content. *Feedback on existing content* was further categorized as *extensive* if it included not only *what* the participants liked or disliked but also *why* they liked it and *how* they felt it would help them. If the feedback on the existing content consisted only of *what* the participants liked or disliked, it was determined to be *minimal*. *Novel suggestions* were *extensive* if they provided detailed content for possible text messages and also offered new types of messages. If novel suggestions were repetitions of text messages already developed by the study and shared with the participant, or if they were unrelated to the purpose or content of the study, they were determined to be *minimal or misaligned*.

Exploratory Analyses

For our secondary analyses, we were interested in the differences in key themes by participant demographics, specifically, health literacy, English proficiency, and digital literacy. Health literacy was measured using the single-item assessment by Chew et al, "How confident are you filling out medical forms by yourself?" As previously validated in a population similar to that of this study, adequate health literacy was defined by answering "extremely," whereas LHL was defined by answers "not at all," "a little bit," "somewhat," or "quite a bit" [21-23]. English proficiency was also measured using a validated single-item assessment, collapsing Spanish-speaking participants with those that reported "moderate," "little," and "very little ability" with speaking English [24]. Owing to the lack of a validated survey measure of digital literacy that is widely used for mobile phone usage, we created a definition of digital literacy based on the interview responses about the current use of mobile phones. We categorized participants with very limited use or engagement and/or self-reported difficulty in using their mobile phones as having LDL. We based our definition in Kayser et al's concept of eHealth literacy [25], incorporating both capabilities and experience/engagement in using technologies. We did not include access to technology in our definition as all participants owned a mobile phone. We then quantitatively compared the prevalence of LDL by participants' health literacy and English proficiency using Fisher exact tests to assess for differences by these 2 independent variables.

Finally, in an exploratory way, we calculated the sum of the number of unique feedback statements or novel suggestions per participant given during the card sorting task. We calculated the means and ranges of the total number of statements given and then preliminary Wilcoxon rank sum tests were conducted to evaluate the differences in these frequencies by health literacy, English proficiency, and digital literacy. We also calculated the number of participants who provided extensive versus minimal feedback statements and novel suggestions and used Fisher exact tests to evaluate the differences in these proportions by health literacy, English proficiency, and digital literacy. Analyses were conducted using Stata/SE 15.0.

Results

Participant Characteristics

The average age of the participants was 59 (SD 8) years; 65% (13/20) were female, 90% (18/20) were nonwhite, 65% (13/20) had LEP (50% (10/20) Spanish-speaking), 45% (9/20) had a high school education or less, and 80% (16/20) had LHL (Table 1). None of the participants were employed full time; 45% (9/20) were disabled, and 20% (4/20) were unemployed. A total of 70% (14/20) of the participants had type 2 diabetes mellitus, and 65% (13/20) self-rated their health status as fair or poor. The proportion of participants with LEP (80% vs 50%) and with diabetes (90% vs 50%) was higher in the DIAMANTE trial than the proportion of participants in the MoodText trial; otherwise there were no differences in the demographic or health data by study.

Table 1. Participant characteristics (N=20).

Characteristics	Values
Age (years), mean (SD)	59.0 (7.7)
Gender, n (%)	
Men	7 (35)
Women	13 (65)
Income (USD), n (%)	
<$20,000	11 (55)
>$20,000	2 (10)
Other/refused	7 (35)
Employment status, n (%)	
Part time	4 (20)
Unemployed	4 (20)
Disabled	9 (45)
Retired	3 (15)
Race or ethnicity, n (%)	
White	2 (10)
Hispanic or Latino	11 (55)
Black or African American	4 (20)
Asian or Pacific Islander	2 (10)
Other	1 (5)
Limited English proficiency, n (%)	13 (65)
Limited health literacy, n (%)	16 (80)
Education, n (%)	
None or primary school	7 (35)
High school graduate or GED[a]	2 (10)
Some college or technical school	9 (45)
College graduate or graduate degree	2 (10)
Diabetes, n (%)	14 (70)
Depression, n (%)	20 (100)
Health status, n (%)	
Fair or poor	13 (65)
Good	7 (35)

[a]GED: General Education Development.

Mobile Phone Use and Digital Literacy

We uncovered major categories of mobile phone usage related to texting networks, mobile phone carriage, and difficulty with the basic features of mobile phones (ie, text messaging). Overall, nearly all participants (18/20) had a smartphone and used text messaging regularly with a network of family and friends (15/20). A total of 5 participants described having limited texting networks, texting only family members or research studies. Most participants (17/20) reported carrying their mobile phones with them whenever leaving home.

Despite the overall high rate of mobile phone usage, many participants were defined as having LDL. Participants were categorized as having LDL if they had limited texting networks, frequently did not carry their mobile phones or turned them off for long periods of time (limited mobile phone carriage), or had difficulty using their phones or sending text messages owing specifically to unfamiliarity with or difficulty typing or using the microphone feature. Overall, 10 participants were categorized as having LDL; others were categorized as having adequate digital literacy. Those who described difficulty using their phones attributed this primarily to being new smartphone owners and still learning how to use their phones. Several also

cited their phones being turned off or forgetting to check their phones, low literacy, and having to use the microphone feature on their smartphone as hindrances to text messaging:

> *Now that I just got this phone, and before I was able to send text messages but now I've forgotten. [...] I didn't want a high-tech phone […] but my kids gifted me a phone and said, "you have to learn, mom."* [translated from Spanish]
>
> *I didn't check my phone or my phone was off for the day. Sometimes I'll turn my phone off and I forget to turn it back on.*
>
> *Yes, because I don't use the keys and sometimes when you speak into the microphone, it doesn't write the words that you want to say correctly. My daughter calls me out on that. So, if I'm going to send a message that's not going to be correct like it should be, I'd feel bad.* [translated from Spanish]

Of the 10 participants with LDL, 7 had LEP and 10 had LHL. There was a similarly high proportion of participants with LEP and/or LHL within each of the mobile phone usage domains we used to define the LDL (Table 2). Specifically, the majority of participants with limited texting network, limited mobile phone carriage, and difficulty using their phone or texting had LEP and/or LHL. Fisher exact test results were not statistically significant.

The other most commonly mentioned barriers to mobile phone usage that were not related to digital literacy included being too busy and forgetting to respond, having a physical impairment (eg, arthritis) that made typing difficult, and not being in the mood to respond to text messages:

> *Um, I got real busy, I didn't hear - I didn't check my phone or my phone was off for the day. Sometimes I'll turn my phone off and I forget to turn it back on and, um, then it's late and then I'm sleepy so I won't answer.*
>
> *It has such small letters and I have a problem with my hands because I have Raynaud's syndrome, where your hands fall asleep.* [translated from Spanish]
>
> *Sometimes, well, I don't pay attention and sometimes I do. Not every day. Sometimes I'm in a bad mood [laughs], sorry for saying it.* [translated from Spanish]

Table 2. Mobile phone digital literacy, as well as domains used to measure digital literacy by English proficiency and health literacy.

Mobile phone digital literacy	Total number of patients, N (%)	LEP[a], n (%)	LHL[b], n (%)	Both LEP and LHL, n (%)
Subdomains of mobile phone digital literacy				
Limited texting network	5 (25)	4 (80)	4 (80)	4 (80)
Limited mobile phone carriage	3 (15)	3 (100)	3 (100)	3 (100)
Difficulty using phone and/or texting	10 (50)	7 (70)	10 (100)	7 (70)
Overall limited digital literacy	10 (50)	7 (70)	10 (100)	7 (70)

[a]LEP: limited English proficiency.

[b]LHL: limited health literacy.

Engagement in Card Sorting Activity

Engagement was examined in 2 ways. First, we determined the frequency of feedback given about the existing content and the frequency of novel suggestions provided. We also used our qualitative coding to identify the extensiveness of both feedback on the existing content and novel suggestions.

Feedback on Existing Text Message Content

Feedback on text message content was elicited by a card sorting activity—half of the participants required a modification of the activity as they either did not understand it or had difficulty reading the cards in the allotted time.

The most common positive feedback was for the text messaging content to encourage participants to reflect on their behaviors or thoughts, give concrete ideas or advice, or be highly positive and motivating. The most common negative feedback was for the content that was viewed as repetitive. Several exemplar quotes demonstrated these positive and negative feedback themes:

> *But most of the times when I received the messages, they came at a good time and it just helped me to evaluate, um, you know, my thinking and my feelings, which if it hadn't came, then I wouldn't have been thinking about, you know, different things* [theme: encourage reflection]
>
> *These are all concrete suggestions of things that you can try because they will probably improve your mood just because you try. […] So I think it's just little suggestions or ways that you could help make yourself feel better. Less negative. Less down.* [theme: concrete advice]
>
> *I also liked the way they were positive.* [translated from Spanish; theme: positive messages]
>
> *They're a little repetitive.* [translated from Spanish; theme: repetitive messages]

For each participant, the number of different reasons they gave for liking or disliking sample text messages was summed, and this defined the frequency of feedback.

We also evaluated the extensiveness of feedback provided by each participant. Half of the participants provided extensive feedback, as evidenced by these comments:

> *[They] just make you stop for a minute. I think that's maybe the key to why it works so well because you*

know the message comes in and you're like well why am I doing this? I'm not being very positive today and I could do something better. I could think of another way to handle this. Maybe I should try that.

Well, they spark the idea of getting started. They spark the idea of why - I mean, they have started to continue in giving me I guess motivational nudges here and there.

It says "every time you think you can't or you won't be able to [do something], remember the times you were able to." Yes, there are times that these messages help me, this type of message helps me because often I say, "I can't," but occasionally I've read these messages and I say, "yes I can. "I'll try it no matter what," "I'll try it," and then I've done it and I have examples. [translated from Spanish]

On the other hand, participants who gave feedback that was minimal, provided little, if any, varied feedback, and did not explain why they liked or disliked the text messages even after being probed by the interviewer:

Yeah. Not interested. Okay.

I liked all of them. [translated from Spanish]

For quantitative analyses (Table 3), we calculated the frequency of feedback by coding each instance of feedback on the sample text messages. We found that the frequency of feedback was lower in participants with LEP, LHL, and LDL than in those with English proficiency, adequate health literacy, or adequate digital literacy, although none of these differences were statistically significant. It appears that the participants who were proficient in English provided more extensive feedback compared with those with LEP (5/7, 71% vs 4/13, 30%), as did those with adequate versus LHL (3/4, 75% vs 7/16, 44%) and those with adequate versus LDL (6/10, 60% vs 4/10, 40%). These differences approached but did not achieve statistical significance.

Table 3. Frequency and extensiveness of the feedback on existing text messaging content by English proficiency, health literacy, and digital literacy.

Patient characteristic	Frequency indicated by median number of unique feedback	Extensiveness indicated by number who provided extensive feedback, n (%)
English proficiency		
Adequate (n=7)	5.5	5 (71)
Limited (n=13)	4	4 (30)
Health literacy		
Adequate (n=4)	5.5	3 (75)
Limited (n=16)	4	7 (44)
Digital literacy		
Adequate (n=10)	5	6 (60)
Limited (n=10)	4	4 (40)

Novel Suggestions

Participants were asked whether they had other ideas for text messages that were not presented during the card sorting activity, as well as whether they had any other general feedback on the text messages or the text messaging interventions. The majority of suggestions were for text message content:

Um, or, um, "I know you're feeling down, so let's go for a walk and maybe that will pick you up a little bit. I've tried that before and it happens, so let's do it."

Hmm. "Rise and shine!" Uh, let's see, okay. "Rise and shine, get off your behind!"

A fewer number of participants provided feedback on broader, structural content of the text messaging interventions:

It could be a little cartoon running, an example of someone walking or running. [translated from Spanish]

There are thousands of illnesses in the world. So, first you need to know which illnesses someone has, and what difficulties they may have in doing physical activity. There are many types of physical activity. Because after my surgery my exercise was to stretch, to lift my arms with a pound of rice. It's moving the body. From here down I couldn't move because of my knee. Stretching, nothing more. Of course. This way, you have to see what the patient has to motivate them. [translated from Spanish]

We evaluated the extensiveness of novel suggestions as well. Extensive novel suggestions were defined as providing detailed content for possible text messages and also offering new types of messages:

And I would send them [a message] saying that life is very beautiful. Life is a thing of beauty and one should be happy and put aside the negative to give life to the positive. [translated from Spanish]

That the gym is not the only place where one can exercise, because you have this mentality of "I'll sign up but it's very expensive." If there are other options for exercise, that's excellent. [translated from Spanish]

Suggestions were categorized as *minimal or misaligned*, if they were vague or identical to the messages presented during card sorting or were unrelated to the physical activity content focus of the text messaging intervention (eg, text messages with the results of laboratory tests):

Probably reminding uh reminding me for doing me or exercise, walking, whatever.

I think that if you're going to send me one of these messages, it's to warn me about something that happened or is about to happen. I mean to say […] let me give an example. The doctor saw […] my results and wanted to tell me how they went. [translated from Spanish]

We coded each separate answer by participants as *novel suggestions*, and the frequency of novel suggestions for each participant was defined as the sum of the coded excerpts. The median number of novel suggestions was 2 per participant (range 0-9; Table 4), and there were no significant differences by English proficiency, health literacy, or digital literacy. All participants with adequate health literacy provided extensive novel suggestions, compared with only half of those with LHL (Table 4). There were no differences in the extensiveness of novel suggestions by English proficiency or digital literacy.

Table 4. Frequency and extensiveness of novel suggestions by English proficiency, health literacy, and digital literacy.

Patient characteristic	Frequency indicated by median number of suggestions	Extensiveness indicated by number that provided extensive suggestions, n (%)
English proficiency		
Adequate (n=7)	2.5	4 (57)
Limited (n=13)	2	6 (56)
Health literacy		
Adequate (n=4)	2	4 (100)
Limited (n=16)	2	8 (50)
Digital literacy		
Adequate (n=10)	2	5 (50)
Limited (n=10)	2	5 (50)

Discussion

Principal Findings

Our study found that the participants representing vulnerable populations engaged differently in the UCD process that we employed and described difficulty with mobile phone usage and text messaging. Although nearly all participants had a smartphone that they carried with them throughout the day and a large majority had wide texting networks, over three-quarters still described difficulty with text messaging. This discrepancy suggests that smartphone ownership, even with daily personal use, does not accurately predict comfort or ability in using the basic features of phones such as text messaging. This finding is supported by emerging data demonstrating limited use (eg, voice communication only) of mobile phones by vulnerable populations [26], as well as by efforts in mobile interaction design to focus on low literacy or illiterate populations [27].

We found variations in engagement in our design process by health literacy, English proficiency, and digital literacy. Digital literacy, defined broadly by the US Department of Education as "digital problem solving" and measured by assessments of basic computer competence, affects 16% of the US population; those who have LDL tend to be older, nonwhite, foreign born, and less educated than those who are digitally literate [28]. This definition, however, does not incorporate the use of other ubiquitous technologies, including smartphones, and has not been evaluated in the context of health and health care. Therefore, we used a de novo classification from the participants' responses to calculate our own definition of mobile phone digital literacy, informed by previous conceptual models in this space. Most individuals we identified as having LDL also had LHL and LEP, suggesting some overlap in these constructs. User characteristics such as age, education, and employment are considered in frameworks for developing information technology systems, and it is likely that health literacy and English proficiency are similarly important user characteristics in health-related digital literacy [25,29]. In fact, health literacy has been included as a component of digital literacy in some conceptual models [30].

Despite following best practice guidelines for card sorting, such as limiting the number of cards and categories and providing uniform, clear directions on how to conduct the task [17,31,32], this method had to be modified in half of the interviews and was therefore not fully effective in this study population. Eliciting novel suggestions was also difficult, with most participants only providing 1 or 2 suggestions. The latter is consistent with studies of engagement in qualitative methods that suggest that end users are often unable to provide anticipatory feedback [16]. However, we expected more consistency in the feedback on existing materials, as end users are usually more engaged in this type of feedback. This may be because in obtaining feedback via a card sorting method, we created the additional challenge of having to provide feedback on hypothetical messages within a set time limit, which has been shown to be a more challenging task [16]. It is also possible that the participant's response bias owing to the interviewer's demand characteristics contributed to the differences in feedback by health and digital literacy and English proficiency, though we tried to minimize this by conducting interviews in Spanish with native Spanish-speaking interviewers [33]. Future work is needed to replicate our findings in larger samples; however, our findings suggest that augmenting card sorting with different

methods in UCD, such as task analysis and usability testing with direct observation, may be more effective in obtaining feedback from vulnerable populations [18,25,29]. Furthermore, observation in environments in which end users are both comfortable and will be using the digital health intervention may be more productive in eliciting feedback and matching the intervention with the end users' needs and preferences.

In the development of digital health interventions to improve outcomes in patients with diabetes, others have also employed user-centered or participatory design and have obtained and successfully incorporated feedback in their development processes [34,35]. Although these studies have enrolled diverse populations, they do not describe whether the feedback they received differed within subgroups of their populations. Our exploratory analyses showed differences in engagement in specific subgroups of patients, suggesting that there may be other factors contributing to these differences. LHL, specifically, has been associated with lower patient activation and engagement in clinical settings, and this may be relevant in research settings as well [36-38]. Our findings underscore the importance of measuring health literacy, English proficiency, and digital literacy and considering their effects in UCD. This is critical as simply recruiting these vulnerable populations to user-centered or participatory design research rather than engaging them in feedback processes in a meaningful way carries the risk of exacerbating existing disparities in the use of HIT [8].

Limitations

There are several limitations to this study. In defining engagement, we assessed the extensiveness of feedback statements and novel suggestions; however, the provision of minimal feedback may have been owing to disinterest in the content being discussed rather than purely a lack of engagement. Given our small sample sizes that are typical in design research [31,32], many comparisons did not reach statistical significance; nevertheless, our quantitative data did allow us to highlight these differences. In addition, we focused specifically on mobile phone–related digital literacy rather than a more comprehensive digital literacy assessment, given the focus of our future text messaging intervention. Furthermore, 10 of the participants had already participated in a trial evaluating a text messaging app and therefore had more familiarity with this type of intervention. Nevertheless, neither successful completion of the card sorting task nor the engagement differed by study. Finally, although we chose to look specifically at health literacy, English proficiency, and digital literacy as the predictors of engagement, there are almost certainly other factors contributing to the differences in engagement noted in these populations, including education, race, culture, patient activation, and the nature and homogeneous content of the interviews [37,39]. Notably, all the study participants had depression, which has been associated with lower patient activation [40]; however, despite this, we were still able to detect differences in engagement within subgroups of the population.

Conclusions

Engagement in our design process varied by health literacy, English proficiency, and digital literacy. The participants of our study represent a diverse population—in race, employment, education, literacy and language, and general health status—that is rarely captured in usability studies for HIT. We believe this to be a major strength of this paper, both in describing the variations in patterns of mobile phone usage and in evaluating the engagement in UCD methods. Our findings highlight the need for a better understanding of how to consider, define, and incorporate digital literacy when developing HIT, as well as continued efforts in better engaging vulnerable populations in research. Our future work will report on the process for incorporating format and content feedback into our final intervention.

Acknowledgments

This project was funded by the Agency for Healthcare Research and Quality (1R01HS025429-01), a career development award by the National Institute of Mental Health (5K23MH0944442), a career development award by the National Cancer Institute (K24CA212294), and a research training grant by the National Research Service Award (T32HP19025). The sponsors did not play a role in the design, methods, subject recruitment, data collection, analysis, and preparation of manuscript.

Conflicts of Interest

None declared.

References

1. 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]
2. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mhealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
3. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

4. 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]
5. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
6. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
7. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and Evaluating mHealth Interventions for Vulnerable Populations: A Systematic Review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21-26, 2018; Montreal, QC, Canada p. 1-17. [doi: 10.1145/3173574.3173589]
8. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
9. Chakkalakal RJ, Kripalani S, Schlundt DG, Elasy TA, Osborn CY. Disparities in using technology to access health information: race versus health literacy. Diabetes Care 2014;37(3):e53-e54 [FREE Full text] [doi: 10.2337/dc13-1984] [Medline: 24558085]
10. Broderick J, Devine T, Langhans E, Lemerise AJ, Lier S, Harris L. National Academy of Medicine. 2014. Designing Health Literate Mobile Apps URL: https://nam.edu/wp-content/uploads/2015/06/HealthLiterateApps.pdf [accessed 2019-03-29] [WebCite Cache ID 77Ds5MvxG]
11. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15(Suppl 2):183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
12. Pew Research Center. 2019. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-10-17] [WebCite Cache ID 77DsReIss]
13. Aguilera A, Lyles CR. The case for jointly targeting diabetes and depression among vulnerable patients using digital technology. JMIR Diabetes 2017 Jan 17;2(1):e1 [FREE Full text] [doi: 10.2196/diabetes.6916] [Medline: 30291080]
14. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
15. Doherty G, Sharry J, Bang M, Alcañiz M, Baños R. Technology in Mental Health. In: CHI'08 Extended Abstracts on Human Factors in Computing Systems. 2008 Presented at: CHI EA'08; April 5-10, 2008; Florence, Italy p. 3965-3968. [doi: 10.1145/1358628.1358968]
16. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
17. Usability. What & Why of Usability: Card Sorting URL: https://www.usability.gov/how-to-and-tools/methods/card-sorting.html [accessed 2019-03-29] [WebCite Cache ID 77DsBkhk8]
18. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. Transl Behav Med 2018 Dec 7 (epub ahead of print). [doi: 10.1093/tbm/iby119] [Medline: 30535343]
19. Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P. Automated text messaging as an adjunct to cognitive behavioral therapy for depression: a clinical trial. J Med Internet Res 2017 May 8;19(5):e148 [FREE Full text] [doi: 10.2196/jmir.6914] [Medline: 28483742]
20. Farris KB, Salgado TM, Batra P, Piette JD, Singh S, Guhad A, et al. Confirming the theoretical structure of expert-developed text messages to improve adherence to anti-hypertensive medications. Res Social Adm Pharm 2016;12(4):578-591 [FREE Full text] [doi: 10.1016/j.sapharm.2015.09.009] [Medline: 26525857]
21. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
22. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? J Am Med Assoc 2010 Jul 7;304(1):76-84. [doi: 10.1001/jama.2010.896] [Medline: 20606152]
23. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. J Gen Intern Med 2011 Mar;26(3):265-271 [FREE Full text] [doi: 10.1007/s11606-010-1552-1] [Medline: 21057882]
24. Karliner LS, Napoles-Springer AM, Schillinger D, Bibbins-Domingo K, Pérez-Stable EJ. Identification of limited English proficient patients in clinical care. J Gen Intern Med 2008 Oct;23(10):1555-1560 [FREE Full text] [doi: 10.1007/s11606-008-0693-y] [Medline: 18618200]

25. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through ehealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
26. Thies IM. SIGCHI Social Impact Award Talk -- Designing for Low-Literate Users. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017 Presented at: CGI EA'17; May 6-11, 2017; Denver, Colorado, USA p. 8-9. [doi: 10.1145/3027063.3058589]
27. Belay EG, McCrickard DS, Besufekad SA. Designing Mobile Interaction for Low-Literacy (D-MILL). In: Proceedings of the First African Conference on Human Computer Interaction. 2016 Presented at: AfriCHI'16; November 21-25, 2016; Nairobi, Kenya p. 251-255. [doi: 10.1145/2998581.2998614]
28. Mamedova S, Pawlowski E. National Center for Education Statistics (NCES). 2018. A Description of US Adults Who Are Not Digitally Literate URL: https://nces.ed.gov/pubs2018/2018161.pdf [accessed 2019-03-29] [WebCite Cache ID 77DsLCqma]
29. Kushniruk A, Turner P. A framework for user involvement and context in the design and development of safe e-health systems. Stud Health Technol Inform 2012;180:353-357. [doi: 10.3233/978-1-61499-101-4-353] [Medline: 22874211]
30. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
31. Paul CL. A modified Delphi approach to a new card sorting methodology. J Usability Stud 2008;4(1):7-30 [FREE Full text]
32. Wood JR, Wood LE. Card sorting: current practices and beyond. J Usability Stud 2008;4(1):1-6 [FREE Full text]
33. Dell N, Vaidyanathan V, Medhi I, Cutrell E, Thies W. Yours is Better: Participant Response Bias in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5-10, 2012; Austin, TX, USA p. 1321-1330. [doi: 10.1145/2207676.2208589]
34. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The design, usability, and feasibility of a family-focused diabetes self-care support mhealth intervention for diverse, low-income adults with type 2 diabetes. J Diabetes Res 2016;2016:7586385 [FREE Full text] [doi: 10.1155/2016/7586385] [Medline: 27891524]
35. Zachary WW, Michlig G, Kaplan A, Nguyen NT, Quinn CC, Surkan PJ. Participatory design of a social networking app to support type II diabetes self-management in low-income minority communities. Proc Int Symp Hum Factors Ergon Healthc 2017 Jun;6(1):37-43 [FREE Full text] [doi: 10.1177/2327857917061010] [Medline: 31157286]
36. Aboumatar HJ, Carson KA, Beach MC, Roter DL, Cooper LA. The impact of health literacy on desire for participation in healthcare, medical visit communication, and patient reported outcomes among patients with hypertension. J Gen Intern Med 2013 Nov;28(11):1469-1476 [FREE Full text] [doi: 10.1007/s11606-013-2466-5] [Medline: 23690237]
37. Martin LT, Schonlau M, Haas A, Derose KP, Rosenfeld L, Buka SL, et al. Patient activation and advocacy: which literacy skills matter most? J Health Commun 2011;16(Suppl 3):177-190 [FREE Full text] [doi: 10.1080/10810730.2011.604705] [Medline: 21951251]
38. McCormack L, Thomas V, Lewis MA, Rudd R. Improving low health literacy and patient engagement: a social ecological approach. Patient Educ Couns 2017 Jan;100(1):8-13. [doi: 10.1016/j.pec.2016.07.007] [Medline: 27475265]
39. Bledsoe KL, Hopson RK. Conducting ethical research and evaluation in underserved communities. In: Mertens DM, Ginsberg PE, editors. The Handbook of Social Research Ethics. Thousand Oaks, California: Sage Publications; 2013:391-406.
40. Blakemore AH, Hann M, Howells K, Panagioti M, Sidaway M, Reeves D, et al. Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. BMC Health Serv Res 2016 Oct 18;16(1):582 [FREE Full text] [doi: 10.1186/s12913-016-1843-2] [Medline: 27756341]

Abbreviations

DIAMANTE: Diabetes and Depression Text Messaging Intervention
HIT: health information technology
LDL: limited digital literacy
LEP: limited English proficiency
LHL: limited health literacy
UCD: user-centered design

Edited by G Eysenbach; submitted 05.04.19; peer-reviewed by L Richardson, M Tremblay, S Chokshi; comments to author 01.06.19; revised version received 11.06.19; accepted 11.06.19; published 29.08.19.

Please cite as:
Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR
Assessing Mobile Phone Digital Literacy and Engagement in User-Centered Design in a Diverse, Safety-Net Population: Mixed Methods Study
JMIR Mhealth Uhealth 2019;7(8):e14250
URL: http://mhealth.jmir.org/2019/8/e14250/
doi:10.2196/14250
PMID:31469083

Original Paper

Mobile Helpline and Reversible Contraception: Lessons From a Controlled Before-and-After Study in Rural India

Sangita Kulathinal[1*], PhD; Bijoy Joseph[2*], PhD; Minna Säävälä[3,4*], MSocSci, PhD

[1]Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

[2]Akra-Numero Ltd, Helsinki, Finland

[3]The Family Federation of Finland (Väestöliitto), Helsinki, Finland

[4]Social and Cultural Anthropology, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland

[*]all authors contributed equally

Corresponding Author:
Minna Säävälä, MSocSci, PhD
The Family Federation of Finland (Väestöliitto)
Kalevankatu 16
Helsinki, 00100
Finland
Phone: 358 405736557
Email: minna.saavala@vaestoliitto.fi

Abstract

Background: Researchers and activists have expressed concerns over the lack of availability and nonuse of reversible, modern, contraceptive methods in India for decades. New attempts to increase access, availability, and acceptance of reversible contraceptives need to be developed, instead of relying solely on female sterilization. Mobile health (mHealth) initiatives may offer one way to serve underprivileged populations who face challenges in sexual and reproductive health (SRH) in countries such as India.

Objective: This study aimed to examine the outcome of an mHealth intervention for enhancing knowledge of, and practices related to, reversible contraceptives in rural Western India.

Methods: We implemented a nonrandomized controlled trial (before-and-after study in an intervention area and a control area) in the Indian state of Maharashtra. The intervention in this case was a mobile-based SRH helpline provided by a nongovernmental organization (NGO). Baseline and follow-up surveys were carried out in two government-run primary health center areas, one each in the intervention and control area, and 405 respondents were surveyed in the two rounds. An interview-based structured questionnaire suitable for a low-literacy environment was used to collect data. The effect of the intervention was estimated using logistic regression, adjusted for gender, by calculating robust standard errors to take into account the clustering of individuals by the area (intervention or control). In each regression model, the effect of intervention was estimated by including a term for interaction between the intervention area and the period before and after the intervention. The exponent of the regression coefficient of the interaction term corresponding to the period after the intervention, along with the 95% CI, is reported here. The odds ratio for the control village multiplied by this exponent gives the odds ratio for the intervention village. Calls received in the intervention were recorded and their topics analyzed.

Results: The current use of reversible contraception (18% increase in intervention area vs 2% increase in control area; 95% CI) has seen changes. The proportion of respondents who had heard of contraception methods from an NGO rose in the intervention area by 23% whereas it decreased in the control area by 1% (95% CI). However, the general level of awareness of reversible contraception, shown by the first contraceptive method that came to respondents' mind, did not improve. Demand for wider SRH information beyond contraception was high. Men and adolescents, in addition to married women, made use of the helpline.

Conclusions: A mobile helpline that one can confidentially approach at a time most convenient to the client can help provide necessary information and support to those who need reversible contraception or other sexual health information. Services that integrate mHealth in a context-sensitive way to other face-to-face health care services add value to SRH services in rural India.

(JMIR Mhealth Uhealth 2019;7(8):e12672) doi:10.2196/12672

KEYWORDS

contraception behavior; family planning services; organizations, nonprofit; cell phone use; mobile phone; information seeking behavior; mHealth; call center; South Asia; India

Introduction

Background

In India, a few reproductive mobile health (mHealth) initiatives have been implemented [1-6], of which the most popular is the Central Government's Mother and Child Tracking System [7]. However, the use of mHealth in family planning and contraceptive services has thus far been limited. In addition, research evidence on mHealth interventions and programs in the field of sexual and reproductive health (SRH) in India remains scarce, although mHealth offers great promise to potentially help cater to the needs of people with limited health care and family planning services [4,5,7]. Telemedicine, emergency services, text messaging services, supervision and support services to the health care service staff, and data collection are among the functions that mobile phones have brought in to improve reproductive health services in less developed countries [5,8].

With its population of 1.3 billion people, India has implemented the Indian family planning program, which relies heavily on female sterilization. Of currently married women, aged 15 to 49 years, 36% are sterilized whereas 11% use reversible modern contraception [9]. Reversible modern contraceptive methods in India mainly include condoms, intrauterine contraceptive devices (IUCDs), oral contraceptive pills and injectable contraceptives. Researchers and activists have expressed concern over the lack of availability and nonuse of reversible modern contraceptive methods in India since decades. The Indian Government expressed its changing focus toward reversible methods in 2012 [10] by expanding the choice of methods, especially encouraging the postpartum adoption of IUCDs. It is expected that increased use of reversible contraceptive methods will help in reducing both maternal and infant mortality and morbidity, as well as slow down population growth by lengthening birth intervals.

The use of reversible contraception remains abysmally low in India. Among the underlying structural factors are the generally low socioeconomic and educational standards [11], and gender and generational asymmetries [12-15]. Practically, lacking information on reversible contraceptive methods or fears related to side effects [16], inaccessibility or poor quality of care [17], and provider-imposed barriers [18] hinder adoption of contraception, and even more so adoption of reversible methods.

Consequently, there is a need to provide personal counseling and information in a gender-sensitive manner, secure access to contraceptives and improve health care services, particularly among the socioeconomically underprivileged groups. Can mHealth assist in providing counseling and information and improve accessibility of contraception? Wireless phone subscribers in India reached a total of 998 million people by March 2018 [19], and mobile phones are increasingly affordable and accessible even among the poorest in rural India [20]. Mobile technology is thus a potential means to also reach out to the disadvantaged.

Purpose of the Study

This study explores gaps in SRH needs of rural, disadvantaged populations, and further examines the outcome of an mHealth intervention in enhancing knowledge and practice of reversible contraception to reduce the gap in rural Western India. An mHealth-based intervention on family planning is examined in a nonrandomized controlled trial (before-and-after study in an intervention area and a control area) in the Indian state of Maharashtra. The intervention was a project by a nongovernmental organization (NGO) offering mobile-based SRH helpline.

Methods

Study Setting

In terms of family planning, the Indian state of Maharashtra represents close to the average Indian situation. In the state, 51% of currently married women are sterilized whereas 11.5% use reversible modern contraceptive methods of which the condom is the most popular. Total unmet need for contraception is 9.7% [21].

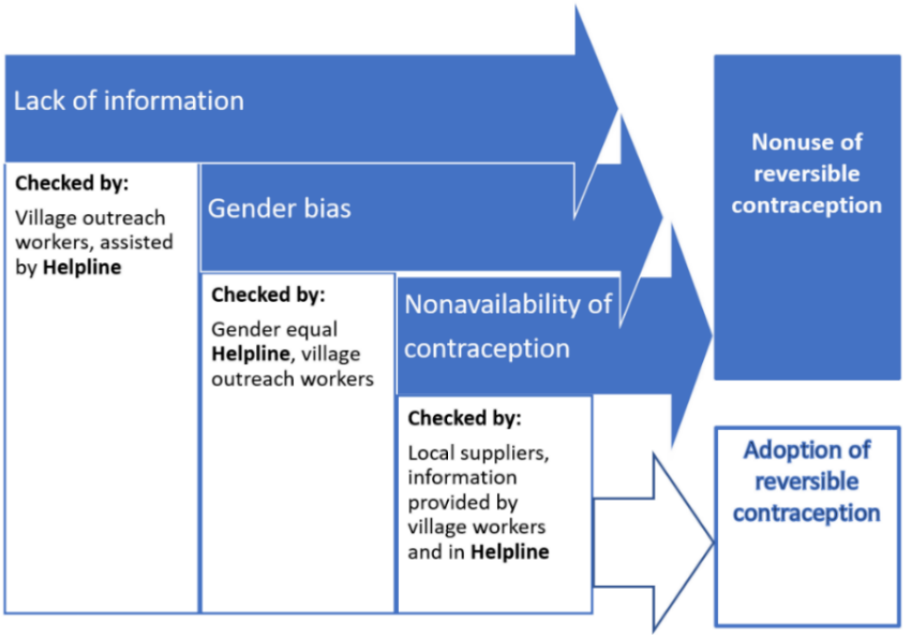
Overall, two sufficiently similarly profiled districts, Thane and Nashik of Maharashtra, were chosen for the study. Although Thane is more urban than Nashik, the subdistricts where primary health center (PHC) areas were chosen for the data collection have a similar level of urbanization (77% in the chosen subdistrict within Thane and 78% in the chosen subdistrict within Nashik [22]). Overall sociodemographic characteristics and contraceptive use patterns were similar enough in the two districts. A PHC area in rural Thane was the intervention area and a PHC area in rural Nashik was the comparable nonintervention area which provided the control group for the study.

Intervention

The intervention introduced a mobile helpline, combined with personal contact with participants by village health workers and with local distribution of contraceptives. The target population was married men and women in the age range of 15 to 35 years.

A toll-free helpline was available to villagers in the project area. A total of 12 gender-equal, frontline, field workers were responsible for communication and branding activities for the call center, and follow-up activities among targeted beneficiaries who sought the intervention services, using both interpersonal communication and mid-media activities (eg, street theater and wall paintings). The calls were attended by trained female and male paramedical staff, one of each, who were fluent in the local dialect. The helpline personnel recorded the received calls, specifically noting the main topic of discussion. This intervention complemented the regular governmental health care services available, and it took place from June 2015 to June 2016.

Figure 1. The role of the helpline in the intervention's theory of change (hindrances in dark arrows and intervention activities in white boxes).



The intervention was based on the theory of change in a situation, when a woman or a couple was motivated to delay or avoid pregnancy, as expressed in Figure 1. On the basis of desk research and grassroots activities in another area, the project identified the following three main hindrances to the adoption of reversible contraceptive methods in a situation when contraception was in demand for a woman or couple: lack of information, gender bias, and unavailability or inaccessibility of contraceptives. Lack of information was to be addressed by providing a mobile helpline and marketing its services, gender bias was to be addressed by providing the helpline, which was an easily accessible source of support and information for the young and the women irrespective of elders' and men's acceptance, and finally, the supply of contraception issue was addressed in the mobile helpline by providing information on the local points of access to condoms, pills, and coils. However, the mobile helpline had to be supported by village workers and could not be implemented without initial personal contact with locals.

The Control Area

The control site was a PHC area in Nashik, an adjacent district. This area was serviced by similar governmental maternal and health care services, including family planning services. For several years, there was a mother and child health program being implemented by an NGO, whose local village workers supported governmental services. Thus, both areas were served by governmental public health, assisted by NGO-based services that involved personal contact with the local people. In the control area, there was no mobile helpline comparable with the one in the intervention area.

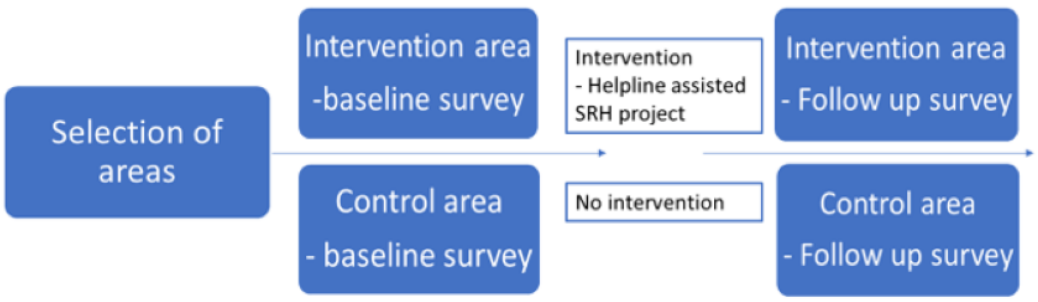
Study Design

The study was a quasi-experimental study, a controlled before-and-after study as described by Reeves et al [23]. The study sample was derived from clusters, one of which was the intervention site, and the other was control site.

Before Study

An interview-based baseline survey was carried out from March to April 2015 in both the study areas to gather data on knowledge of, and practices related to, reversible contraceptive methods, needs related to SRH, and willingness to use mHealth services. The study design is shown in Figure 2.

Figure 2. Study design. SRH: sexual and reproductive health.



After Study

To assess the outcome of the mobile helpline on SRH, a follow-up survey was carried out from November to December 2016. The follow-up survey was conducted in both areas about 18 months after the introduction of mHealth-assisted SRH services in the intervention area. The influence of external activities and information, such as government services and media programs toward family planning, was hypothesized to be similar in both areas. An NGO working with issues related to SRH was present in both areas but a mobile helpline was promoted only in the intervention area. Wider socioeconomic confounders were controlled using the control area. Hence, any differences in SRH observed between the areas after the intervention were analyzed as related to the potential outcome of the helpline intervention.

Study Sampling and Recruitment

The target sample size was 100 married men and 100 married women from each area. The intervention survey was intended to represent the area or cluster level and not the individual level, meaning that the study participants in the two survey rounds were not necessarily the same persons.

The sampling was carried out in two stages. A total of 10 villages (about 50% from each area) were randomly selected from each selected PHC using systematic sampling with a random start. A list of villages with a total population size of at least 500, which were arranged in ascending population sizes, served as the sampling frame from which sample selection was made. From each sampled village, 20 households were selected by a systematic sampling process using the left-hand rule, and 1 study subject was selected per household. If a household had more than 1 eligible subject, then one of them was chosen randomly. Trained research investigators were responsible for recruiting the study subjects. The inclusion criteria for selection of a subject were: age between 15 and 35 years, married, and lived permanently in the settlement. If the respondent was a visitor (eg, a daughter come down for delivery to her natal home or a visiting guest) then she was excluded from the study and the interview was terminated.

Measurements and Outcomes

This study used an interview-based data collection method, a structured questionnaire which suits low-literacy environments. The primary measurements were knowledge and practice of family planning, current use of contraception, intent for further use, and any changes in these after the intervention. The questionnaire included background socioeconomic and demographic information, questions on mobile phone ownership and use, union status and hygiene practices; knowledge, attitude, and practice of family planning; children and decision making, current use of contraception, intent for further use, and a contraceptive tracking sheet showing dynamics of contraceptive use of the study subjects. The data collection tool was first developed in English and then translated into the local language Marathi. A 3-day intensive training program was organized for 14 interviewers. A team of interviewers always included both a man and a woman. The interviews were conducted in privacy in the local (Marathi) language.

The main issue examined was whether a mobile helpline would improve people's awareness of reversible contraceptive methods, and as a result of this would the use of such methods increase? The awareness of reversible contraception was operationalized by the question: *When you think of a family planning method, which is the first method that comes to your mind?*

The proportion of responses, other than female sterilization, was taken as a measure of awareness of reversible methods among the respondents. The respondents were also asked to give a list of all other contraceptive methods that they knew without being prompted. To examine the contraceptive prevalence and use of reversible methods, the respondent was first asked if she or he was currently doing anything to prevent or postpone her, or his wife's, next pregnancy and then what they were currently doing to prevent it. The responses were coded as reversible methods (condom, oral contraceptive pills, IUCDs, and injectable contraceptives), permanent methods (female and male sterilization), and other methods (eg, withdrawal, abstinence, and herbs).

Data Analysis

The intervention's potential influence on the use of contraception based on the level of knowledge on contraceptive methods and SRH, access to contraception, and acceptability of mHealth support for SRH, were analyzed. The effect of intervention was estimated using logistic regression, adjusted for gender, by calculating robust standard errors to take into account clustering of individuals by the area (intervention or control). In each regression model, the effect of intervention was estimated by including a term for interaction between the intervention area and the period before and after the intervention. The exponent of the regression coefficient of the interaction term corresponding to the period after the intervention, along with the 95% CI, is reported here. The odds ratio for the control village multiplied by this exponent gives the odds ratio for the intervention village. The difference between the proportions after and before for each area is also reported. All analyses were performed with the statistical environment R (a free software environment supported by the R Foundation for Statistical Computing) [24].

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki of 1975, as revised in 2000, as well as with all applicable legal regulations governing data collection in India. The data did not contain any biological specimens but instead consisted of data on individual experiences, practices, attitudes, and knowledge, in addition to personal socioeconomic and demographic information. Informed consent was obtained in writing from every study subject after the nature and possible consequences of the study were explained. Participation was voluntary. All respondents were married. In the event of a respondent being below 18 years of age, consent was taken from a parent or spouse to include the underaged respondent in the study, with complete details of the study given to them, in addition to reading out the verbal consent form to the respondent.

The study was implemented in relation to the activities of two local NGOs [25] that gave their approval for the data collection. The study was approved by the local government authorities. As the study was carried out by independent scholars, the option of an institutional review board acceptance by a university was not available. Thus, an *ad hoc* ethical committee, consisting of three Indian members with expertise in social work and law and a social and health scientist, was formed with the help of the two NGOs to review the ethics of the study protocol and data collection plan. The ethical committee consented to the data collection and signed a statement of their approval.

Results

Descriptive Statistics

Table 1 presents background information of the surveys and respondents.

Table 1. Descriptive characteristics (source: baseline and follow-up surveys in two public health center areas in Maharashtra).

Characteristics	Intervention area	Control area
Sample size, n		
Baseline survey		
Women	103	100
Men	102	100
Total	205	200
Follow-up survey		
Women	90	101
Men	88	101
Total	178	202
Respondent age (years), median (range), baseline survey		
Women	26 (24-30)	26 (22-30)
Men	29 (26-34)	29 (26-33)
Respondent age (years), median (range), follow-up survey		
Women	27 (25-31)	27 (23-31)
Men	30 (27-35)	30 (27-34)
Age at marriage (years), median (range), baseline survey		
Women	19 (13-28)	18 (14-30)
Men	22 (17-30)	21 (16-29)
Age at marriage (years), median (range), follow-up survey		
Women	19 (15-28)	18 (12-28)
Men	23 (17-30)	22 (18-29)
Literacy[a] rate, n (%)		
Women[b]	87 (84.4)	76 (74.5)
Men[b]	86 (84.3)	90 (90.0)
Disadvantaged group, n (%)[b]		
Scheduled tribes and scheduled castes	132 (64.3)	150 (75.4)
Other backward classes	45 (22.0)	34 (17.1)
Access to mobile phone in household, n (%)[c]		
Women	87 (84.5)	93 (93.0)
Men[c]	81 (79.4)	86 (86.0)

[a]Able to both read and write, according to own statement.

[b]Information collected in baseline survey only.

[c]Having at least one mobile phone in household.

A total of 405 married men and women aged 15 to 35 years participated in the baseline survey before the intervention, of which 200 were from the control area. Gender-wise distributions are presented in Table 1. In the postintervention survey, the total number of participants was 380, of which 202 were from the control area. 87.3% (152/179) and 50.0% (101/202) of the participants were in both the surveys of intervention and control areas, respectively.

The median age of men and women was 29 and 26 years, respectively, in both areas before the intervention. It was 30 and 27 years, respectively, during the follow-up survey (detailed descriptive tables are available in [26]). The quartile age ranges for men and women in the baseline survey were 26 to 34 and 22 to 30, respectively. Median age at marriage for women was 18 years. Most men (84.4%; 87/103) and women (84.3%, 86/102) in the intervention area could read and write, whereas slightly more men (90.0%; 90/100) and somewhat less women (74.5%, 76/102) could do so in the control area. Literacy was recorded as reported by the study subjects; it was recorded separately for reading and writing, and it did not differ significantly for men and women.

In both areas, the caste composition was characterized by a high proportion of socially disadvantaged populations. Both areas have atypically high proportions of tribal populations, and in the Thane district, the intervention area, the proportion is even higher than in Nashik. Scheduled caste (formerly called the untouchables) and scheduled tribe populations are the most disadvantaged social categories in India and their combined proportion was 64.3% (132/205) in the intervention area, and 75.4% (150/199) in the control area. In both areas, the main occupation was farming and unskilled labor. A majority of the respondents in both areas reported having access to a mobile phone, with only 21.6% (21/102) of men and 15.5% (16/103) of women reporting not having a mobile phone in the household in the intervention area, and even less in the control area, with 14.0% (14/100) of men and 7.0% (7/100) of women reporting similarly. The difference between the proportion of men (P=.30) and women (P=.10) having access to mobile phones in the two study areas was not significant. The fact that women appear to have slightly more access to mobile phones than men seems somewhat surprising.

The summary statistics of the survey responses (n [%] for men, women, and all) are provided in Table 2. Both areas manifested a considerable increase in the general awareness of contraception: in the follow-up survey, 75.8% (135/178) of respondents in the intervention area and 81.6% (160/196) in the control area were aware of means to avoid pregnancy. The proportion of those who had heard of a contraceptive method from an NGO rose in the intervention area from 5.0% (6/121) to 33.5% (59/176), whereas in the control area, it was 1.7% (2/118) before and 1.2% (2/172) after. The proportion of respondents using reversible contraception rose in both areas.

Table 2. Knowledge on contraceptive methods and sexual and reproductive health, access to contraception, and acceptability of mobile health support for sexual and reproductive health by study areas before and after the intervention (source: baseline and follow-up surveys, two primary health care areas in Maharashtra).

Outcome	Intervention area, n (%)		Control area, n (%)	
	Before	After	Before	After
Has heard of any contraception method				
Men	49 (48.0)	76 (86.4)	46 (46.0)	85 (84.2)
Women	72 (69.9)	59 (65.6)	72 (72.0)	75 (78.9)
All	121 (59.0)	135 (75.8)	118 (59.0)	160 (81.6)
Has heard of contraception method from NGO[a]				
Men	3 (6.1)	16 (18.6)	2 (4.8)	2 (2.2)
Women	3 (4.2)	43 (47.8)	0 (0.0)	0 (0.0)
All	6 (5.0)	59 (33.5)	2 (1.7)	2 (1.2)
Reversible method first mentioned[b]				
Men	28 (59.6)	5 (6.8)	24 (52.2)	38 (46.3)
Women	22 (32.4)	14 (21.9)	43 (67.2)	36 (48.0)
All	49 (42.6)	19 (13.8)	67 (60.9)	74 (47.1)
Uses contraception now				
Men	30 (29.4)	54 (61.4)	54 (54.0)	25 (24.8)
Women	59 (57.3)	37 (41.1)	69 (69.0)	55 (57.9)
All	89 (43.4)	91 (51.1)	123 (61.5)	80 (40.8)
Using reversible method[c]				
Men	8 (26.7)	22 (40.7)	15 (27.8)	14 (56.0)
Women	14 (23.7)	17 (45.9)	26 (37.7)	14 (25.5)
All	22 (24.7)	39 (42.9)	41 (33.3)	28 (35.0)
Willing to call sexual health helpline[d]				
Men	44 (56.4)	88 (100.0)	84 (95.5)	88 (91.7)
Women	100 (100.0)	86 (95.6)	75 (81.2)	88 (94.7)
All	144 (80.9)	174 (97.8)	159 (88.3)	176 (93.1)

[a]NGO: nongovernmental organization.

[b]First contraceptive method that comes to mind is a reversible method.

[c]Using some method other than sterilization, that is, a reversible method.

[d]Willing to call a male or female health worker to anonymously ask about sexual problems.

The Intervention and Contraception

Table 3 shows the changes in the outcomes (proportions) in the intervention and control areas, and the exponent of the regression coefficient corresponding to the interaction of intervention area and the period after intervention, adjusted for gender. In terms of practice, the change in the direction toward reversible methods is evident in the intervention area. Both the current use of contraception (8% increase in the intervention area vs 21% decrease in the control area) and the use of reversible contraception (18% increase in the intervention area vs 2% increase in the control area) have increased in the intervention area compared with the control area.

The general level of awareness of reversible contraception, shown by the first contraceptive method that came to respondents' mind being a reversible method, neither improved in the intervention area nor in the control area. On the contrary, fewer respondents in both areas mentioned a reversible method as the first method that came to their mind (decrease of 29% in the intervention area vs decrease of 14% in the control area). This means that, in the follow-up survey in both areas, it had become more common to mention female sterilization as the first method that came to mind. In the intervention area, the acceptability of contacting a helpline for SRH needs rose by 17% compared with the control area's increase of 5%.

Table 3. Odds ratios (95% CIs) for interaction between intervention area and period after intervention and mean change for intervention versus control area on knowledge and practice of sexual and reproductive health (source: baseline and follow-up surveys, two primary health center areas in Maharashtra).

Outcome	Odds ratio (95% CI)	Change (after and before), intervention versus control area
Has heard of any contraception method	0.85 (0.844-0.855)	20 versus 23
Has heard of contraception method from NGO[a]	14.203 (13.265-15.208)	29 versus −1
Reversible method first mentioned[b]	0.363 (0.36-0.366)	−29 versus −14
Uses contraception now	3.207 (3.037-3.388)	8 versus −21
Using reversible method[c]	2.053 (1.856-2.271)	18 versus 2
Willing to call sexual health helpline[d]	6.038 (4.7-7.759)	17 versus 5

[a]NGO: nongovernmental organization.

[b]The first contraceptive method that comes to mind is a reversible method.

[c]Using method of contraception other than sterilization or traditional method, that is, a reversible method.

[d]Willing to call a male or female health worker to ask anonymously about sexual problems.

In the intervention area, 274 individuals aged between 15 and 52 years made a total of 964 calls to the mobile helpline on issues pertaining to SRH. This meant that repeat calls averaged around 3.5 calls per client. The typical questions dealt with were as follows:

- Information about family planning and contraceptive methods (their contraindications and side effects if any) and access to services (where, how far, and what cost): 31.0% (85/274) of calls.
- Sexual health issues such as itching in genitals, concerns regarding masturbation, wet dreams, condoms access, and use: 43.0% (118/274) of calls.
- Issues pertaining to maternal and child health (vaccination schedules, nutrition-related questions, and other related matters): 13.1% (36/274) of calls.
- Menstrual health and hygiene: 13.1% (36/274) of calls.

Questions on masturbation, condoms, and symptoms in genitals were mainly from men, and questions on menstrual hygiene were from adolescent girls. Questions on family planning and other contraceptive methods, other than condoms, came mainly from married women.

Discussion

Principal Findings

Both the current use of contraception and the use of reversible contraception have increased in the intervention area compared with the control area. Contraceptive knowledge increased in both areas, whereas fewer people first mentioned a reversible method in the follow-up survey than in the baseline survey, in both areas. Thus, the study suggests that the mobile helpline has had a bearing on the practice of contraception, whereas there is less evidence of effect on knowledge.

This apparent lack of improvement in awareness of reversible contraceptives might relate partly to the measurement used for the awareness: whether the first method that came to a respondent's mind was a reversible method. In India, the universal connotation of family planning is female sterilization, a permanent method, which might turn out to be slow to change. Other types of questions to measure awareness of reversible methods might have painted a different picture. Moreover, this growing awareness of female sterilization compared with reversible methods in the follow-up survey might reflect governmental health service campaigns for female sterilization which the health workers were discussing in the communities in the months just before the follow-up survey, leading to an influence on the answers in the survey.

The study found that contraceptive use (both reversible methods and sterilization combined) at the time of the survey increased in the intervention site whereas it fell in the control site. This decrease found in the control area is rather unusual and needs further examination. In addition, the latest National Family Health Survey [21] showed a slight decline in the prevalence of family planning use among currently married women in Maharashtra. However, the decline witnessed in this study in the control area is more pronounced. It is unlikely that demand for family planning would have so strongly reduced within 18 months. It is possible that a temporary structural or administrative condition in the control area, such as reduced supply of contraceptives or absence of health care service personnel, could have caused a decline in contraceptive prevalence. However, this decrease in general contraceptive use (in practice, of female sterilization) in the control area does not explain away the increase in reversible contraception in the intervention area. Reversible contraception remained about as popular in the control area at both points of measurement, whereas in the intervention area the popularity of reversible contraception increased considerably. The adoption of female sterilization and reversible contraception are governed partly by different dynamics, with sterilization being a terminal method while reversible methods are mainly used for spacing births. This means that their developments are not necessarily interrelated. Consequently, the main result of the analysis, that the mobile SRH helpline is associated with increased use of reversible contraceptives, would not have to be compromised despite the decline in general contraceptive prevalence in the control area.

In both the intervention and in the control areas, NGOs working with issues related to SRH were present along with the standard local governmental services. The two NGOs provide basic

information and services related to maternal and reproductive health to the relatively underprivileged populations of the two areas. The main difference between the two areas was the presence and promotion of a mobile helpline on SRH only existing in the intervention area. The results show that the intervention building upon a mobile helpline on SRH was associated with the increased use of reversible contraception compared with the control area. Although awareness of family planning generally strengthened, the awareness of reversible methods did not increase in the intervention area more than in the control area. This would imply that the governmental services and information channels have increased local people's awareness on female sterilization in both areas.

A review of studies dealing with the impact of interventions that aim at influencing knowledge, attitudes, beliefs, and discussions regarding family planning and in increasing contraceptive use found that impacts are often a result of programs that have considered the importance of varied approaches to reaching women and couples [27]. The intervention examined here manifested this approach by including face-to-face activities, an anonymous helpline, and support to contraceptive availability in local settings.

The helpline received acceptance in the intervention area so that a 17–percentage point rise in the acceptability of such service provided by an NGO in case of SRH problems was evident. Villagers and helpline customers had expressed their appreciation of the fact that they could call whenever it was a convenient time and place for them. Most rural families in the project area, as in much of India, share a phone so that it stays at home and many family members can use it [20]. The need to consider mHealth interventions from a relational perspective, not only as the choice of an individual, becomes essential here [28]. Thus, a mobile helpline that can be confidentially approached at a time most convenient to the client is essential for a successful mHealth service. Clients particularly preferred a voice-based personal service, as they felt insecure about the terminology on issues of SRH. Many could not write or read fluently in the local language, Marathi. The experiences from this mHealth intervention study point to the need for program developers and designers to explore contextual and implementation factors seriously [6].

A considerable proportion of the questions raised in the mobile helpline touched on sexuality and issues other than contraception. There is clearly a great need in the area to provide more general SRH services and information than only family planning services to the local population, as well as to men and adolescents. A comprehensive SRH approach works better than a narrow family planning approach in winning local people's confidence, which is essential for sustainable results.

It was evident from the intervention that among rural, socioeconomically underprivileged populations, women and men have an unmet need for reversible family planning methods [29]. Although the governmental programs have started to pay more attention to contraceptive choices instead of sole reliance on female sterilization, there is still a long way to go. This intervention study shows that services that integrate mHealth in a context-sensitive way to face-to-face health care services can provide better results in rural India and assumedly also in other contexts in less developed societies.

Strengths and Limitations

This study was a rare attempt to examine the outcome of an mHealth intervention by making use of a control area. However, the conclusions would have been stronger if the samples would have been larger and the time frame would have been longer. The outcome was assessed after only half a year after the end of the intervention, which makes it difficult to say much about the perseverance of the changes. Changes in perceptions take a longer time to emerge in a measurable form.

The apparent inconsistency in the intervention area findings, in that knowledge of reversible methods seems to have fallen but the use of reversible contraception nevertheless increased substantially, can point to the problems in the instrumentalization of knowledge on reversible methods. Relying on the first thing that comes to mind question may have rendered the study vulnerable to a government campaign on female sterilization. If another measure of knowledge on reversible contraception were used this apparent inconsistency might have disappeared. The study design was neither a true panel design (interviewing the same respondent both in the baseline and the follow-up surveys) nor purely a successive cross-sectional design. Of all the respondents in the follow-up survey in the intervention area, 85% had also been interviewed in the baseline, whereas this proportion was 50% in the control area, meaning that a larger proportion of respondents in the intervention than control area had been interviewed earlier. This unplanned asymmetry was an outcome of some unforeseen practical imperatives in the field study. However, this fact has had only a minor influence on the age range of the respondents that appears nearly similar in the follow-up survey in the two areas. We do not see any other logic behind how this partial asymmetry in study design would have significantly influenced the results of the logistic regression analysis. The intervention implemented in the study was a community-level intervention and not an individual-level intervention, as the entire villages were targeted by the intervention. This is the main reason why it does not make a difference if the baseline subjects differ from the postintervention subjects. We are interested in population-level changes, and these can be studied by having different subjects pre and postintervention.

Acknowledgments

This study was funded by Kone Foundation's generous support for which the authors present their gratitude. The authors thank two NGOs, U-Respect Foundation, Dr Arundhati Char and Ms Arti Kandala, Mumbai, Vachan and the team, and Nashik for their essential contribution in training and data collection and making the study possible. Special thanks are expressed to the field data collectors and office staff of the abovementioned NGOs that presented their invaluable contribution. The authors thank all the

study participants who spared their time in answering to the surveys. The members of the Ethics Committee are thanked for evaluating the data collection procedures. They also thank Akranumero Ltd for providing a platform for storing the offline survey data.

Authors' Contributions

SK was the main author and contributed to data analysis. MS was the principal investigator and contributed to the literature review. BJ contributed to data management and was responsible for creating a mobile app for the data collection. All authors participated in the designing of the study, content of research instruments, interpretation of results, and the writing process.

Conflicts of Interest

None declared.

References

1. Shet A, Arumugam K, Rodrigues R, Rajagopalan N, Shubha K, Raj T, et al. Designing a mobile phone-based intervention to promote adherence to antiretroviral therapy in South India. AIDS Behav 2010 Jun;14(3):716-720. [doi: 10.1007/s10461-009-9658-3] [Medline: 20054634]
2. World Health Organization. 2011. mHealth: New Horizons for Health Through Mobile Technologies. Based on the Findings of the Second Global Survey on eHealth URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [WebCite Cache ID 71S1Iuult]
3. Mascarenhas A. The Indian Express. 2014. State Gets First-Of-Its-Kind 'Adolescent Helpline' URL: https://indianexpress.com/article/cities/pune/state-gets-first-of-its-kind-adolescent-helpline/ [accessed 2018-08-06] [WebCite Cache ID 71S1VegyV]
4. Ilozumba O, Dieleman M, Kraamwinkel N, van Belle S, Chaudoury M, Broerse JE. 'I am not telling. The mobile is telling': factors influencing the outcomes of a community health worker mhealth intervention in India. PLoS One 2018;13(3):e0194927 [FREE Full text] [doi: 10.1371/journal.pone.0194927] [Medline: 29584773]
5. Reynolds NR, Satyanarayana V, Duggal M, Varghese M, Liberti L, Singh P, et al. MAHILA: a protocol for evaluating a nurse-delivered mhealth intervention for women with HIV and psychosocial risk factors in India. BMC Health Serv Res 2016 Dec 4;16:352 [FREE Full text] [doi: 10.1186/s12913-016-1605-1] [Medline: 27491288]
6. Lerma K, Reyes G, Tiwari S, Tewari A, Hastings C, Blumenthal PD. Acceptability of a text message-based fertility awareness application for family planning in Lucknow, India. Int J Gynaecol Obstet 2018 Jul;142(1):104-107. [doi: 10.1002/ijgo.12488] [Medline: 29574716]
7. Majumdar A, Kar SS, Kumar G, Palanivel C, Misra P. mHealth in the prevention and control of non-communicable diseases in India: current possibilities and the way forward. J Clin Diagn Res 2015 Feb;9(2):LE06-LE10 [FREE Full text] [doi: 10.7860/JCDR/2015/11555.5573] [Medline: 25859473]
8. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26(6):CD011159 [FREE Full text] [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
9. National Family Health Survey. India Fact Sheet 2015-16 URL: http://rchiips.org/NFHS/pdf/NFHS4/India.pdf [accessed 2018-08-06] [WebCite Cache ID 71S2bDQoP]
10. National Health Mission. 2013. A Strategic Approach to Reproductive, Maternal, Newborn, Child and Adolescent Health (RMNCH+A) in India URL: http://nhm.gov.in/images/pdf/programmes/rmncha-strategy.pdf [accessed 2018-08-06] [WebCite Cache ID 71S2rfTEG]
11. de Oliveira IT, Dias JG, Padmadas SS. Dominance of sterilization and alternative choices of contraception in India: an appraisal of the socioeconomic impact. PLoS One 2014;9(1):e86654 [FREE Full text] [doi: 10.1371/journal.pone.0086654] [Medline: 24489759]
12. Raj A, Ghule M, Ritter J, Battala M, Gajanan V, Nair S, et al. Cluster randomized controlled trial evaluation of a gender equity and family planning intervention for married men and couples in rural India. PLoS One 2016;11(5):e0153190 [FREE Full text] [doi: 10.1371/journal.pone.0153190] [Medline: 27167981]
13. Char A, Saavala M, Kulmala T. Assessing young unmarried men's access to reproductive health information and services in rural India. BMC Public Health 2011 Jun 17;11:476-485 [FREE Full text] [doi: 10.1186/1471-2458-11-476] [Medline: 21682899]
14. Char A, Saavala M, Kulmala T. Influence of mothers-in-law on young couples' family planning decisions in rural India. Reprod Health Matters 2010 May;18(35):154-162 [FREE Full text] [doi: 10.1016/S0968-8080(10)35497-8] [Medline: 20541094]
15. Rimal RN, Sripad P, Speizer IS, Calhoun LM. Interpersonal communication as an agent of normative influence: a mixed method study among the urban poor in India. Reprod Health 2015 Aug 12;12:71 [FREE Full text] [doi: 10.1186/s12978-015-0061-4] [Medline: 26265221]

16. Diamond-Smith N, Campbell M, Madan S. Misinformation and fear of side-effects of family planning. Cult Health Sex 2012;14(4):421-433. [doi: 10.1080/13691058.2012.664659] [Medline: 22390371]
17. Samal J, Dehury RK. Family planning practices, programmes and policies in India including implants and injectables with a special focus on Jharkhand, India: a brief review. J Clin Diagn Res 2015 Nov;9(11):LE01-LE04 [FREE Full text] [doi: 10.7860/JCDR/2015/16152.6714] [Medline: 26674943]
18. Calhoun LM, Speizer IS, Rimal R, Sripad P, Chatterjee N, Achyut P, et al. Provider imposed restrictions to clients' access to family planning in urban Uttar Pradesh, India: a mixed methods study. BMC Health Serv Res 2013 Dec 23;13:532 [FREE Full text] [doi: 10.1186/1472-6963-13-532] [Medline: 24365015]
19. Telecom Regulatory Authority of India, Government of India. 2018. Highlights of Telecom Subscription Data as on 31st March, 2018 URL: https://www.trai.gov.in/sites/default/files/PR_No.56_Eng.pdf [accessed 2018-08-06] [WebCite Cache ID 71S3eOC7T]
20. Tenhunen S. A Village Goes Mobile: Telephony, Mediation, and Social Change in Rural India. Oxford: Oxford University Press; 2018.
21. National Family Health Survey. State Fact Sheet: Maharashtra, 2015-16 URL: http://rchiips.org/NFHS/pdf/NFHS4/MH_FactSheet.pdf [accessed 2018-08-06] [WebCite Cache ID 71S767QHF]
22. Data for All. Census Info: India 2011 URL: http://www.dataforall.org/dashboard/censusinfoindia_pca/
23. Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. J Clin Epidemiol 2017 Sep;89:30-42 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.02.016] [Medline: 28351692]
24. R project. The R Project for Statistical Computing URL: https://www.r-project.org/ [WebCite Cache ID 71S6mFfay]
25. U-Respect Foundation. URL: http://www.urespect.org/ [accessed 2018-08-06] [WebCite Cache ID 71S6gyYOE]
26. Akra-Numero. 2018. mHealth Study: Contraception and Mobile Reproductive Health Services in India URL: https://akranumero.fi/publ/ [accessed 2018-08-05] [WebCite Cache ID 71S6qGJZL]
27. Mwaikambo L, Speizer IS, Schurmann A, Morgan G, Fikree F. What works in family planning interventions: a systematic review. Stud Fam Plann 2011 Jun;42(2):67-82 [FREE Full text] [doi: 10.2307/41310713] [Medline: 21834409]
28. Nahar P, Kannuri N, Mikkilineni S, Murthy G, Phillimore P. mHealth and the management of chronic conditions in rural areas: a note of caution from southern India. Anthropol Med 2017 Apr;24(1):1-16 [FREE Full text] [doi: 10.1080/13648470.2016.1263824] [Medline: 28292206]
29. Tapare VS, Parande MA, Borle PS. Unmet need for contraception among married women of reproductive age in rural Maharashtra. Int J Community Med Public Health 2017 Aug 23;4(9):3365. [doi: 10.18203/2394-6040.ijcmph20173846]

Abbreviations

IUCD: intrauterine contraceptive device
mHealth: mobile health
NGO: nongovernmental organization
PHC: primary health center
SRH: sexual and reproductive health

Edited by G Eysenbach; submitted 01.11.18; peer-reviewed by O Ilozumba, S Jejeebhoy; comments to author 02.03.19; revised version received 24.04.19; accepted 05.07.19; published 09.08.19.

Please cite as:
Kulathinal S, Joseph B, Säävälä M
Mobile Helpline and Reversible Contraception: Lessons From a Controlled Before-and-After Study in Rural India
JMIR Mhealth Uhealth 2019;7(8):e12672
URL: https://mhealth.jmir.org/2019/8/e12672
doi:10.2196/12672
PMID:31400103

Original Paper

A Novel Mobile Health Tool for Home-Based Identification of Neonatal Illness in Uganda: Formative Usability Study

Madison Vanosdoll[1*], BSE, MSE; Natalie Ng[1*], BS, MSE; Anthony Ho[1*], BS, MSE; Allison Wallingford[1], BS, MSE; Shicheng Xu[1], BEng, MSE; Shababa Binte Matin[1], ScB; Neha Verma[1], MS; Azadeh Farzin[2,3], MD; W Christopher Golden[2], MD; Youseph Yazdi[1,2,4], MBA, PhD; Peter Waiswa[5], MBChB, MPH, PhD; Alain Labrique[6], MHS, PhD; Soumyadipta Acharya[1], MBBS, MSE, PhD

[1]Center for Bioengineering Innovation and Design, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States
[2]School of Medicine, Johns Hopkins University, Baltimore, MD, United States
[3]Pediatrix Medical Group, Rockville, MD, United States
[4]Global mHealth Initiative, Johns Hopkins University, Baltimore, MD, United States
[5]School of Public Health, Makerere University, Kampala, Uganda
[6]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
[*]these authors contributed equally

Corresponding Author:
Soumyadipta Acharya, MBBS, MSE, PhD
Center for Bioengineering Innovation and Design
Department of Biomedical Engineering
Johns Hopkins University
3400 N Charles Street
Baltimore, MD, 21218
United States
Phone: 1 4105167074
Email: acharya@jhu.edu

Abstract

Background: While early identification of neonatal illness can impact neonatal mortality rates and reduce the burden of treatment, identifying subtle clinical signs and symptoms of possible severe illness is especially challenging in neonates. The World Health Organization and the United Nations Children's Fund developed the Integrated Management of Neonatal Childhood Illness guidelines, an evidence-based tool highlighting seven danger signs to assess neonatal health. Currently, many mothers in low-resource settings rely on home visits from community health workers (CHWs) to determine if their baby is sick. However, CHWs visit infrequently, and illness is often detected too late to impact survival. Thus, delays in illness identification pose a significant barrier to providing expedient and effective care. Neonatal Monitoring (NeMo), a novel neonatal assessment tool, seeks to increase the frequency of neonatal screening by task-shifting identification of neonatal danger signs from CHWs to mothers.

Objective: This study aimed to explore the usability and acceptability of the NeMo system among target users and volunteer CHWs by assessing ease of use and learnability.

Methods: Simulated device use and semistructured interviews were conducted with 32 women in the Iganga-Mayuge districts in eastern Uganda to evaluate the usability of the NeMo system, which involves a smartphone app paired with a low cost, wearable band to aid in identification of neonatal illness. Two versions of the app were evaluated using a mixed methods approach, and version II of the app contained modifications based on observations of the first cohort's use of the system. During the posed scenario simulations, participants were offered limited guidance from the study team in order to probe the intuitiveness of the NeMo system. The ability to complete a set of tasks with the system was tested and recorded for each participant and closed- and open-ended questions were used to elicit user feedback. Additionally, focus groups with 12 CHWs were conducted to lend additional context and insight to the usability and feasibility assessment.

Results: A total of 13/22 subjects (59%) using app version I and 9/10 subjects (90%) using app version II were able to use the phone and app with no difficulty, despite varying levels of smartphone experience. Following modifications to the app's audio instructions in version II, participants' ability to accurately answer qualitative questions concerning neonatal danger signs improved

by at least 200% for each qualitative danger sign. All participants agreed they would trust and use the NeMo system to assess the health of their babies. Furthermore, CHWs emphasized the importance of community sensitization towards the system to encourage its adoption and regular use, as well as the decision to seek care based on its recommendations.

Conclusions: The NeMo system is an intuitive platform for neonatal assessment in a home setting and was found to be acceptable to women in rural Uganda.

(JMIR Mhealth Uhealth 2019;7(8):e14540) doi:10.2196/14540

KEYWORDS

neonatal; community health workers; maternal; Uganda; World Health Organization; smartphone app; digital health; mobile health; telemedicine

Introduction

Background

Each year, 2.5 million newborns die in the first 28 days of life, with 80% of these deaths occurring within the first seven days of age [1]. In resource-limited settings in low- and middle-income countries (LMICs), a majority of these deaths occur within homes largely due to preventable causes such as pneumonia, sepsis, and other illnesses [2,3]. The neonatal mortality rate in Uganda is 26 per 1000 live births [4]. In rural districts in eastern Uganda where 34 in every 1000 newborns die in their first month of life, neonates are even more vulnerable [3]. Many of these deaths could be averted by timely identification and referral to treatment [5,6].

While as many as 84% of neonatal infection-related deaths could be prevented with available interventions, identifying subtle clinical signs and symptoms of severe illness is especially challenging in neonates [5]. Thus, delays in illness identification pose a significant barrier to providing expedient care [6-8]. An analysis of 64 neonatal deaths in eastern Uganda showed that the highest contributing delays to newborn intervention were challenges in problem recognition and in the decision to seek care (50%), with a median time of 3 days from onset of illness to seeking care outside the home [3].

In order to improve recognition of neonatal illness, the World Health Organization (WHO) has developed the Integrated Management of Neonatal and Childhood Illness (IMNCI) guidelines [9-11]. Studies have shown the following seven danger signs predict severe illness in neonates: (1) difficulty feeding; (2) convulsions; (3) lethargy; (4) chest indrawing; (5) respiratory rate of 60 breaths per minute or more; (6) temperature above 37.5°C; or (7) temperature below 35.5°C [9]. Presentation of a danger sign indicates the need for immediate medical attention [9,12]. Although effective identification of these signs at the community level can intercept illness and incite care-seeking behavior capable of impacting child mortality, the tools and training needed to assess indicators of illness are lacking in low-income settings [6,13-17].

Currently, rural health care systems in Uganda and other LMICs rely on village health teams (VHTs) of volunteer community health workers (CHWs) to visit mothers at home to provide postnatal care. While women visited by CHWs are 87% less likely to lose their newborns, these volunteers are often overburdened, and limited by both availability and bandwidth [18]. Thus, CHWs are often unable to complete the three postnatal home visits recommended by the WHO during the critical first week of life. As few as 5% of Ugandan newborns receive a CHW visit in the first 48 hours of life and sick infants are often identified too late to impact survival [19]. Task-shifting neonatal assessment from CHWs to mothers is a promising strategy to reduce neonatal mortality [6,7,20]. While the aforementioned danger signs can have a sensitivity of 85% and specificity of 75% when used by a trained primary care worker to identify neonatal illness, unassisted maternal recognition of these symptoms in community settings has failed to exceed a sensitivity of 20-24% [7,9,21,22].

The NeMo System: A Low-Cost Technology Platform for Assessment of Neonatal Danger Signs

The opportunity to improve identification of neonatal illness in home settings was identified by a team of faculty and biomedical engineering graduate students at the Johns Hopkins University Center for Bioengineering Innovation and Design (CBID) and School of Public Health, based on over 18 years of community research in rural Bangladesh [23-25]. Engineering design teams had the opportunity to confirm these population research findings through community and rural health system immersion in Iganga, Uganda. Primary observations from ethnographic research conducted in Iganga highlighted delays in identification of neonatal illness due to a systemic reliance on overburdened CHWs. Teams also visited rural communities in Kenya and Bangladesh and observed similar patterns. The design teams identified the need for a low-cost technology to regularly, objectively, and accurately assess the seven previously mentioned IMNCI danger signs during the first week of life to expedite detection of neonatal illness in LMICs. By empowering care providers within the family unit (ie, mothers) to detect signs of illness, the team hypothesized that timely care-seeking behavior could be triggered when it otherwise might not be. Neonatal Monitoring (NeMo), a proposed solution to the need for earlier identification of neonatal illness, is a two-part system designed to empower mothers to effectively assess the seven validated IMNCI danger signs.

The NeMo system consists of a novel, wearable sensing band (the NeMo band) and a low-cost smartphone preloaded with the custom NeMo app (Figure 1). The NeMo band fastens around the neonate's abdomen using a generic hook-and-loop fastener like Velcro (herein after referred to as velcro). It is equipped with sensors to measure temperature and respiratory rate, which are difficult to assess without appropriate training or tools. These sensors are housed in a small plastic enclosure

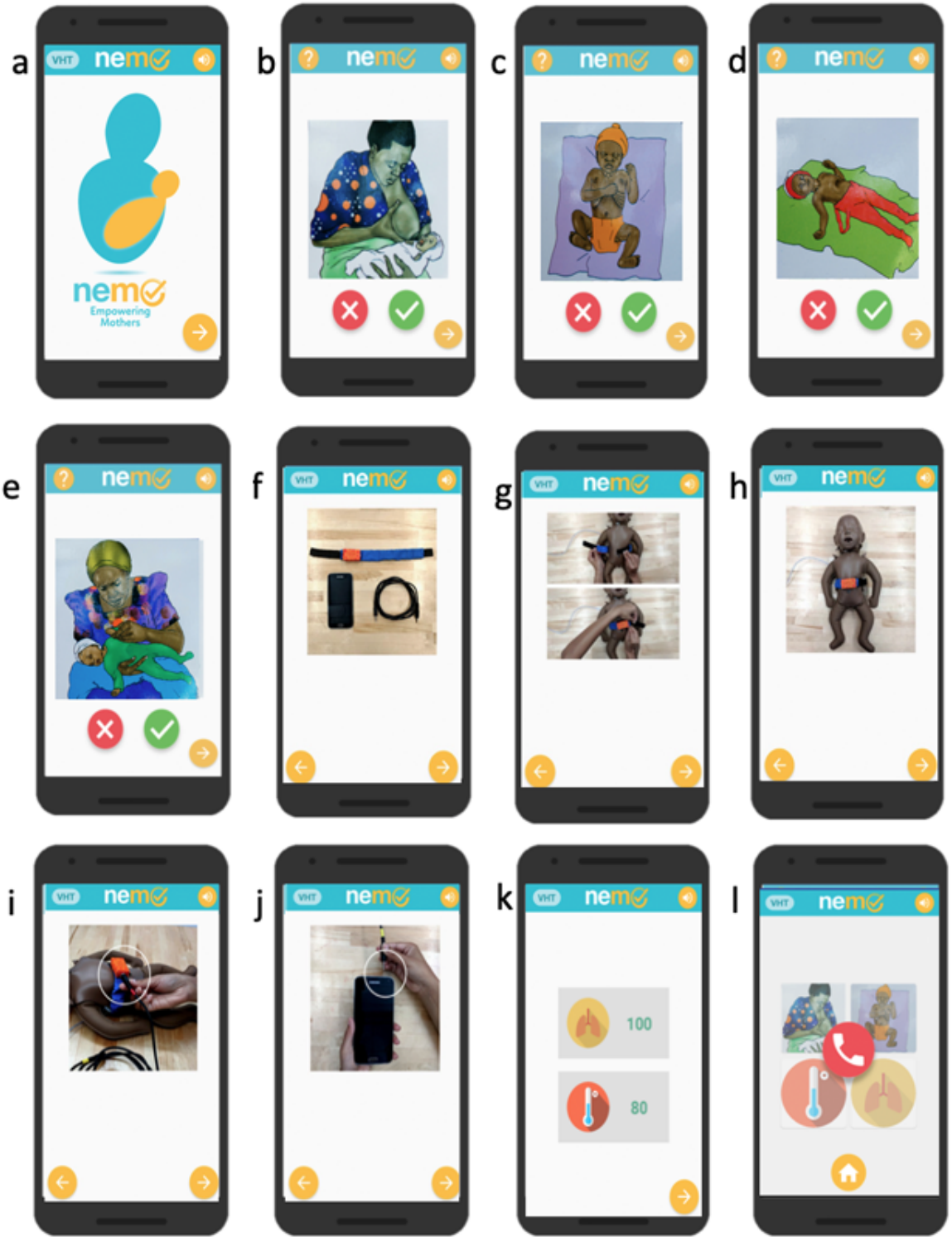
which must be placed centrally on the neonate's abdomen (in between the nipples and midway between the umbilical cord and the nipples) (Figure 1). Information from the respiratory and temperature sensors is transmitted to the phone for processing via a standard audio cable which plugs into the audio jack on the phone and via an identical jack embedded into the enclosure on the NeMo band. The app uses audio and visual cues in the local language to enable any mother, regardless of literacy, to assess the qualitative danger signs and place the NeMo band (Figure 2). By responding to a series of interactive audio prompts concerning her newborn's health, the mother is asked to indicate whether or not her baby has any danger signs by selecting either a check or X on the app interface (ie, "Has your baby been refusing to breastfeed? Press check if yes, press X if no."). The app does not require internet connectivity. By synthesizing the results from the quantitative and qualitative assessments, NeMo alerts mothers of any danger signs detected and advises users if the newborn requires medical attention.

When using the system, the mother is first taken to a home screen (Figure 2a), which introduces the navigational features of the app. The mother is then guided through individual screens which ask her to assess the presence of the four qualitative danger signs: difficulty breastfeeding (Figure 2b), chest indrawing (Figure 2c), convulsions (Figure 2d), and lethargy (Figure 2e). The app then uses audio instructions and images to walk the mother through each step of properly placing and securing the band around her baby and connecting the band to the phone with the audio cord (Figure 2f-j). Once positioned around the neonate's abdomen, the NeMo band measures temperature and respiratory rate (Figure 2k). Following the quantitative assessment, the app displays representative images of any danger signs detected and, if indicated, produces a red phone icon to prompt the mother to call a CHW for further assessment and help (Figure 2l). The mother only needs to press the icon to initiate the phone call.

NeMo is based on a CHW-coordinated, phone-sharing framework. In this model, mothers are provided a smartphone by a CHW for the week following birth and purchase their own low cost, wearable band. After the first seven days of the newborn's life, the phone will be returned to the CHW and the band discarded to prevent the spread of infection between newborns. By making both this technology and education to support clinical evaluation accessible in community settings, NeMo could empower mothers to assess neonatal illness at home and without support. While evidence supporting the acceptability of phone use among CHWs is abundant, evaluation of a community phone-sharing model and maternal smartphone use to enable neonatal mobile health (mHealth) interventions remains lacking [26-28]. Therefore, for any digital health intervention targeting mothers in LMICs directly, there is a need to validate acceptability and usability among target users.

Figure 1. The NeMo (Neonatal Monitoring) band goes around the neonate's abdomen to acquire respiratory rate and temperature data. The data is then transmitted via an audio cable and processed in the NeMo app, where it is then integrated with mothers' responses regarding qualitative danger signs to provide a recommendation on whether a mother should seek care for her neonate.

Figure 2. Illustration of the NeMo app, with qualitative danger sign assessment screens: a) home screen; b) difficulty breastfeeding; c) chest indrawing; d) convulsions; and e) lethargy. The mother can select the X or check symbols to respond as to whether her child is sick or healthy. Instructions on how to place the band are shown in f-j. Respiratory rate and temperature are displayed in k). Danger signs detected, and need to call a CHW are shown in l. NeMo: Neonatal Monitoring; CHW: community health worker.



Objectives

While several studies have established the utility of the seven danger signs in identifying neonates in need of care, the ability of mothers to utilize a simple, digital health strategy to guide effective assessment of these signs has not been established [9,12]. Thus, this study was designed to evaluate the usability and acceptability of the NeMo smartphone app and wearable sensor among both women of child-bearing age and CHWs in the Iganga-Mayuge districts of Uganda. Further, this study seeks to evaluate women's and CHWs' perceptions of the phone-sharing model and their willingness to use the NeMo device to evaluate the health of their neonates.

Methods

Overview

This study was undertaken in eastern Uganda in the Iganga-Mayuge districts, over the course of two weeks, in five villages. These districts were selected for their representation of rural Uganda, offering socioeconomic diversity and corresponding literacy rates [3,29,30]. Fieldwork was supported by Uganda Development and Health Associates (UDHA), which has a regional headquarters based in Iganga. Villages in these districts were visited to capture perceptions of the usability and acceptability of the NeMo system within populations at high risk for vulnerability [29,31]. For both study participants and CHWs, a convenience sampling strategy was utilized. Village leaders randomly selected female citizens from their village within reproductive age (18-49 years old) to be interviewed. CHWs were randomly selected via their connection with UDHA.

This study was conducted following protocols approved by the Johns Hopkins Homewood Institutional Review Board (IRB00005144) and by the Institutional Review Board of Makerere University School of Public Health (MakSPH IRB) in Uganda. Written consent was obtained from all parties including study subjects and CHWs before each interview and was available in Lusoga and English.

Simulated Use of the NeMo System

The qualitative study consisted of semistructured interviews with all women (n=32) and volunteer CHWs (n=12). With each subject, a preliminary structured interview was conducted to collect demographic information. Subsequently, each subject participated in a verbally posed simulated scenario to assess the usability of the NeMo system. Subjects utilized the NeMo app preloaded onto an Android smartphone, along with a lookalike prototype of the NeMo device. The steps of the simulation were carried out on a NeoNatalie inflatable simulator (Laerdal Medical, Norway). A comprehensive list of the simulated tasks can be found in Table 1.

The usability of the app interface was measured based on the subjects' ability to navigate through the app and accurately answer the qualitative questions based on the posed scenarios. Additionally, the ability of each participant to position the device on the NeoNatalie's chest and correctly insert the audio cord into both audio jacks was used as a measure of usability. The usability assessment was further rounded out by a series of closed- and open-ended questions, followed by a series of questions answered using a Likert scale.

Subjects were shown two alternative versions of the band and two alternative versions of the app. They were then asked to indicate their preferences based on their perceptions of ease of use, helpfulness, and safety.

This study was conducted in two phases, reflecting version I and version II of the app interface. The first phase (n=22) followed the methods detailed above and used app version I. The second phase (n=10) utilized version II of the app, which was updated based on qualitative and quantitative results from phase one. All study methods remained consistent between the two phases of the study. Women enrolled in the study were only given either version I or version II of the app.

Statistical significance was determined using a Fisher's Exact Test to determine subjects' association with a green check and red X and between mothers utilizing each version of the app. A Likert scale was used to gauge subjects' opinions on the NeMo system and converted to a composite score (1=Strongly Disagree, 5=Strongly Agree). All other responses were analyzed using descriptive statistics.

Table 1. Outline of study tasks and questions asked regarding the subjects' experience using the NeMo system.

Task #	Type	Description
1	Consent	Study team members obtained written consent from subject.
2	Assessment of symbol comprehension	Before the subject was introduced to the phone app, she was shown printed images of a green check and a red X and asked to point to the image that meant “yes” or “no,” respectively.
3	Ability to use smartphone	The subject was shown how to unlock the phone and open the app, but not how to navigate through the NeMo[a] app. She was then asked to unlock the smartphone device, open the app, and begin clicking through it, following the audio prompts. If the subject appeared to be stuck on a screen, she was prompted to replay the audio cue rather than guided on how to proceed, and her hesitation was noted by the study team. Specific tasks assessed included: • unlocking the phone • opening the app • navigating through the app from homescreen
4	Answering questions regarding qualitative danger signs	Before the subject answered each of the app’s questions regarding qualitative danger signs, the interviewer stated whether the NeoNatalie was hypothetically afflicted with the relevant condition. For example, the interviewer would state, “The baby is convulsing,” or, “The baby is healthy; she is not convulsing,” in reference to the NeoNatalie. The subject was then asked to respond to the app’s questions based on the posed scenario given by the interviewer.
5	Device placement & audio cord insertion	The subject’s ability to correctly insert the audio cord into the device and the smartphone audio jack, as well as her ability to place the band in the center of the simulation mannequin’s chest when guided only by the app’s audio and visual cues was evaluated. Specific tasks assessed included: • device placement on the simulation mannequin • successful connection of audio cord into the device • successful connection of audio cord into the phone
6	User feedback	The subject was asked a series of closed- and open-ended questions, followed by a series of questions answered using a Likert scale to probe her perceptions of the NeMo system. Questions covered the following themes: • Would you use this device on your baby? • Were there any danger signs mentioned on the phone that you did not understand? • How much, if anything, would you be willing to pay for the device?
7	Intent to act	To qualitatively assess subjects’ intent to act on NeMo’s recommendation on whether or not to seek care, the subject was asked to respond to two hypothetical questions: • If the device says your baby is sick, but you think your baby is healthy, what would you do? • If the device says that your baby is healthy, but you believe your baby is sick, what would you do?
8	Alternative band embodiments	The subject was shown two additional versions of the NeMo band with different fastening mechanisms and asked to practice placing them on the NeoNatalie. She was then asked to answer open-ended questions concerning which embodiment she felt was safest, easiest to use, and which one she would be most likely to use on her own baby.
9	Alternative app embodiments	The subject was shown two additional versions of the NeMo app and asked to answer similar questions pertaining to ease of use and helpfulness for each version.

[a]NeMo: Neonatal Monitoring

Community Health Worker Focus Groups

Three focus groups, facilitated by 2-3 study team members and a translator, were conducted with groups of 3-5 CHWs. A total of 12 CHWs participated in focus group discussions. Each focus group lasted approximately 60 minutes, with question themes including mothers’ ability to use the system, the feasibility of a CHW-lead training and sensitization initiative, and the acceptability of a community phone-sharing model. Table 2 details key topics discussed with CHWs.

Five English-speaking study team members were trained to conduct the interviews and were provided interview guides. The majority of the interviews were conducted in Lusoga through the aid of two translators, however, some study participants and CHWs were able to understand and respond to the interview questions in English. All interviews were video recorded with participant consent.

Table 2. Outline of themes explored in focus group discussions with community health workers (CHWs).

Theme	Examples
1 CHW background	• Personal experience, training, responsibilities • Current practices for neonatal assessment
2 Training mothers	• Current practices providing antenatal care • Danger sign recognition training for mothers
3 Acceptability of technology	• Smartphone experience • Comfort level training mothers on smartphone
4 Business model validation	• Perception of phone sharing model • Willingness to coordinate band sales and phone sharing

Results

Study Population Demographics

The study cohort consisted of 32 women from 5 different villages across the Iganga-Mayuge districts. Women were screened to be of child-bearing age or older, so they ranged from 19 to 44 years of age, with an average age of 27. They had each given birth to a range of 0-6 children, averaging 2.6 children.

Of all the subjects, 30/32 (94%) either owned or had access to a cellphone. However, the majority of these were not smartphones, with only 10/32 subjects (32%) having previously used a smartphone. In acknowledgement of the importance of neonatal danger sign recognition, the Ugandan government has mandated that danger signs be taught in antenatal care education. All subjects enrolled in phase two of the study received antenatal and postnatal education. Two of the women interviewed in phase two of the study did not have children and were not included in this survey analysis. Despite the expectation of proficiency in identifying signs of illness following this training, subjects' retention was low: only 3 of the 8 mothers (37.5%) recognized fever as a danger sign. Furthermore, only 3 mothers were able to correctly identify one additional danger sign (Table 3). These results emphasize the gap in mothers' education and knowledge surrounding the assessment of neonatal illness and highlight the potential value of an effective tool to guide identification of danger signs in the home setting.

Table 3. Mothers' recall of neonatal danger signs in study phase two (n=8).

Danger signs	Mothers, %
Fever	37.5
Chest indrawing	12.5
Increased respiratory rate	12.5
Lethargy	12.5
Hypothermia	0
Refusing to breastfeed	0
Convulsions	0

Assessment of Symbol Comprehension

Of all the subjects, 30/32 (94%; $P<.001$) correctly identified the green check to mean yes and the red X to mean no, while two women were unfamiliar with this symbology. These two women thus had the meaning of each symbol explained to them.

Ability to Use Smartphone

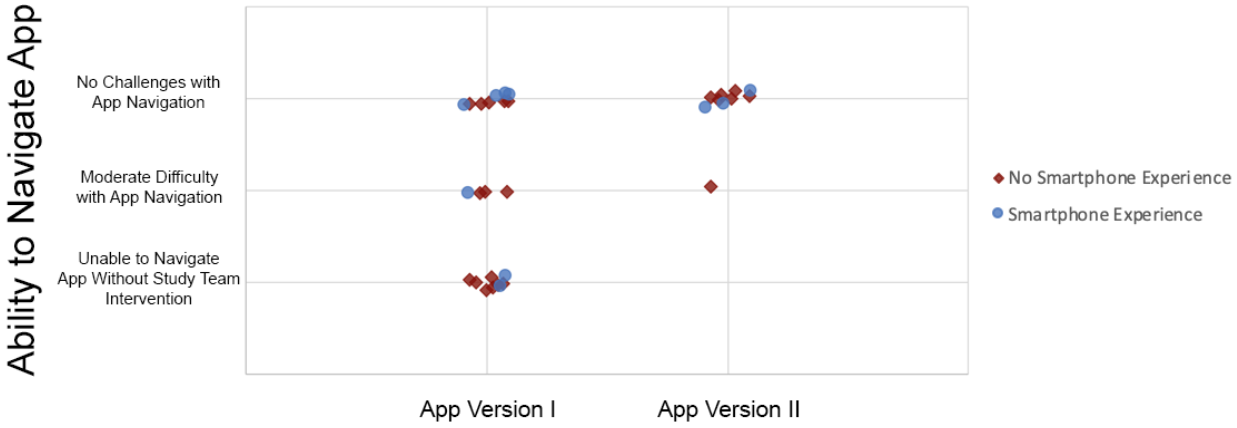
Of the 22 subjects that had never used a smartphone before, 18 (82%) could unlock the phone and open the app without additional guidance and 20 (91%) could do at least one or the other without additional guidance.

App Version I

App Navigation

In total, 13/22 subjects (59%) enrolled in phase one required no external prompting or direction from the interviewer or translator to navigate from the homescreen to begin the assessment (Figure 3). A lack of specificity in the verbal cues indicating what button should be clicked to proceed often prevented subjects from navigating through the app.

Figure 3. Comparison of mothers' ability to navigate version I and version II of the app based on previous smartphone experience. Mothers using version II of the app showed a significant improvement in app navigation (*P*=.02).

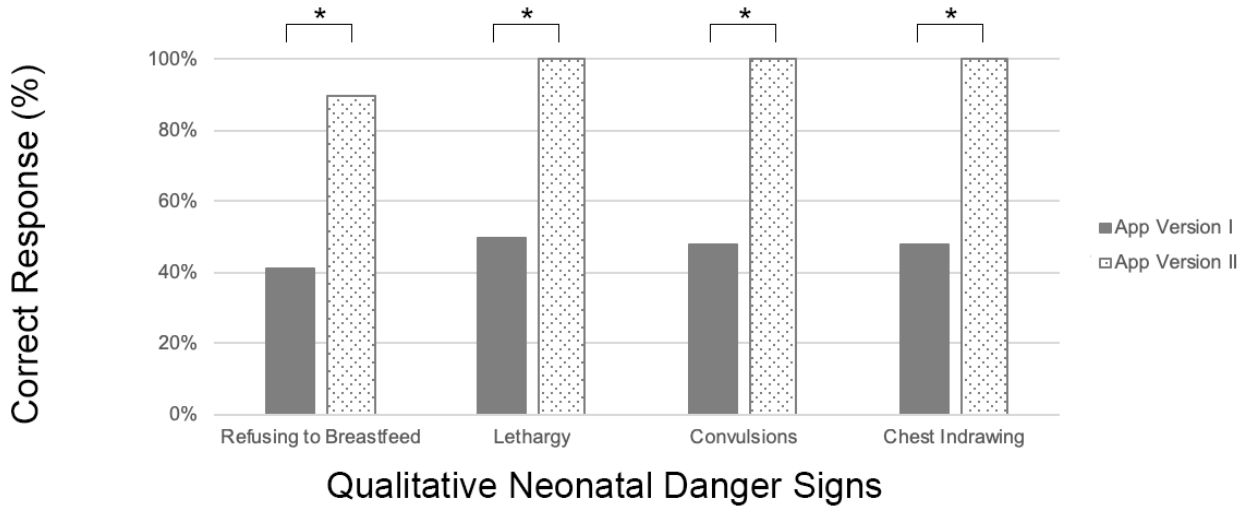


Answering Questions Regarding Qualitative Danger Signs

Across both versions of the app, all subjects responded that there were no danger signs introduced by the app they had not understood. Although this was the first time two subjects had heard about chest indrawing, 28/32 women (88%) responded that they were confident or very confident in their ability to identify those danger signs in their own infants. However, this study only assessed a subject's ability to utilize the NeMo system to input information regarding their newborn's health. Determining their accuracy in identifying qualitative danger signs is beyond the scope of this study. All 32 subjects responded that the voice directions in the app were intuitive and easy to understand.

In version I of the app, only one subject was able to correctly answer all 4 of the qualitative questions based on the posed scenarios and only 6/22 subjects (27%) were able to correctly answer 3 of the 4 questions (Figure 4).

Figure 4. Comparison of mothers' ability to correctly respond to verbal scenarios using app versions I and II. Their ability to answer qualitative questions significantly improved across all four danger signs. * indicates *P*<.05.



Device Placement & Audio Cord Insertion

As correct placement of the NeMo device on the abdomen of the neonate is critical to obtaining accurate measurements for both respiratory rate and temperature, a subject's ability to properly position the device around the NeoNatalie simulator using only the verbal instructions from the app was used as a measure of ease of use of the device. Using version I of the app, 18/22 subjects (82%) were able to properly position the device with only the verbal instructions and no prior instruction or training. Additionally, 20/22 subjects (91%) were able to successfully connect the audio cable to both the device and smartphone.

Changes made Between App Version I and Version II

Subjects stuck on the homescreen in version I often attempted to press different areas on the screen (including the check mark in the NeMo logo) in an attempt to advance to the next screen. To avoid this challenge, in version II the team added more detailed audio instructions explaining the location of the button to continue to the next screen.

As a result of the poor ability of subjects to answer qualitative questions based on the simulated scenarios using version I of the app, the team evaluated what factors might be at fault. One translator suggested that subjects might be responding in reference to their own infants rather than to the posed NeoNatalie scenarios. The team also hypothesized that questions were being posed in a misleading way based on the positive and negative connotations of the check and X symbols. Originally, the app instructed subjects to press the green check if the baby had the danger sign ("Yes, my baby has this danger sign") and the red X if the baby did not have the danger sign ("No, my baby does not have this danger sign"). However, it was suggested that subjects might not intuitively associate a green check with illness and a red X with a healthy baby. Thus, in app version II, the team rephrased the app audio, instructing subjects to press the green check if their baby was healthy and did not have the sign and the red X if their baby was sick and did have the sign.

Finally, while most women using version I of the app were able to correctly place the device, the NeMo app's audio cues were updated to further assist subjects with device placement. Greater detail was given about anatomical structures on the newborn, specifying the position of the device relative to the nipples and umbilical cord in version II.

App Version II

App Navigation

All subjects utilizing version II of the app, regardless of previous smartphone exposure, were able to turn on the smartphone and successfully navigate through the app without intervention by the study team. Figure 3 illustrates a comparison between the different app versions and subjects' exposure to smartphones.

For both versions of the app, subjects were assessed on ease of app navigation based on observed hesitation and need for further explanation and prompting by the study team. The introduction of version II resulted in significant improvement in subjects' ability to navigate the app interface, as assessed by comparing the number of women that were able to navigate the app without any challenges to those who could not (P=.02).

Answering Questions Regarding Qualitative Danger Signs

Using version II of the app, 9/10 (90%) of the remaining subjects correctly answered all four of the qualitative questions based on the posed scenarios. The remaining subject answered three of the four questions correctly. This modification resulted in significant improvement in subjects' ability to respond to each of the qualitative danger sign questions using the NeMo app interface: breastfeeding (P=.02), lethargy (P=.006), convulsions (P=.005), and chest indrawing (P=.005). A comparison between subjects' performance using version I and version II of the app is shown in Figure 4.

Placement and Audio Cord Insertion Comparison

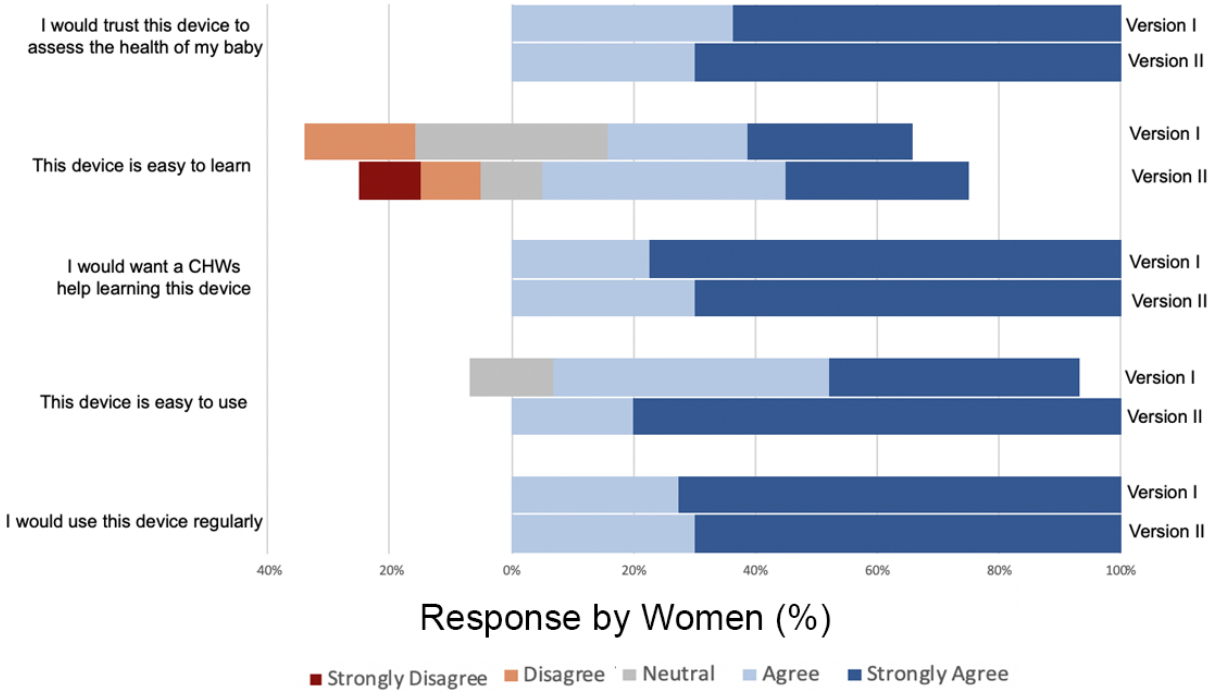
Using improved audio instruction in app version II, all subjects were able to correctly position the band on the NeoNatalie. While these results showed improvement from the 18/22 subjects (82%; P=.28) that were able to place the band correctly and the 20/22 subjects (91%; P>.99) that were able to connect the audio cable properly, these changes were not found to be statistically significant.

Perception of Usability and Acceptability

Ease of Use and Learnability

Each subject was asked about her perception of the system's ease of use, and they responded using a Likert scale. A comparison between responses from women enrolled in phase one and phase two was carried out, but their responses were not found to be significantly different. Responses provided by subjects in both phases of the study are displayed in Figure 5. Using a composite scoring system (Strongly Agree=5, Strongly Disagree=1), the device had an ease of use score of 4.34 (median 5, interquartile range [IQR] 1) and a learnability score of 3.56 (median 4, IQR 2). In addition, all subjects surveyed agreed or strongly agreed they would like a CHW's help learning how to use the device, giving a composite score of 4.75 (median 5, IQR 1).

Figure 5. Mothers' responses to statements regarding perception of the NeMo device scored on a Likert scale. CHW: community health worker.



Trust in Technology and Intent to Use

All subjects surveyed indicated they agree or strongly agree they would trust NeMo to assess the health of their baby, giving a composite score of 4.66 (median 5, IQR 1). Additionally, all subjects agree or strongly agree that they would use NeMo regularly, giving a composite score of 4.72 (median 5, IQR 1).

Intent to Act

A total of 29/32 subjects (91%) stated they would take their baby to a health care facility if the device told them their baby was sick, even if they believed the baby was healthy. Alternatively, 24/32 subjects (75%) responded that they would take their baby to a health care facility if they believed their baby was sick, even if the device told them their baby was healthy. These findings support subjects' trust in the device and the potential for NeMo to trigger care-seeking behavior, while also demonstrating women's agency in seeking medical attention even when NeMo does not signal illness.

Perception of Band Embodiments

In addition to evaluating usability and acceptability of the current versions of the NeMo system and band embodiment, the team developed two additional bands to evaluate which method of fastening was most acceptable to study subjects. For each of the three band variations shown in Figure 6, subjects placed the band on the NeoNatalie simulator and then evaluated which band they liked best, found easiest to use, were most likely to use, and thought would be safest. The velcro band utilizes a press-seal–fixation mechanism, while the velcro-with-loop uses similar velcro fixation with the addition of a loop to aid adjustments of tightness. The adjustable strap is slid over the newborn's lower half and utilizes a friction fastening mechanism that slides along the strap to tighten. Results reported by the cohort are displayed in Table 4. A majority of 22/32 subjects (69%) found the velcro-with-loop band to be their favorite. Similarly, 17/32 subjects (53%) considered the velcro-with-loop easiest to use, while 14/32 subjects (44%) preferred the velcro band.

Figure 6. NeMo band embodiments: (A) velcro; (B) velcro-with-loop; (C) adjustable strap.

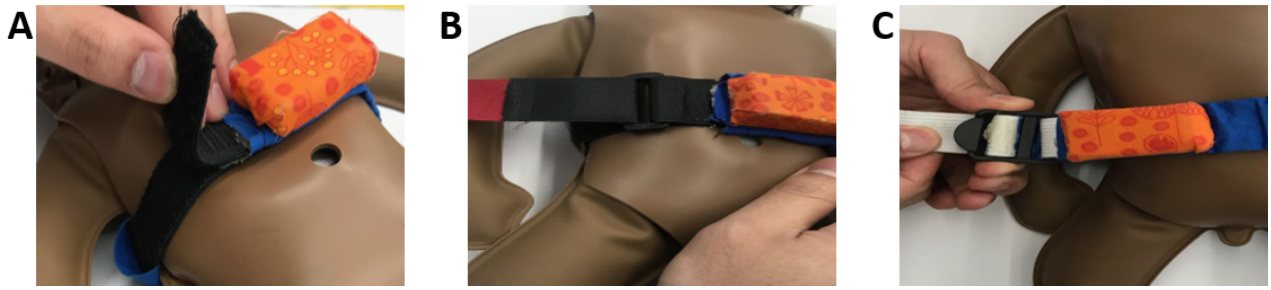


Table 4. Subjects' perceptions of different band embodiments (n=32). All values are given as n (%).

Embodiment	Velcro, n (%)	Velcro-with-loop, n (%)	Adjustable strap, n (%)
Liked best	9 (28)	22 (69)	1 (3)
Easiest to use	14 (44)	17 (53)	1 (3)
Safest	11 (34)	13 (41)	8 (24)

Perception of App Embodiments

Subjects were also shown three different variations of the phone app and asked to rank which was the easiest to use, which was most helpful, and which one they liked the best (Figure 7). The distribution of their preferences among the app interfaces is displayed in Table 5. The new interfaces developed alongside the users were favored. Subjects indicated the animated graphic interchange format (GIF) app was the easiest, while the clickable pictures interface was the most liked and most helpful variation. 31 subjects responded to which app was easiest to use, and 30 subjects responded to which app was liked best and most helpful.

Figure 7. NeMo app embodiments: (A) illustrations; (B) animated GIF; (C) clickable pictures.

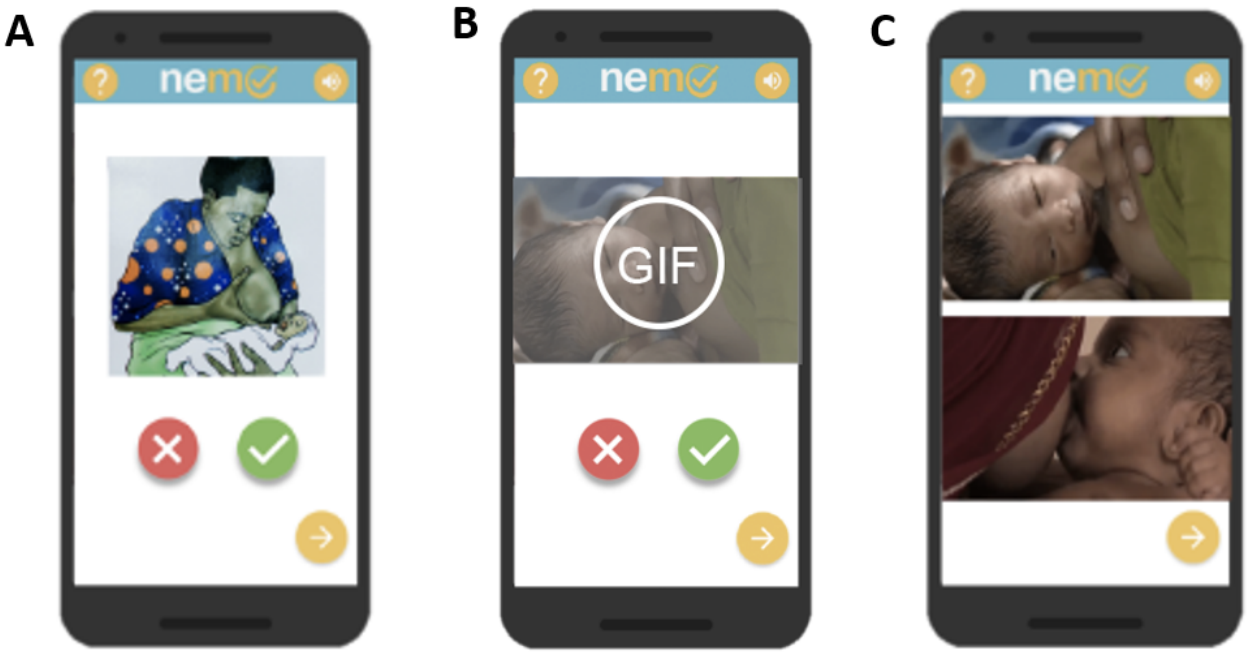


Table 5. Subjects' perception of different app embodiments. All values are given as n (%).

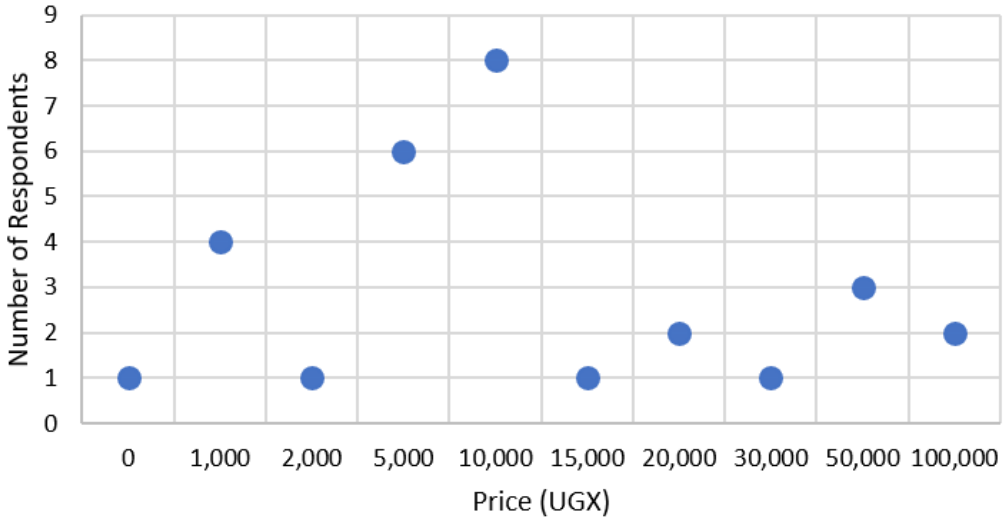
App interface	Illustrations, n (%)	Animated GIF[a], n (%)	Clickable pictures, n (%)
Easiest	8 (26)	12 (39)	11 (35)
Liked best	8 (27)	10 (33)	12 (40)
Most helpful	4 (13)	15 (50)	11 (37)

[a]GIF: graphic interchange format.

Amount Women are Willing to Pay for NeMo Device

In addition to assessing the usability of the band and the phone app, subjects were asked questions regarding the amount they would be willing to pay for the device. A total of 30/32 subjects (94%) were willing to make a one-time payment to purchase a NeMo band usable for the seven days following birth. The amount they were willing to pay varied between 1,000 to 100,000 Ugandan shillings (UGX), or approximately $0.27 to $27.00 United States dollars (USD). The median price was 10,000 UGX, or approximately $2.70, and the mean was 19,000 UGX, or $5.14. Figure 8 illustrates the distribution of prices subjects indicated they would pay for the NeMo band. While most subjects indicated they would be willing to pay for the device, participants were less receptive to the prospect of making a security deposit to borrow the smartphone. In total, 20/27 subjects (74%) indicated they would be comfortable making a security deposit they could collect after the device was returned.

Figure 8. Distribution of mothers' willingness to pay for NeMo device.



Insights from Community Health Worker Interviews

Three focus groups conducted with a total of twelve CHWs highlighted critical logistical considerations for the successful implementation of the NeMo system in Uganda. Through these conversations, it was confirmed that CHWs are expected to visit mothers three times during the first week of life. Additionally, CHW assessments of critical quantitative danger signs, including respiratory rate and temperature, are most often only qualitative in nature and are conducted by visually observing the baby's breathing as well as touching the baby to determine temperature. Regarding environmental factors, CHWs were optimistic about mothers' abilities to charge the phones, especially considering that most families own at least one phone and solar power is becoming increasingly more available in village homes.

Community Health Workers' Perceptions of Mothers' Ability to Use the NeMo System

CHWs were generally supportive of the NeMo system and believed that the device could help reduce neonatal mortality. However, a few participants expressed concern that some mothers might either lack the education to identify danger signs or opt to visit traditional birth assistants (TBAs) rather than a CHW or health facility. Despite these doubts, all twelve CHWs agreed they would be eager and willing to lead training programs in which mothers would be instructed in the use of the NeMo system.

Community Sensitization

The CHWs illustrated the importance of sensitizing the community to the utility and value of the device in order to ensure regular use and to encourage familial agreement to purchase the device. This community sensitization will also be paramount in determining whether or not mothers act on the suggestions of the NeMo system regarding their decision to seek care. Furthermore, it may dictate the safety of the phones in the community, reducing the likelihood of other community members attempting to steal them.

Payment for the System

After expressing concern about requiring mothers to pay for the sensing bands, numerous CHWs questioned whether the band could be used on multiple babies or for a duration longer than a week in order to give mothers greater incentive to purchase the device. A number of senior CHWs spoke of a poverty grading tool they utilized to determine the price point for a voucher that would cover the costs of pregnancy and postnatal care. A similar tool could be used to evaluate the cost of the NeMo system on a case-by-case basis.

Phone-Sharing Model

Focus group participants expressed hesitation over the phone-sharing model, claiming some mothers would lose the phone or that the phone was likely to be damaged. One CHW believed a phone-sharing model would not work because villagers would not understand who the phone belonged to and thus who was responsible for it. Again, emphasis was placed on the need to sensitize the community to the importance and utility of the device and phone-sharing model in order to gain acceptance. Because CHWs in Uganda are volunteer health care providers, it is frowned upon for them to accept money for their services. Thus, multiple CHWs rejected a proposed model in which CHWs would purchase the phones with their own money and sell the band above cost to recuperate the funds they spent on the phones. As reported by focus group participants, very few CHWs own smartphones. Thus, expecting these individuals to purchase one or more smartphones for their communities is likely unreasonable. For similar reasons, it was suggested that anything CHWs were responsible for selling to mothers (such as the NeMo sensing bands) needed to have an irremovable price tag so that there would be no concern that CHWs were potentially overcharging to make a profit.

Discussion

Principal Findings

While early identification of neonatal illness can impact neonatal mortality rates and reduce the burden of treatment, identifying subtle clinical signs and symptoms of possible severe illness is

especially challenging in neonates. Thus, delays in illness identification pose a significant barrier to providing expedient care. Implementation of IMNCI and danger sign recognition have been shown to improve inequities in neonatal mortality and care-seeking behavior; however, the success of such interventions may be limited by the consistency of CHW follow-up and training [15,16,32]. Further, the current lack of sensitivity in maternal screening limits the feasibility of at-home danger sign screening [7]. The proposed NeMo system has the potential to empower mothers in community settings to assess their neonates' health from home and without support. Further validation of mothers' use during the newborn's first week of life, behavior change catalyzed by triage information, and integration into the existing health care system must still be conducted.

When asked about their prior training in antenatal care, all subjects indicated they had received some antenatal education. However, the observed discrepancy between training and recollection of danger signs emphasizes the value the NeMo device might add as an educational tool and assessment aid.

Although many subjects had never used a smartphone before the study, the majority were able to unlock the phone and navigate through the NeMo phone app without external guidance. Improvements made in the app between version I and version II pointed to the importance of precise verbal instruction. Small changes to how questions were framed and how instructions were articulated, without any change to the visual appearance of the app, significantly increased subjects' ability to correctly answer qualitative questions based on a posed scenario (Figure 4). In version II of the app, no subjects required prompting or further guidance from the study team to navigate the app, despite their unfamiliarity with smartphones prior to the study (Figure 3). This indicates sufficient verbal instruction and intuitive app design can overcome the barrier of introducing unfamiliar technology. Further, these results suggest lack of prior smartphone experience may not be a barrier to successful use of the NeMo system.

Many participants expressed their trust in the device's evaluation, with agreement to the statement, "I would trust this device to assess the health of my baby," receiving a score of 4.66 on the 5-point Likert scale. Women also indicated willingness to engage with the proposed system of phone-sharing. This acceptance suggests NeMo and other such frameworks may be acceptable in low-income, eastern Ugandan communities. While 29/32 subjects (91%) thought the device was easy to use, perceptions around learnability were not as conclusive. In addition to these results, unanimous desire for CHW support indicates the value of engaging CHWs in training programs and device management in order to successfully shift neonatal assessment to the home setting.

One challenge facing implementation of the NeMo system is creating trust in the device output sufficient enough to trigger behavior change and promote care-seeking. While care-seeking is a multifactorial and complex decision, preliminary questions were asked to gauge the subjects' initial response to the NeMo system. In order to assess the ability to incite care-seeking, the team asked subjects how they would respond if the device said the baby was sick when they believed it was healthy. As 29/32 (91%) women indicated they would seek care from a CHW or hospital, this suggests that the NeMo device may be able to trigger a behavior change to reduce time to care for sick neonates. One potential risk posed by the device is mothers ignoring their own intuition and relying too heavily on NeMo. The results of this study indicated 24/32 subjects (75%) would take their baby to a facility if they thought it was sick, even if the device indicated it was healthy. This majority suggests most mothers retain autonomy in their decision making.

The ability of the NeMo device to improve care-seeking for neonates will rely heavily on the accessibility of the technology in the home setting for LMICs. In order to successfully transition assessment from CHWs to mothers, the smartphone and device must be available at a low enough cost to enable use in low-income households. In the eastern Ugandan districts where the study was conducted, subjects indicated a median price point of 10,000 UGX ($2.70) for the sensor. To improve accessibility of the smartphone, the study team developed a community phone-sharing framework in which CHWs would manage phones for all mothers in their village to borrow in the week following birth. Interviews with CHWs revealed concern about coordinating a phone-sharing network if the CHW is receiving payment directly, and some CHWs worried about the phones being lost or broken. However, focus groups also emphasized the ability of community sensitization to educate mothers on the value of the device so as to minimize these risks and increase likelihood of purchasing the device.

Despite the promising acceptability and usability results in this study and potential for improved maternal recognition to reduce neonatal mortality, the study revealed several barriers to implementing and scaling the NeMo system. One of these issues is the phone-sharing framework, which attempts to minimize the barriers to technology in rural, low-income regions by distributing phones on a community-by-community basis. However, in order to initiate the implementation of this phone-sharing system, government or private foundations will need to provide funding and support. Without their aid, the likelihood of the populations studied herein accessing this technology is low until smartphone penetration increases substantially. Although implementation of the NeMo system could significantly increase the frequency of neonatal screening, the device distribution system proposed herein would continue to require CHW interaction and support. Incorporation of distribution and training in existing antenatal care programs could enable this contact for many women. As access to antenatal care is limited in many areas, alternative intervention strategies centered around the opportunity for training and distribution around the time of birth may be employed to overcome barriers to CHW access. Further, while this study demonstrates promising results regarding perceived intent to act on a recommendation provided by the device, past researchers have found that barriers, including money and transportation, often bar mothers from following through with referrals to health facilities [3,33]. Even after a mother acts on NeMo's recommendation to seek care, infrastructural or other barriers may prevent sick newborns from receiving the necessary treatments to improve neonatal mortality. This suggests initial

intervention in a region with more robust infrastructure may foster a greater impact.

Limitations

While the study was conducted in multiple subcounties and villages within the Iganga-Mayuge districts with a notably wide range of socioeconomic backgrounds, the limited sample size prevents greater generalizability of the study results. When posed hypothetical situations of neonatal illness during the simulated scenarios, a few subjects were noted to have responded about the condition of their own child instead of following the study member's prompts, leading to incorrect responses to the qualitative questions posed by the app. Due to the use of a convenience sampling methodology, there was potential for positive bias and interviewer bias. Steps, including the use of structured questionnaires and predetermined simulated use procedures, were implemented to minimize this bias. Trained local interpreters were used to minimize interviewer bias. Because two villages overlapped between implementation of version I and version II of the app, there exists the possibility of study contamination between these populations; however, as the NeMo system never remained with the women beyond the interview and study members gave no indication of actions or questions being completed correctly, the risk of contamination was considered to be very low.

Conclusions

The results of this study indicate that women are receptive to and capable of using NeMo to screen their newborns for neonatal danger signs. This study suggests that the NeMo system could be an acceptable, easy to use resource capable of expediting identification of neonatal illness to decrease delays in care-seeking behavior. While additional studies are required to assess the NeMo system's ability to safely and effectively decrease delays in accessing neonatal care, this system, if further developed, could be an approach to decrease preventable neonatal deaths by empowering and educating mothers to detect danger signs in their newborn.

Acknowledgments

This study was supported by a grant from the Bill and Melinda Gates Foundation and the Global Social Venture Competition. Additional support was provided by the Johns Hopkins Center for Bioengineering Innovation and Design (CBID), the Johns Hopkins Center for Global Health (CGH), the Johns Hopkins Global mHealth Initiative (JHU-GMI), Makerere University School of Public Health, and Uganda Development and Health Associates.

Conflicts of Interest

None declared.

References

1. UNICEF. 2018 Mar. Neonatal Mortality URL: https://data.unicef.org/topic/child-survival/neonatal-mortality/ [accessed 2019-06-19]
2. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 2012 Jun 09;379(9832):2151-2161 [FREE Full text] [doi: 10.1016/S0140-6736(12)60560-1] [Medline: 22579125]
3. Waiswa P, Kallander K, Peterson S, Tomson G, Pariyo GW. Using the three delays model to understand why newborn babies die in eastern Uganda. Trop Med Int Health 2010 Aug;15(8):964-972. [doi: 10.1111/j.1365-3156.2010.02557.x] [Medline: 20636527]
4. Mbonye AK, Sentongo M, Mukasa GK, Byaruhanga R, Sentumbwe-Mugisa O, Waiswa P, Uganda Decade of ChangeFuture Implications Analysis Group. Newborn survival in Uganda: a decade of change and future implications. Health Policy Plan 2012 Jul;27 Suppl 3:iii104-iii117. [doi: 10.1093/heapol/czs045] [Medline: 22692413]
5. Bhutta ZA, Das JK, Bahl R, Lawn JE, Salam RA, Paul VK, Lancet Newborn Interventions Review Group, Lancet Every Newborn Study Group. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? Lancet 2014 Jul 26;384(9940):347-370. [doi: 10.1016/S0140-6736(14)60792-3] [Medline: 24853604]
6. Zaman K, Zeitlyn S, Chakraborty J, de Francisco A, Yunus M. Acute lower respiratory infections in rural Bangladeshi children: patterns of treatment and identification of barriers. Southeast Asian J Trop Med Public Health 1997 Mar;28(1):99-106. [Medline: 9322291]
7. Choi Y, El Arifeen S, Mannan I, Rahman SM, Bari S, Darmstadt GL, Projahnmo Study Group. Can mothers recognize neonatal illness correctly? Comparison of maternal report and assessment by community health workers in rural Bangladesh. Trop Med Int Health 2010 Jun;15(6):743-753 [FREE Full text] [doi: 10.1111/j.1365-3156.2010.02532.x] [Medline: 20406425]
8. Stoll BJ. The global impact of neonatal infection. Clin Perinatol 1997 Mar;24(1):1-21. [Medline: 9099499]
9. Young Infants Clinical Signs Study Group. Clinical signs that predict severe illness in children under age 2 months: a multicentre study. Lancet 2008 Jan 12;371(9607):135-142. [doi: 10.1016/S0140-6736(08)60106-3] [Medline: 18191685]
10. World Health Organization. 1997 Sep. First Global Review and Coordination Meeting on Integrated Management of Childhood Illness (IMCI) URL: https://apps.who.int/iris/bitstream/handle/10665/64629/WHO_CHD_97.11.pdf?sequence=1&isAllowed=y [accessed 2019-04-02]

11. World Health Organization. Integrated Management of Childhood Illness. In: Distance Learning Course IMCI. Switzerland: WHO Press; 2014.
12. Opiyo N, English M. What clinical signs best identify severe illness in young infants aged 0-59 days in developing countries? A systematic review. Arch Dis Child 2011 Nov;96(11):1052-1059. [doi: 10.1136/adc.2010.186049] [Medline: 21220263]
13. Awasthi S, Verma T, Agarwal M. Danger signs of neonatal illnesses: perceptions of caregivers and health workers in northern India. Bull World Health Organ 2006 Oct;84(10):819-826 [FREE Full text] [Medline: 17128362]
14. Amarasiri de Silva MW, Wijekoon A, Hornik R, Martines J. Care seeking in Sri Lanka: one possible explanation for low childhood mortality. Soc Sci Med 2001 Nov;53(10):1363-1372. [Medline: 11676406]
15. Bhandari N, Mazumder S, Taneja S, Sommerfelt H, Strand TA, IMNCI Evaluation Study Group. Effect of implementation of Integrated Management of Neonatal and Childhood Illness (IMNCI) programme on neonatal and infant mortality: cluster randomised controlled trial. BMJ 2012 Mar 21;344:e1634 [FREE Full text] [doi: 10.1136/bmj.e1634] [Medline: 22438367]
16. Costello A, Dalglish S, Aboubaker S, Bahl R, Boschi-Pinto C, Daelmans B, et al. World Health Organization. 2016 Nov. Towards a Grand Convergence for Child Survival and Health URL: https://apps.who.int/iris/bitstream/handle/10665/251855/WHO-MCA-16.04-eng.pdf?sequence=1 [accessed 2019-04-01]
17. Taneja S, Bahl S, Mazumder S, Martines J, Bhandari N, Bhan MK. Impact on inequities in health indicators: Effect of implementing the integrated management of neonatal and childhood illness programme in Haryana, India. J Glob Health 2015 Jun;5(1):010401 [FREE Full text] [doi: 10.7189/jogh.05.010401] [Medline: 25674350]
18. Kananura RM, Tetui M, Mutebi A, Bua JN, Waiswa P, Kiwanuka SN, et al. The neonatal mortality and its determinants in rural communities of Eastern Uganda. Reprod Health 2016 Feb 16;13:13 [FREE Full text] [doi: 10.1186/s12978-016-0119-y] [Medline: 26883425]
19. McPherson R, Hodgins S. Postnatal home visitation: Lessons from country programs operating at scale. J Glob Health 2018 Jun;8(1):010422 [FREE Full text] [doi: 10.7189/jogh.08.010422] [Medline: 29977530]
20. Jitta J, Kyaddondo D. The Republic of Uganda: Ministry of Health. 2008 Feb. Situation Analysis of Newborn Health in Uganda URL: http://chdc.mak.ac.ug/sites/default/files/Situation%20Analysis%20%20of%20Newborn%20Health%202008.pdf [accessed 2019-04-02] [WebCite Cache ID 77LA3U2Ir]
21. Darmstadt GL, Baqui AH, Choi Y, Bari S, Rahman SM, Mannan I, Bangladesh Projahnmo-2 (Mirzapur) Study Group. Validation of community health workers' assessment of neonatal illness in rural Bangladesh. Bull World Health Organ 2009 Jan;87(1):12-19 [FREE Full text] [Medline: 19197400]
22. Bazzano AN, Kirkwood BR, Tawiah-Agyemang C, Owusu-Agyei S, Adongo PB. Beyond symptom recognition: care-seeking for ill newborns in rural Ghana. Trop Med Int Health 2008 Jan;13(1):123-128 [FREE Full text] [doi: 10.1111/j.1365-3156.2007.01981.x] [Medline: 18291010]
23. Yazdi Y, Acharya S. A new model for graduate education and innovation in medical technology. Ann Biomed Eng 2013 Sep 14;41(9):1822-1833. [doi: 10.1007/s10439-013-0869-4] [Medline: 23943068]
24. Labrique A, Christian P, Klemm R, Rashid M, Shamim A, Massie A, et al. A cluster-randomized, placebo-controlled, maternal vitamin A or beta-carotene supplementation trial in Bangladesh: design and methods. Trials 2011 Apr 21;12:102 [FREE Full text] [doi: 10.1186/1745-6215-12-102] [Medline: 21510905]
25. Klemm RDW, Labrique AB, Christian P, Rashid M, Shamim AA, Katz J, et al. Newborn vitamin A supplementation reduced infant mortality in rural Bangladesh. Pediatrics 2008 Jul;122(1):e242-e250. [doi: 10.1542/peds.2007-3448] [Medline: 18595969]
26. Betjeman TJ, Soghoian SE, Foran MP. mHealth in Sub-Saharan Africa. Int J Telemed Appl 2013;2013:482324 [FREE Full text] [doi: 10.1155/2013/482324] [Medline: 24369460]
27. Thondoo M, Strachan DL, Nakirunda M, Ndima S, Muiambo A, Källander K, et al. Potential Roles of Mhealth for Community Health Workers: Formative Research With End Users in Uganda and Mozambique. JMIR Mhealth Uhealth 2015;3(3):e76 [FREE Full text] [doi: 10.2196/mhealth.4208] [Medline: 26206419]
28. O'Donovan J, Kabali K, Taylor C, Chukhina M, Kading JC, Fuld J, et al. The use of low-cost Android tablets to train community health workers in Mukono, Uganda, in the recognition, treatment and prevention of pneumonia in children under five: a pilot randomised controlled trial. Hum Resour Health 2018 Dec 19;16(1):49 [FREE Full text] [doi: 10.1186/s12960-018-0315-7] [Medline: 30231894]
29. Uganda Bureau of Statistics. 2017 Apr. National Population and Housing Census 2014 - Area Specific Profiles - Iganga District URL: https://www.ubos.org/wp-content/uploads/publications/2014CensusProfiles/IGANGA.pdf [accessed 2019-04-01] [WebCite Cache ID 77LAKMywN]
30. UNICEF. Statistics by Country: Uganda URL: https://www.unicef.org/infobycountry/uganda_statistics.html [accessed 2019-04-02] [WebCite Cache ID 77LAadGBP]
31. Rafa M, Moyer J, Wang X, Sutton P. USAID. 2017 Nov. Estimating District GDP in Uganda URL: https://pardee.du.edu/sites/default/files/Estimating%20District%20GDP%20in%20Uganda%20.pdf [accessed 2019-04-02] [WebCite Cache ID 77LArjgsI]
32. Gerensea H, Kebede A, Baraki Z, Berihu H, Zeru T, Birhane E, et al. Consistency of Integrated Management of Newborn and Childhood Illness (IMNCI) in Shire Governmental Health Institution in 2017. BMC Res Notes 2018 Jul 16;11(1):476 [FREE Full text] [doi: 10.1186/s13104-018-3588-y] [Medline: 30012196]

33. Peterson S, Nsungwa-Sabiiti J, Were W, Nsabagasani X, Magumba G, Nambooze J, et al. Coping with paediatric referral--Ugandan parents' experience. Lancet 2004 Jun 12;363(9425):1955-1956. [doi: 10.1016/S0140-6736(04)16411-8] [Medline: 15194257]

Abbreviations

CBID: Johns Hopkins University Center for Bioengineering Innovation and Design
CGH: Johns Hopkins Center for Global Health
CHW: community health worker
GIF: graphic interchange format
IMNCI: Integrated Management of Neonatal and Childhood Illness
IQR: interquartile range
JHU-GMI: Johns Hopkins Global mHealth Initiative
LMIC: low- and middle-income country
MAKSPH IRB: Institutional Review Board of Makerere University School of Public Health
mHealth: mobile health
NeMo: Neonatal Monitoring
TBA: traditional birth assistant
UDHA: Uganda Development and Health Associates
UGX: Ugandan shilling
VHT: village health team
WHO: World Health Organization

Edited by G Eysenbach; submitted 30.04.19; peer-reviewed by M Wiens, J Kabakyenga, C Perrin; comments to author 21.05.19; revised version received 15.07.19; accepted 19.07.19; published 15.08.19.

Please cite as:
Vanosdoll M, Ng N, Ho A, Wallingford A, Xu S, Matin SB, Verma N, Farzin A, Golden WC, Yazdi Y, Waiswa P, Labrique A, Acharya S
A Novel Mobile Health Tool for Home-Based Identification of Neonatal Illness in Uganda: Formative Usability Study
JMIR Mhealth Uhealth 2019;7(8):e14540
URL: http://mhealth.jmir.org/2019/8/e14540/
doi:10.2196/14540
PMID:31418428

Original Paper

Analysis of Digital Documentation Speed and Sequence Using Digital Paper and Pen Technology During the Refugee Crisis in Europe: Content Analysis

Kai Kehe[1,2], MD, MBA; Roland Girgensohn[3], PhD; Walter Swoboda[4], MSc, MD; Dan Bieler[5], MD; Axel Franke[5], MD; Matthias Helm[6], MD; Martin Kulla[6], MD; Kerstin Luepke[7], MD; Thomas Morwinsky[7], MSc; Markus Blätzinger[8], MSc; Katalyn Rossmann[7], DVM, MSc

[1]Department F, Bundeswehr Medical Academy, Munich, Germany

[2]Walther Straub Institute of Pharmacology and Toxicology, Ludwig-Maximilians-University of Munich, Munich, Germany

[3]Department E, Bundeswehr Medical Academy, Munich, Germany

[4]University of Applied Science Neu-Ulm, Neu-Ulm, Germany

[5]Department of Trauma Surgery and Orthopaedics, Reconstructive Surgery, Hand Surgery, Burn Medicine, German Armed Forces Central Hospital Koblenz, Koblenz, Germany

[6]Department of Anaesthesiology, Intensive Care, Emergency Medicine and Pain Medicine, German Armed Forces Hospital, Ulm, Germany

[7]Department VI-2.2, Bundeswehr Medical Service Headquarters, Munich, Germany

[8]Akademie der Unfallchirurgie, Munich, Germany

Corresponding Author:
Kai Kehe, MD, MBA
Department F
Bundeswehr Medical Academy
Ingolstädter Str 240
Munich, 80939
Germany
Phone: 49 31687422
Fax: 49 31687439
Email: kai.kehe@lrz.uni-muenchen.de

Abstract

Background: The Syria crisis has forced more than 4 million people to leave their homeland. As a result, in 2016, an overwhelming number of refugees reached Germany. In response to this, it was of utmost importance to set up refugee camps and to provide humanitarian aid, but a health surveillance system was also implemented in order to obtain rapid information about emerging diseases.

Objective: The present study describes the effects of using digital paper and pen (DPP) technology on the speed, sequence, and behavior of epidemiological documentation in a refugee camp.

Methods: DPP technology was used to examine documentation speed, sequence, and behavior. The data log of the digital pens used to fill in the documentation was analyzed, and each pen stroke in a field was recorded using a timestamp. Documentation time was the difference between first and last stroke on the paper, which includes clinical examination and translation.

Results: For three months, 495 data sets were recorded. After corrections had been made, 421 data sets were considered valid and subjected to further analysis. The median documentation time was 41:41 min (interquartile range 29:54 min; mean 45:02 min; SD 22:28 min). The documentation of vital signs ended up having the strongest effect on the overall time of documentation. Furthermore, filling in the free-text field clinical findings or therapy or measures required the most time (mean 16:49 min; SD 20:32 min). Analysis of the documentation sequence revealed that the final step of coding the diagnosis was a time-consuming step that took place once the form had been completed.

Conclusions: We concluded that medical documentation using DPP technology leads to both an increase in documentation speed and data quality through the compliance of the data recorders who regard the tool to be convenient in everyday routine. Further analysis of more data sets will allow optimization of the documentation form used. Thus, DPP technology is an effective tool for the medical documentation process in refugee camps.

(JMIR Mhealth Uhealth 2019;7(8):e13516) doi:10.2196/13516

KEYWORDS

digital documentation; digital pen; digital paper; refugee camp; refugee crisis; Europe; Germany; epidemiology

Introduction

Since 2010, the geopolitical situation in the Middle East, as well as in Africa, has become more and more fragile. An increasing number of refugees, asylum seekers and migrants are attempting to reach the European Union [1]. A refugee is defined as a person fleeing persecution or lack of protection [2], and the protection of refugees is set out in the 1951 Refugee Convention and its 1967 Protocol [3]. An asylum seeker is someone who is claiming refugee status but whose status has not yet been determined; however, the term migrant is loosely defined. United Nations High Commissioner for Refugees (UNHCR) defines migrants as people who (temporarily) change their place of residence to another country or administrative unit [4]. Europe is still facing a huge influx of refugees and migrants [5], with the Syria crisis in particular forcing more than 4 million people to leave their homeland [6]. In 2015, approximately 1 million refugees entered the European Union [7].

Refugees come from devastated countries with poor health standards, and they cross several borders and live under poor conditions during their journey. Due to Europe's overcrowded reception stations, with their limited sanitary resources, there is a constant risk of the spread of communicable diseases. Thus, these stations pose a substantial health risk to Europe that needs to be controlled. European (as well as national) health security depends on individual health security to protect public health [8,9].

A medical information technology (IT) system dealing with the movements of significant numbers of patients requires several components: patient tracking, patient regulating, medical documentation management and exchange, medical asset tracking, medical capability assessment and sustainability analysis, provision of epidemiological statistics, report generation, and provision of relevant medical data [10]. The refugee crisis therefore called for the rapid establishment of a health surveillance system, and especially an early warning and detection system in order to meet minimum quality standards as defined by the criteria above. An early warning surveillance system is an effective tool to reduce outbreaks in refugee camps, with examples such as Bill Gates recommending building up such a warning and response system in the context of the Ebola crisis [11]. Syndromes definitions used in this paper were developed by the NATO Centre of Excellence for Military Medicine, Munich, and will be published elsewhere (personal communication by Katalyn Rossmann, 18.02.2018).

A syndromic surveillance system was implemented in Italy during a similar situation. However, documentation involved a purely paper-based system from which reporting data had to be extracted manually, and reports were being sent by e-mail to the Italian Ministry of Health or the National Centre for Epidemiology, Surveillance and Health Promotion of the National Institute of Health (CNESPS-ISS). Napoli et al concluded that this system had limitations regarding data quality and the rapid preparation of reports, so a web-based system was recommended to overcome these obstacles [12].

Other studies have demonstrated the effectiveness of web-based surveillance systems such as SurvNet developed by the Robert Koch-Institute [13] or the QSurveillance system used by Public Health England in the United Kingdom [14].

Adequate documentation is essential to ensure proper medical records and information flow between medical facilities. Furthermore, these data can be used for quality management systems and for health surveillance. Good data quality is essential for a scientific analysis of medical data but meeting these quality requirements poses a challenge for frequently used paper-based documentation systems.

A transfer from analogous data to digital data normally requires an additional step after patient contact, which is the manual typing of data into a local computer system, or web-based front-end of a client, or server-driven hospital information system, or a registry. This additional step has a negative effect on data quality, performance and process costs, as it is both time consuming and additional effort for the data recorder, without any instantaneously recognizable advantages [15]. It also introduces an additional possibility for data errors.

To improve the speed of documentation and to reduce errors, the intuitive layout of the documentation sheet is crucial [16]. However, paper-based documentation does not allow an analysis of the speed and sequence of filling the form. Digital paper and pen (DPP) technology aims to overcome this shortcoming, and several studies have been published demonstrating the effectiveness of DPP technology in various settings.

In Germany, DPP technology has been tested in several studies, especially in air rescue and emergency room data recording. The technology allowed medical documentation to be carried out with an electronic (digital) pen on special paper. After completing the form, the data sets from the pen were transferred to a computer system, approved by the physician or qualified medical personnel, and then stored in a hospital information system. The digital pen provided a timestamp for each data entry field on the form, which allowed in-depth analysis of the process of completing the form. On average, the approval time was less than 2 minutes. Data quality for core data was greater than 95% and superior to handwritten documentation, and both checkbox and numerical data fields were correctly recorded at 99.8% [15-18].

In addition to civilian organizations, the Bundeswehr also developed a near real-time syndromic surveillance system, known as the Visitor and Immigrant Health Surveillance and Information Tool (VISIT). This system was implemented at the refugee camp in Bad Fallingbostel (Lower Saxony, Germany), which was the first time that DPP technology for patient

recording and automated epidemiological data collection for an early warning system had been used.

The aim of the present study was to analyze the data sets concerning the speed and sequence of medical documentation to improve data quality, process performance, and costs.

Methods

Digital Paper and Pen Technology

Digital Pen

The digital pen was purchased from Diagramm Halbach (Schwerte, Germany), and the technology was developed by Anoto (Lund, Sweden). The digital pen consists of a standard ink cartridge, a front-facing infrared camera, a microprocessor, a storage chip, Bluetooth connectivity and a universal serial bus (USB) interface. A pressure sensor starts the data-capture process. The front camera is located under the ink cartridge and captures 50 frames per second with a timestamp, while the positioning of the pen on the paper is calculated by a special pattern of dots on the paper. Less than 2 square millimeters are needed for the correct calculation of the position, and every stroke is stored in the pen. The storage capacity of each pen is up to 50 filed protocols, which was never exceeded during the field trial. A battery of a charged pen lasts up to 15 h. Finally, every pen has a unique ID.

Digital Paper

All digital paper has a slight (almost invisible) pattern of black dots on it. This pattern allows the digital pen to calculate the coordinates of the pen tip on the paper as well as the unique ID of the paper. Thus, digital paper is an essential part of the digital process. The VISIT forms were printed with this pattern. The position of each data field was encoded with dotforms software (Diagramm Halbach, Germany).

Data Transfer and Validation

A DPP docking station was connected to a standard personal computer (PC) via a USB cable. After connecting the digital pen to this station, the data were transferred to the PC. Once the raw data had been transferred to a notebook, a lookup table matched the recorded coordinates with the relevant field. In free-text fields (eg, blood pressure), optical character recognition (OCR) of handwriting was used to create a digital entry. Every marked field and OCR result was recorded in an XML-file which allowed further analysis, and then all the datasets were exported as XML-files. DPP technology could store the time of documentation for every sheet and pen, thus building up a database for further analysis of the medical documentation process.

Data Collection

The data were collected over 3 months, from May 2, 2016 to August 2, 2016, during the field trial of the VISIT project at the refugee center in Bad Fallingbostel. The center is composed of two camps with a total capacity of 2000 persons. The German Red Cross (DRK) and the Johanniter-Unfallhilfe (JUH) run each camp, respectively. After study approval, the medical personnel in both camps were quickly trained to use the DPP technology. The study group consisted of two reception lines which were on duty every day, and during the study period a total of 495 data sets were collected. The Bundeswehr provided support for both camps with regard to medical personnel and medical documentation.

Medical personnel filled in the standardized protocol during medical examinations, and signs and symptoms, the clinical diagnosis, and sociodemographic characteristics (eg, year of birth, nationality and gender) were collected from the patients. The DPP technology was used consistently; however, personal information was not stored in the DPP system by paper design (no digital pattern) in order to protect patient privacy. Thus, only anonymous data were processed.

For each patient, one protocol (sheet of digital paper) was completed. Each protocol consisted of 277 fields into which data could be entered (either text or symbols, eg, check marks). Of course, for any given patient, many of the possible fields were left blank as they did not pertain to that specific patient.

At the end of each day, collected medical data were then transmitted, via a protected transfer, to central servers in Koblenz and Ulm, Germany. A data subset concerning syndromic surveillance was encrypted and transmitted to the Bundeswehr University in Munich and to the Deployment Health Surveillance Capability (DHSC) for further analysis by the specialists of the Bundeswehr Medical Service Headquarters as well as the DHSC. The DHSC provides epidemiological surveillance for the North Atlantic Treaty Organization (NATO) deployment areas and is a satellite branch of the NATO Centre of Excellence for Military Medicine (MILMED COE) in Budapest, Hungary.

Process of Medical Documentation

Documentation time was the difference between first and last stroke on the paper, which includes clinical examination and translation. In the initial phase of the study, the medical documentation process was not uniform. In one medical treatment facility, only the documentation sheet accompanied the patient. This resulted in entries being produced by different pens on one documentation sheet with multiple XML-files. As for all the other documentation processes, a unique digital pen and a documentation sheet accompanied the patient, resulting in a single XML-file per person.

Data Set

The data set was developed by the DHSC and the Bundeswehr Medical Service Headquarters. The implemented data set is a subset of the German national data set regarding anything emergency department–related that is relevant for primary health care [19]. This collection was complemented by adapting the syndrome-specific parameter, resulting in a total of 64 sets of data. For further statistical analysis, the diagnosis was matched to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) code, according to the World Health Organization (WHO) list.

The VISIT protocol sheets are available upon request from the authors.

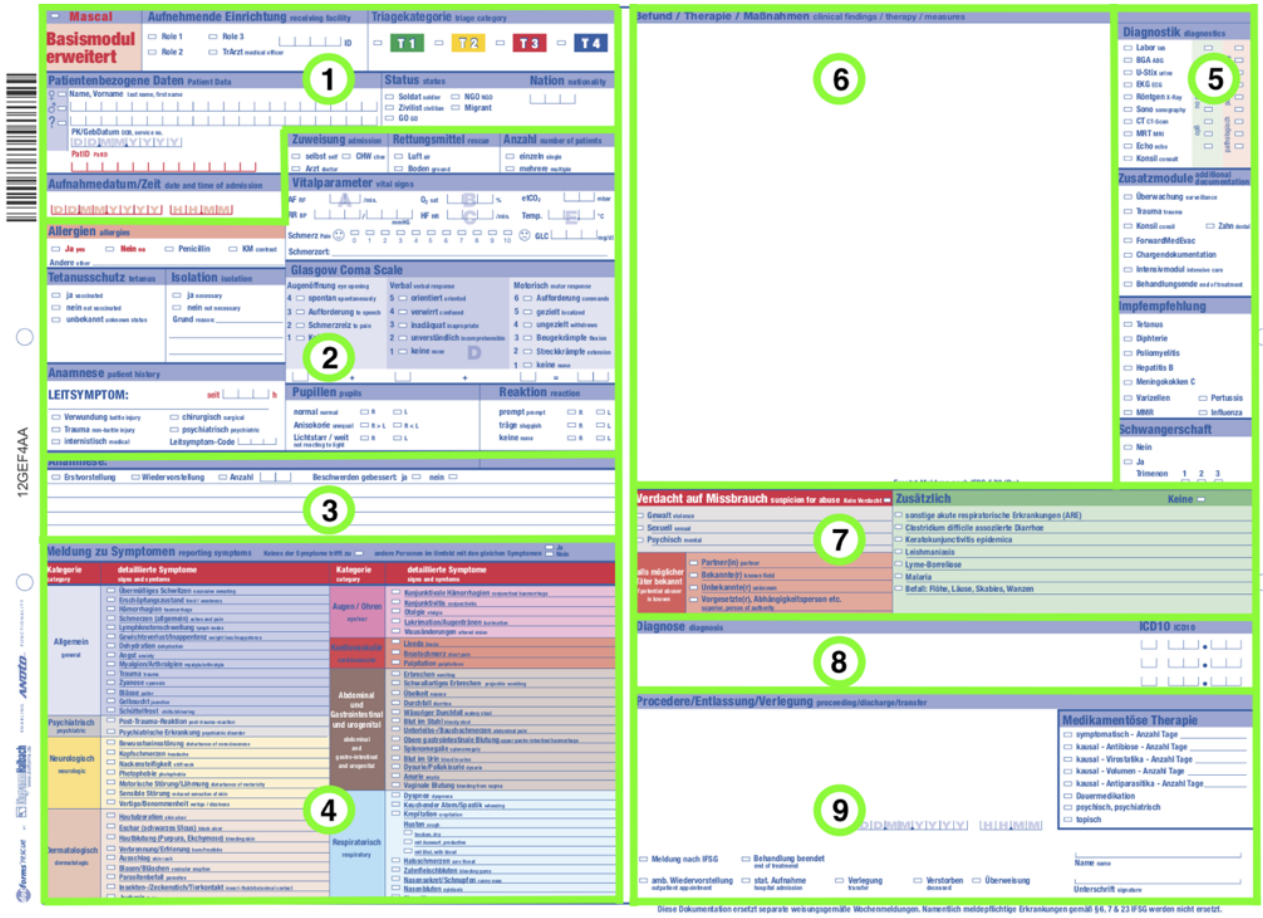
Analysis of Time Stamps

With the goal of optimizing the processing time, we asked which fields had the strongest effect on total examination time for a patient. For this analysis, we grouped the 277 fields into 70 categories, pooling fields covering different aspects of the same topic into one category. For example, the category "symptoms eye/ear" consists of five fields covering symptoms from conjunctival hemorrhaging to otalgia (see upper right corner of section 4 in Figure 1). Then, a Mann-Whitney *U* test was conducted to compare the documentation time of reports where a specific category was completed (at least one of the fields filled out) with those where this specific category was empty.

For a more detailed analysis, we also asked which fields required the most time for completion. This analysis is possible since, for each field, the time stamps of the first and the last entry into the field are stored. This is not corrected for cases where the examiner goes back and forth between fields. If, for example, the examiner starts filling out field A, then makes entries into field B, and then finally goes back to field A and completes it, then the time stamp would count the whole time from the first change in field A to the last. In this sense, our analysis captures those fields which are the most complicated to fill out.

Finally, the design of the documentation sheet determines the speed at which the form can be completed. As the digital pen stores a timestamp for each entry, the sequence of documentation can be analyzed. The total time of documentation for each sheet was set at 1, so each timestamp was normalized and expressed as a value between 0 and 1. The fields of the documentation sheet were then grouped into nine sections (Figure 1). For each sheet, the normalized time of the first entry into each section was noted, thus, the order in which the sections were worked on can be analyzed. For example, the median of these normalized times for a section expresses the time point (relative to the total time needed for the full medical examination) where half of the examiners have started on that section.

Figure 1. Form for essential documentation. Fields are clustered into nine groups. Group 1: patient information. Group 2: patient history and vital data. Group 3: free text patient history. Group 4: symptoms and signs. Group 5: diagnostic and vaccination recommendation. Group 6: free-text documentation. Group 7: sexual abuse and rare diseases. Group 8: free text diagnosis with coding. Group 9: proceeding, recommendation, and transfer.



Statistical Analysis

The XML files were imported into Microsoft Excel (version 15.29) and prepared for further analysis. Statistical analysis was performed using standard statistical software (SPSS version 24 for Windows), depicted as box-and-whisker plots. The effect of different pens on time of documentation was analyzed with the Mann-Whitney *U* test.

Ethical Approval

Prior to the investigation, ethical clearance was obtained from the Ethics Commission of the Ludwig-Maximilians-University of Munich (143-16). Verbal informed consent was received from the refugees.

Results

Effect of Optimizing Digital Pen and Paper Handling

Altogether, 495 data sets (XML files of individual documentation sheets, each documenting one patient) were generated. However, after an initial phase of training personnel and optimizing the documentation process, the procedure was adapted. During the initial phase, digital pens were available at each station. This ended up causing multiple entries per sheet being made with different digital pens, which caused problems with respect to merging data sets and data integrity. This technical problem was solved when the digital paper and pen accompanied the patient during the whole medical examination process. As a result, 74 data sets, generated during this initial phase with more than one digital pen ID, were excluded from further analysis, leaving 421 data sets to be analyzed.

Time of Documentation

The total time of documentation (time difference between first and last stroke on a documentation sheet) ranged between about 90 seconds and over 2 hours, with a median of 41:41 minutes and an interquartile range (IQR) of 29:40 minutes (mean 45:02; SD 22:28 min). The distribution of these times is skewed to the right, and the Kolmogorov-Smirnov test rejects the null hypothesis of the normal distribution (P=.03). The median time of documentation for individual pens varied from 31:44 min to 48:20 min, and the number of completed forms per pen ranged between 28-62. A summary of documentation times per pen is presented in Figure 2; however, the time of documentation for the different pens showed no significant difference, according to the Kruskal-Wallis test (P=.54).

Fields with a Significant Effect on the Speed of Documentation

Overall, 21 out of 70 categories have a significant effect on documentation time. The data are shown in Table 1. The documentation of vital signs (eg, blood pressure, blood oxygenation, heart rate, temperature, and blood glucose) had the highest impact on documentation time, with 5 out of the 21 significant categories belonging to this group. Additionally, anything connected to the calculation of the Glasgow Coma Scale also had a significant effect on overall documentation time. Interestingly, sheets where eye, ear or dermatological symptoms were documented had a significantly shorter completion time. A Mann-Whitney U test was conducted to compare the documentation time of reports with entries in a category and reports without entries in the same category. Significant effects were considered where P<.05.

Figure 2. Time of documentation. The median time of documentation for each pen was calculated. Boxplots show the lowest and highest values of time of documentation, a quantitative measure for the length of the medical process. Boxes represent the inter-quartile range (25th to 75th percentile), and whiskers indicate the minimum and maximum of the data except for outliers (shown as circles). The thick horizontal line within each box represents the median. ID: identification.

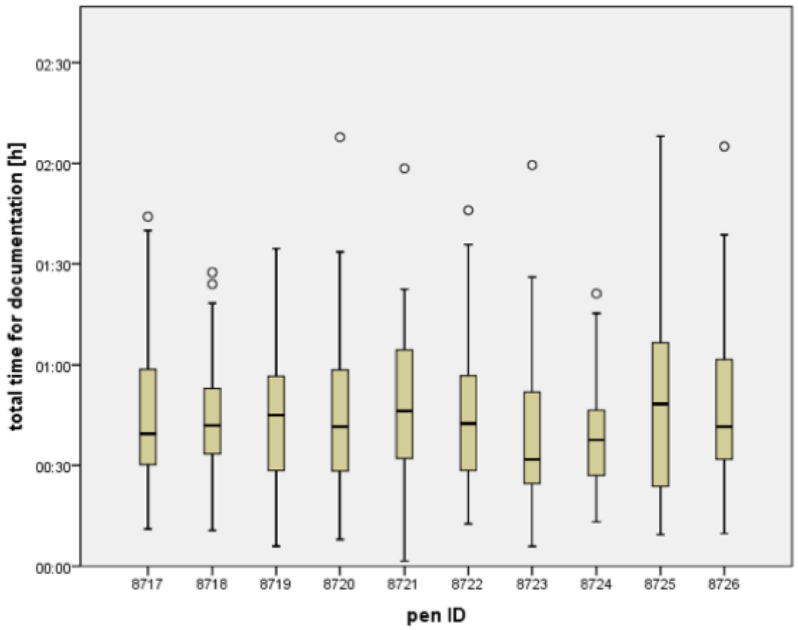


Table 1. List of field categories with significant effect on time of documentation.

Categories with significant effect	Reports, n	Average difference, min	*P* value[a]
Blood pressure	269	15:49.9	<.001
Oxygen saturation	249	15:23.9	<.001
Heart rate	250	14:35.2	<.001
Temperature	218	06:11.0	<.001
ICD[b] code			
Field 2	221	08:14.5	<.001
Field 3	95	04:38.8	.03
Urine stick	75	09:41.3	<.001
Patient history	190	07:40.4	.001
Glasgow Coma Scale			
Eye	150	04:55.6	.02
Motor	150	04:55.6	.02
Verbal	150	04:55.6	.02
Sum	88	08:14.2	.002
Symptoms			
Eye or ear	35	–08:10.0	.008
Other person with same symptoms	79	–05:25.4	.008
Dermatological	19	–08:12.9	.03
Cardiovascular	4	21:11.4	.045
Blood glucose	39	08:16.5	.03
Medication			
Symptomatic or acute	162	05:03.1	.013
Long-term	113	06:07.6	.02
Antibiotics	67	–05:12.3	.04
ECG[c]	9	17:02.5	.04

[a]Value was calculated using the *U* test.

[b]ICD: International Statistical Classification of Diseases and Related Health Problems.

[c]ECG: electrocardiogram.

Time Needed per Field

For each field, the time needed to fill it was calculated. Table 2 shows the nine fields with the highest median times (4 seconds or more) for completion. The large free-text field called clinical findings, or therapy, or measures, needed the most time (median 8:06 min; IQR 23:56 min). The next four most time-consuming fields were also free-text fields: patient history, proceeding, diagnosis and discharge.

Sequence of Documentation

The data presented in Figure 3 shows the sequence of documentation. Almost every medical examiner started with the first section (patient personal data and manner of transfer to the medical center). Section 2 (patient vital signs and medical history) usually came next. Following that, the diagnostic sections 3, 5, 6, and 7 were started more or less interchangeably and at the same time. Interestingly, half of the medical examiners made their first entry into section 4 (symptoms) only after the other diagnostic sections, namely at roughly the same time as the concluding sections 8 (diagnosis with ICD code; this was often the last section to be started) and 9 (further proceeding, discharge or transfer). It should be noted, however, that these times (except for section 1) displayed large variances as seen in Figure 3.

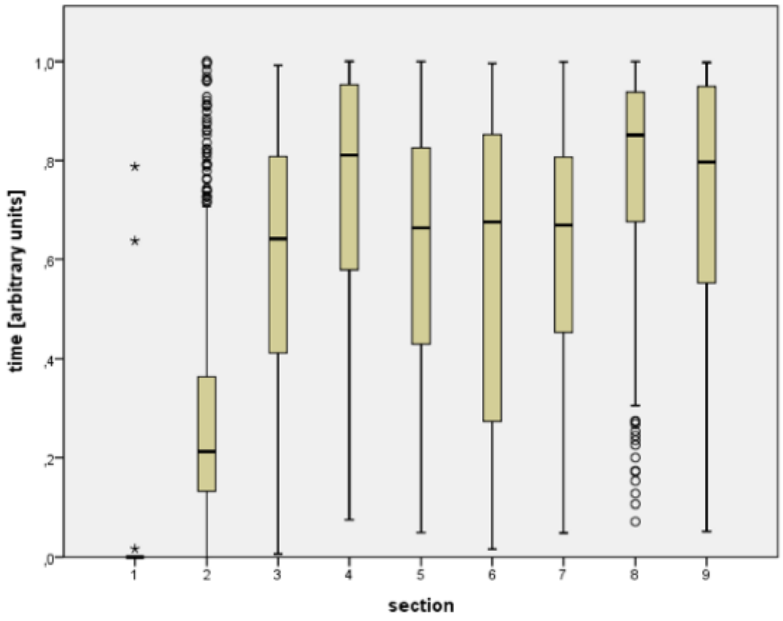
Table 2. Time needed per field. The table lists all fields with a median documentation time of 4 seconds or more (N=421).

Field	n	Median (IQR[a]), min	Mean (SD), min
Clinical findings	412	8:06 (23:56)	16:49 (20:33)
Patient history	407	2:07 (5:39)	4:35 (5:59)
Proceeding	293	1:06 (4:06)	3:59 (6:28)
Diagnosis	409	0:26 (2:35)	3:04 (6:14)
Discharge	113	0:13 (0:18)	1:01 (3:16)
Transfer	50	0:06 (0:12)	0:21 (0:57)
Date of admission	418	0:06 (0:02)	0:07 (0:06)
Allergies (other)	17	0:05 (0:05)	0:09 (0:12)
Patient ID[b]	416	0:04 (0:02)	0:11 (1:49)

[a]IQR: interquartile range.

[b]ID: identification.

Figure 3. Sequence of documentation. Fields are clustered into nine sections. Section 1: patient information. Section 2: patient history and vital signs. Section 3: free text patient history. Section 4: symptoms and signs. Section 5: diagnostics and vaccination recommendation. Section 6: free-text documentation. Section 7: sexual abuse and additional rare diseases. Section 8: free text diagnosis with coding. Section 9: proceeding, recommendation and transfer.



Number of Data entries per Category

Data are shown in Table 3. The categories more frequently completed (i.e. more than 90%) were patient data (99.8%), receiving facility (99.5%), time of admission (99.3%), proceeding date (99.0%), status (98.1%), first ICD-10 code (97.1%), patient history (96.2%), nationality (96.0%) and self-admission (95.2%). 17 of the 70 categories had no entry in all the analyzed data sets; 11 had only one entry.

Table 3. Number of data sets with entries in the field categories. Table shows categories with entries in more than 95% of data sets as well as categories which have been edited at most once (N=421).

Category	n (%)
Patient related data	420 (99.8)
Reception facility	419 (99.5)
Recording date	418 (99.3)
Procedure date	417 (99.0)
Status	413 (98.1)
ICD[a] code 1	409 (97.1)
Patient history (details)	405 (96.2)
Nationality	404 (96.0)
Self-assigning	401 (95.2)
Allergy	1 (0.2)
Number of patients	1 (0.2)
Diagnostics	
CT[b] scan	1 (0.2)
Consult	1 (0.2)
ABG[c]	0 (0)
Echocardiogram	0 (0)
MRI[d]	0 (0)
Sonography	0 (0)
X-ray	0 (0)
Vaccination	
Meningococcus	1 (0.2)
Measles, mumps, rubella	1 (0.2)
Diphtheria	0 (0)
Hepatitis B	0 (0)
Influenza	0 (0)
Pertussis	0 (0)
Poliomyelitis	0 (0)
Tetanus	0 (0)
Varicella	0 (0)
Procedure	
Transfer	1 (0.2)
Dead	1 (0.2)
Suspicion of sexual abuse	1 (0.2)
Vital CO_2[e] parameter	1 (0.2)
Consultation	1 (0.2)
MASCAL[f]	0 (0)
Medication	
Virostatic agents	0 (0)
Volume	0 (0)
Rescue	0 (0)

Category	n (%)
Triage category	0 (0)

[a]ICD: International Statistical Classification of Diseases and Related Health Problems.

[b]CT: computed tomography.

[c]ABG: arterial blood gas.

[d]MRI: magnetic resonance imaging.

[e]CO_2: carbon dioxide.

[f]MASCAL: mass casualties.

Discussion

Principal Findings

Taking handwritten notes is one of the least technical ways to collect medical data. Generations of medical doctors have been trained to use paper-based documentation and to note their clinical findings [20], however, the enormous medicolegal need for documentation poses a considerable challenge [21]. Additionally, paper-based documentation should be phased out and instead be digitized for long-term storage, integration into hospital information systems, and for further analysis with respect to quality management [22].

The speed of handwriting is limited and slows down the documentation process even further, as the use of other devices (eg, smartphones, typewriting or keyboard-based typing on personal computers) increases [23]. Therefore, forms are needed to ensure both reliable data collection as well as to speed up the process of filling them in through structured protocols. To overcome the obstacles of paper-based documentation, new technologies are emerging to support the documentation process. Speech recognition [24,25], automated data recording [26,27], and DPP technology [14] are examples of this. Handwriting recognition is improving and being implemented in standard office software suites, but it can still lead to errors [28,29]. The same holds true for speech recognition, although further training is needed as well as a relatively silent ambience around the user [23]. Voice recognition technology often shows no advantage over manual transcription [30]; however, the rise of artificial intelligence (AI) or deep machine learning in combination with voice recognition is promising new technologies to further speed up the documentation process.

In the case of mass casualties during a crisis, it is crucial to increase the speed of medical documentation without losing quality. Processing speed and data quality should be as high as possible to cope with a large number of patient casualties [31]. In this context, valid data for health surveillance are of utmost importance [11], but the recorded information is also of interest for evaluating complex interventions [32]. The data imported into automated health surveillance systems should be as simple as possible, and in the best case it should be an automatic byproduct of the legally required patient documentation. DPP technology has proven to be a good combination of both paper-based records and electronic documentation [14,33].

No statistical difference was identified between the digital pens used. Thus, both reception lines worked equally effectively. Apart from the first week, use of the technology was very reliable. The pen strokes from the 421 collected data sets allowed a reconstruction of the sequence in which the fields were filled in the form. The graphical representation of the sequence showed that the design of the report seems to be appropriate. As stated above, the diagnostic sections were recorded earlier than the final diagnosis field, including the ICD-10 code that was often the last field to be worked on. A request submitted to the medical documentation team revealed that matching the diagnosis to the ICD-10 code required additional time, and thus the ICD-10 code was added as the last step of medical documentation. This personal observation is backed up by the fact that the step to match the diagnosis with the ICD code had a significant impact on overall documentation time. Additionally, the diagnosis and proceedings fields were among the five fields with the longest documentation time.

Another time-consuming block involved everything related to the examination of the patient. The clinical findings, or therapy, or measures free-text field required the most time for documentation per field. Furthermore, the recording of vital signs had the highest impact on documentation time. A total of 98% of the clinical findings, or therapy, or measures field, and approximately 60% of the significant vital sign fields, were filled in. Thus, the optimization of this step has great time-saving potential.

The fields with the most data entries are assumed to be more relevant. By contrast, fields with little or no data entry must be considered less important and are thus probably dispensable in further studies. The recorded pen strokes clearly showed that data sets were recorded during the patient examination process and not afterwards. It has been shown in previous studies that the completeness of data sets is superior to any kind of retrospective documentation [34,35]. Furthermore, total documentation time is reduced using DPP technology compared to paper-based documentation, where secondary data entry in a computer system is necessary and errors may occur. The time required to transfer all the data sets from a pen is less than 10 min. Given that 30 min per form are needed for secondary data entry, the time saved is estimated to be 210 hours in total, which is approximately 26 days of work, or more than one month, for one documentation assistant. Thus, DPP technology has a return on investment that will increase the longer this technology is used. Additionally, the time-saving effect in epidemiological surveillance is crucial, as with DPP technology more surveillance reports can be generated in less time compared to traditional paper-based systems [32].

The strength of this study lies in its description of how DPP technology can be used to analyze the speed of medical

documentation, as well as the behavior of the medical documentation team [36]. In particular, it was possible to conduct detailed analysis of the documentation sequence and speed with regard to the form used. Furthermore, the study highlighted the possibility of training medical personnel very quickly to use DPP technology. However, the main limitation of this study lies in the limited timeframe of data capture. It could be seen as a longitudinal cross-sectional study with the inherent weaknesses of this study design, but to acquire more valid data, it is necessary to analyze more data sets. Furthermore, it may be of interest to see whether a learning curve is detectable as users adapt to this new technology. A cohort design is likely to produce more reliable and precise data regarding these questions.

Conclusion

We conclude that medical documentation using DPP technology leads to an increase in documentation speed and quality. Further analysis of more data sets will allow optimization of the documentation form used. DPP technology is an effective tool for the medical documentation process in refugee camps.

Acknowledgments

The research project was funded by the German Bundeswehr (grant #34K3-S-171515).

Conflicts of Interest

None declared.

References

1. Daynes L. The health impacts of the refugee crisis: a medical charity perspective. Clin Med (Lond) 2016 Oct;16(5):437-440 [FREE Full text] [doi: 10.7861/clinmedicine.16-5-437] [Medline: 27697805]
2. International Committee of the Red Cross. 1958. Treaties, States parties, and Commentaries - Geneva Conventions of 1949 and Additional Protocols, and their Commentaries URL: https://ihl-data-ba-ses.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=ECCFBF83D7C4D815C12563CD0042C1E9 [accessed 2017-02-11]
3. UNHCR. The UN Refugee Agency. 2015. Refugees/Migrants Emergency Response: Mediterranean URL: http://data.unhcr.org/mediterranean/regional.php [accessed 2017-02-11] [WebCite Cache ID 6oCm653XY]
4. UNHCR. The UN Refugee Agency. 2013 Oct. Expert roundtable: Interpretation of the extended refugee definition contained in the 1984 Cartagena Declaration on Refugees URL: https://www.unhcr.org/protection/expert/53bd4d0c9/summary-conclusions-interpretation-extended-refugee-definition-1984-cartagena.html?query=refugee%20definition [accessed 2019-08-07]
5. Kulla M, Josse F, Stierholz M, Hossfeld B, Lampl L, Helm M. Initial assessment and treatment of refugees in the Mediterranean Sea (a secondary data analysis concerning the initial assessment and treatment of 2656 refugees rescued from distress at sea in support of the EUNAVFOR MED relief mission of the EU). Scand J Trauma Resusc Emerg Med 2016 May 20;24:75 [FREE Full text] [doi: 10.1186/s13049-016-0270-z] [Medline: 27206483]
6. Brannan S, Campbell R, Davies M, English V, Mussell R, Sheather J. The Mediterranean refugee crisis: ethics, international law and migrant health. J Med Ethics 2016 Mar 22;42(4):269-270. [doi: 10.1136/medethics-2016-103444]
7. The UN Refugee Agency. 2017. Refugees and migrants arrivals to Europe in 2017 URL: https://data2.unhcr.org/en/documents/download/62023 [accessed 2018-02-18] [WebCite Cache ID 6xJxHUR7a]
8. Gonzalez-Block MA. Health policy and systems research agendas in developing countries. Health Res Policy Syst 2004 Aug 05;2(1):6 [FREE Full text] [doi: 10.1186/1478-4505-2-6] [Medline: 15296509]
9. Pottie K, Hui C, Rahman P, Ingleby D, Akl EA, Russell G, et al. Building Responsive Health Systems to Help Communities Affected by Migration: An International Delphi Consensus. Int J Environ Res Public Health 2017 Feb 03;14(2) [FREE Full text] [doi: 10.3390/ijerph14020144] [Medline: 28165380]
10. NATO. 2013 Jun. NATO Standard AJMedP-5 Allied Joint Doctrine for Medical Communications and information systems (MedCIS) Internet; 2013?46 URL: https://www.coemed.org/images/Courses_pdf/stanags/02_AJMEDP/2562_AJMedP-5%20EDA%20V1%20E.pdf [accessed 2019-08-03] [WebCite Cache ID 6xN5WNfRB]
11. Gates B. The next epidemic--lessons from Ebola. N Engl J Med 2015 Apr 09;372(15):1381-1384. [doi: 10.1056/NEJMp1502918] [Medline: 25853741]
12. Napoli C, Riccardo F, Declich S, Dente MG, Pompa MG, Rizzo C, National Working Group. An early warning system based on syndromic surveillance to detect potential health emergencies among migrants: results of a two-year experience in Italy. Int J Environ Res Public Health 2014 Aug 20;11(8):8529-8541 [FREE Full text] [doi: 10.3390/ijerph110808529] [Medline: 25140999]
13. Krause G, Altmann D, Faensen D, Porten K, Benzler J, Pfoch T, et al. SurvNet electronic surveillance system for infectious disease outbreaks, Germany. Emerg Infect Dis 2007 Oct;13(10):1548-1555 [FREE Full text] [doi: 10.3201/eid1310.070253] [Medline: 18258005]

14. Harcourt SE, Smith GE, Elliot AJ, Pebody R, Charlett A, Ibbotson S, et al. Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. Epidemiol Infect 2012 Jan;140(1):100-105. [doi: 10.1017/S095026881100046X] [Medline: 21473803]
15. Despont-Gros C, Bœuf C, Geissbuhler A, Lovis C. The digital pen and paper technology: implementation and use in an existing clinical information system. Stud Health Technol Inform 2005;116:328-333. [Medline: 16160280]
16. Helm M, Hauke J, Schlechtriemen T, Renner D, Lampl L. [Primary documentation quality for paper-assisted digital mission data documentation. Initial results from the air rescue service]. Anaesthesist 2009 Jan 10;58(1):24-29. [doi: 10.1007/s00101-008-1435-6] [Medline: 19132331]
17. Helm M, Hauke J, Schlechtriemen T, Renner D, Lampl L. [Paper-assisted digital Mission documentation in air rescue services. Quality management in preclinical emergency medicine]. Anaesthesist 2007 Jun 07;56(9):877-885. [doi: 10.1007/s00101-007-1215-8] [Medline: 17554515]
18. Helm M, Kulla M, Birkenmaier H, Hauke J, Lampl L. Improved Data Quality by Pen Computer-Assisted Emergency Room Data Recording Following Major Trauma in the Military Setting. Eur J Trauma Emerg Surg 2007 Feb;33(1):52-58. [doi: 10.1007/s00068-007-6187-5] [Medline: 26815975]
19. Kulla M, Röhrig R, Helm M, Bernhard M, Gries A, Lefering R, Sektion Notaufnahmeprotokoll der DIVI. [National data set "emergency department": development, structure and approval by the Deutsche Interdisziplinäre Vereinigung für Intensivmedizin und Notfallmedizin]. Anaesthesist 2014 Mar;63(3):243-252. [doi: 10.1007/s00101-014-2295-x] [Medline: 24615292]
20. Barr MS. The clinical record: a 200-year-old 21st-century challenge. Ann Intern Med 2010 Nov 16;153(10):682-683. [doi: 10.7326/0003-4819-153-10-201011160-00015] [Medline: 21079228]
21. Busch J. Electronic records: 21st century EMS: who likes paperwork anyway? EMS World 2012 Feb;41(2):35. [Medline: 22413698]
22. Krumm R, Semjonow A, Tio J, Duhme H, Bürkle T, Haier J, et al. The need for harmonized structured documentation and chances of secondary use - results of a systematic analysis with automated form comparison for prostate and breast cancer. J Biomed Inform 2014 Oct;51:86-99 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.008] [Medline: 24747879]
23. Kiefer M, Schuler S, Mayer C, Trumpp NM, Hille K, Sachse S. Handwriting or Typewriting? The Influence of Pen- or Keyboard-Based Writing Training on Reading and Writing Performance in Preschool Children. Adv Cogn Psychol 2015;11(4):136-146 [FREE Full text] [doi: 10.5709/acp-0178-7] [Medline: 26770286]
24. Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of Documentation Speed Using Web-Based Medical Speech Recognition Technology: Randomized Controlled Trial. J Med Internet Res 2015 Nov 03;17(11):e247 [FREE Full text] [doi: 10.2196/jmir.5072] [Medline: 26531850]
25. Ahlgrim C, Maenner O, Baumstark MW. Introduction of digital speech recognition in a specialised outpatient department: a case study. BMC Med Inform Decis Mak 2016 Oct 18;16(1):132 [FREE Full text] [doi: 10.1186/s12911-016-0374-4] [Medline: 27756284]
26. Krakamp B, Kirschberg O, Scheding A, Emmerich D, Klein S, Saers T. Computer-aided recording of automatic endoscope washing and disinfection processes as an integral part of medical documentation for quality assurance purposes. BMC Gastroenterol 2010 Jul 08;10:76 [FREE Full text] [doi: 10.1186/1471-230X-10-76] [Medline: 20615248]
27. Chan TC, Griswold WG, Buono C, Kirsh D, Lyon J, Killeen JP, et al. Impact of wireless electronic medical record system on the quality of patient documentation by emergency field responders during a disaster mass-casualty exercise. Prehosp Disaster Med 2011 Aug;26(4):268-275. [doi: 10.1017/S1049023X11006480] [Medline: 21993045]
28. Shawahna R, Rahman N, Ahmad M, Debray M, Yliperttula M, Declèves X. Impact of prescriber's handwriting style and nurse's duty duration on the prevalence of transcription errors in public hospitals. J Clin Nurs 2013 Feb;22(3-4):550-558. [doi: 10.1111/j.1365-2702.2012.04076.x] [Medline: 23228072]
29. Mathaiyan J, Jain T, Dubashi B, Batmanabane G. Prescription, Transcription and Administration Errors in Out- Patient Day Care Unit of a Regional Cancer Centre in South India. Asian Pac J Cancer Prev 2016;17(5):2611-2617 [FREE Full text] [Medline: 27268639]
30. Suvalsky S, Boelman V. Voice recognition technology vs. manual transcription: Is voice recognition technology superior to manual transcription for surgical pathology--in terms of return on investment? A formal study says the answer is no. MLO Med Lab Obs 2012 Sep;44(9):36, 38-36, 39. [Medline: 23002510]
31. Dugas M, Dugas-Breit S. A generic method to monitor completeness and speed of medical documentation processes. Methods Inf Med 2012;51(3):252-257. [doi: 10.3414/ME09-01-0085] [Medline: 21915436]
32. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
33. Mathewson AA, Daly ER, Cavallo SJ, Alic A. Use of Digital Pens for Rapid Epidemiologic Data Collection During a Foodborne Outbreak Investigation. Disaster Med Public Health Prep 2015 Aug;9(4):349-353. [doi: 10.1017/dmp.2015.43] [Medline: 25939887]
34. Davis GJ. Forensic documentation in trauma resuscitations: practical considerations. Int J Trauma Nurs 1999;5(4):134-135. [Medline: 10661144]

35. Menke JA, Broner CW, Campbell DY, McKissick MY, Edwards-Beckett JA. Computerized clinical documentation system in the pediatric intensive care unit. BMC Med Inform Decis Mak 2001;1:3 [FREE Full text] [Medline: 11604105]
36. Helm M, Jaehun K, Lampl L, Hauke J. Zum Dokumentationsverhalten von Notärzten während des Einsatzes. Notfall Rettungsmed 2011 Apr 10;15(2):127-135. [doi: 10.1007/s10049-011-1450-z]

Abbreviations

AI: artificial intelligence
CNESPS-ISS: National Centre for Epidemiology, Surveillance and Health Promotion of the National Institute of Health
DHSC: Deployment Health Surveillance Capability
DPP: digital pen and paper
DRK: German Red Cross
ICD-10: International Statistical Classification of Diseases and Related Health Problems–10th revision
IQR: interquartile range
IT: information technology
JUH: Johanniter-Unfallhilfe
MILMED COE: NATO Centre of Excellence for Military Medicine
NATO: North Atlantic Treaty Organization
OCR: optical character recognition
PC: personal computer
UNHCR: United Nations High Commissioner for Refugees
USB: universal serial bus
VISIT: Visitor and Immigrant Health Surveillance and Information Tool
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.01.19; peer-reviewed by V Guardabasso, V Tiase; comments to author 19.02.19; revised version received 31.05.19; accepted 21.07.19; published 19.08.19.

Please cite as:
Kehe K, Girgensohn R, Swoboda W, Bieler D, Franke A, Helm M, Kulla M, Luepke K, Morwinsky T, Blätzinger M, Rossmann K
Analysis of Digital Documentation Speed and Sequence Using Digital Paper and Pen Technology During the Refugee Crisis in Europe: Content Analysis
JMIR Mhealth Uhealth 2019;7(8):e13516
URL: http://mhealth.jmir.org/2019/8/e13516/
doi:10.2196/13516
PMID:31429420

Viewpoint

The Elusive Path Toward Measuring Health Outcomes: Lessons Learned From a Pseudo-Randomized Controlled Trial of a Large-Scale Mobile Health Initiative

Patricia Mechael[1*], PhD; Nadi Nina Kaonga[2,3*], MHS, MD; Subhashini Chandrasekharan[2*], PhD; Muthu Perumal Prakash[4], MSc; Joanne Peter[5], MD; Aakash Ganju[6*], MD; Nirmala Murthy[7*], DSC

[1]HealthEnabled, Washington, DC, United States
[2]All of Us Research Program, National Institutes of Health, Bethesda, MD, United States
[3]Tufts University School of Medicine, Boston, MA, United States
[4]Foundation for Research in Health Systems, Chennai, India
[5]Johnson & Johnson, Cape Town, South Africa
[6]Saati Health, Mumbai, India
[7]Foundation for Research in Health Systems, Bengaluru, India
[*]these authors contributed equally

Corresponding Author:
Patricia Mechael, PhD
HealthEnabled
4101 Albemarle St NW Apt 521
Washington, DC, 20016
United States
Phone: 1 9732228252
Email: pmechael@gmail.com

Abstract

Mobile health (mHealth) offers new opportunities to improve access to health services and health information. It also presents new challenges in evaluating its impact, particularly in linking the use of a technology intervention that aims to improve health behaviors with the health outcomes that are impacted by changed behaviors. The availability of data from a multitude of sources (paper-based and electronic) provides the conditions to facilitate making stronger connections between self-reported data and clinical outcomes. This commentary shares lessons and important considerations based on the experience of applying new research frameworks and incorporating maternal and child health records data into a pseudo-randomized controlled trial to evaluate the impact of mMitra, a stage-based voice messaging program to improve maternal, newborn, and child health outcomes in urban slums in India.

(JMIR Mhealth Uhealth 2019;7(8):e14668) doi:10.2196/14668

KEYWORDS

India; mHealth; maternal health; child health; research

Background

Over the past 10 years, there has been a rapid increase in the adoption and use of mobile technology by the health sector globally as a tool to increase access to health services and health information, strengthen health systems, improve the quality of care by health professionals, and increase efficiency in the delivery of health services. From the earliest days of mobile health (mHealth), the peer-reviewed literature has primarily focused on usability and feasibility of apps and has largely been published in the computer science literature [1,2]. Few, if any, studies focused on patient-level outcomes, especially in resource-poor settings in low- and middle-income countries [3]. Increasingly, mHealth has been embraced to strengthen delivery of maternal, neonatal, and child health (MNCH) services. There are numerous implementations globally; however, the evidence remains mixed.

Several systematic reviews highlight the lack of rigorous experimental and quasi-experimental trials; they also cite the lack of research measuring health outcomes as a key challenge to the advancement of the field of mHealth [4-9]. However, individual studies evaluating mHealth interventions have

demonstrated an increased uptake of proven home- and facility-based practices and MNCH services. For example, an external evaluation of the *Chipatala Cha Pa Foni* project in Malawi, using a pre- and posttest design, demonstrated increased use of home- and facility-based maternal health practices and home-based child health practices among women exposed to the mHealth intervention (a toll-free case management hotline and automated and personalized mobile messages) [10]. Similarly, Lund et al used cluster randomized controlled trials to assess the *Wired Mothers* program in Zanzibar and demonstrated a significant increase of skilled delivery attendance and completion of four antenatal care visits among women receiving the mHealth intervention [11,12]. Although the observed improvements in proven practices may positively impact maternal and child health outcomes, no maternal or child health outcomes were directly measured in these studies.

Therefore, in this paper, we will share our experience and lessons learned from the evaluation of a large-scale mHealth voice message service implementation in India, where *real-world* health outcomes were assessed. We will also briefly highlight new possibilities for policy makers, implementers, and researchers to leverage electronic data to move toward *real-time* evaluation of outcomes from mHealth and other health interventions.

The mMitra Program

The Mobile Alliance for Maternal Action (MAMA) was a 4-year public-private partnership focused on harnessing the power of mobile technology to send stage-based health messages to pregnant women and new mothers. There were country implementations of MAMA in Bangladesh and South Africa followed by India and Nigeria. Thus, the implementation and evaluation in India built on the program design and lessons from the earlier MAMA implementations and research studies.

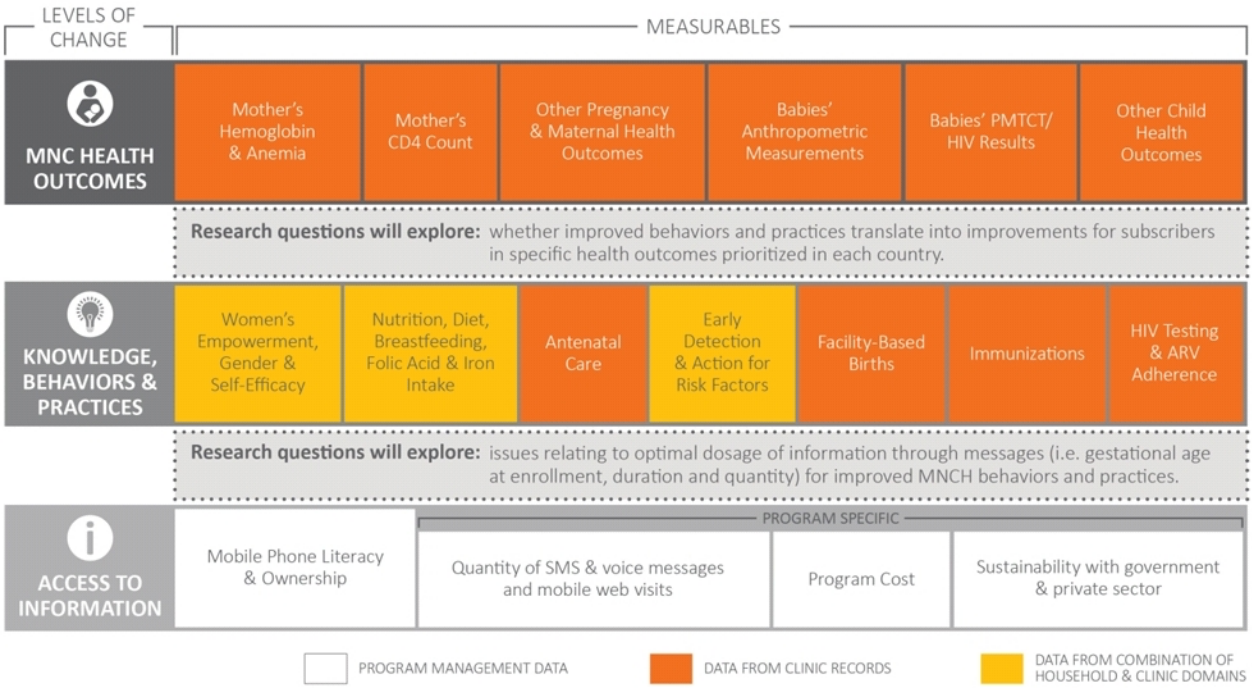
In 2015, the MAMA partnered with the nonprofit organization Advancing Reduction in Mortality and Morbidity of Mothers, Children and Neonates (ARMMAN) to implement a mobile phone–based voice messaging service in the urban slums of Mumbai, India, called mMitra [13]. mMitra aimed to leverage mobile phone technology to support improvements in self-care among pregnant women living in urban slums and infant care.

A total of 145 voice messages were developed. The voice was female, and there were 2 languages available—Hindi and Marathi—the languages commonly spoken in the urban slums of Mumbai. The message development process was rigorous and involved BabyCenter, ARMMAN, and representatives from the Federation of Obstetric and Gynecologic Societies of India and the Indian Academy of Pediatrics. They were field tested with local health experts and community focus groups. The messages were delivered as a prerecorded phone call and covered the period from 6 weeks of pregnancy to an infant's first birthday. A call-back service was integrated into the package for women to access within 2 days after the original call.

Designing the Study

A MAMA research agenda was developed to provide a standardized approach to evaluating the impact of the intervention in different countries [14]. A theory of change (see Figure 1) and targeted health outcomes for each of the countries were integrated into the agenda and helped inform the research design. All studies were designed to track and compare changes in key MNCH knowledge, attitudes, practices, and outcomes.

Figure 1. Mobile Alliance for Maternal Action theory of change and priority outcomes. ARV: antiretroviral; HIV: human immunodeficiency virus; MNC: maternal, newborn, and child; MNCH: maternal, newborn, and child health; PMTCT: prevention of mother to child transmission (of HIV); SMS: short message service.



The underlying hypothesized pathway to change for mMitra was that if women receive educational messages that are interesting, easy to understand, and aligned with their physiological state during pregnancy and post delivery, then they would be motivated to take the needed self-care and seek the needed health services. There was an assumption that women would also have access to other sources of similar information, such as community health workers who enroll them for pregnancy care; health care providers who treat them at health centers; and mass media messages on radio, television, and posters.

As the MAMA studies assessed behavior, it was important to collect data from real-world sources, triangulate the data as a method of data validation, and juxtapose the self-reported data with objective clinical data (see Table 1).

Table 1. Prioritized health and behavior outcomes for Mobile Alliance for Maternal Action programs in Bangladesh, South Africa, and India with data sources.

Prioritized Health Outcomes	Bangladesh	South Africa	India
MNCH[a] Biomarkers of Health Outcomes			
Hemoglobin levels/ anemia in mothers	—[b]	—	Clinic records
Babies' anthropometric measurements	—	—	Clinic records
CD4[c] count from mothers with HIV	—	Clinic records	—
WHO[d] stage of mothers with HIV	—	Clinic records	—
Tuberculosis status of mothers with HIV	—	Clinic records	—
Babies HIV test result	—	Clinic records	—
MNCH Behaviors, Practices, and Service Uptake			
Mother's nutrition/diet/folic acid/iron tablets	Self-report	Self-report	Self-report
Breast-feeding–exclusive and colostrum	Self-report	Self-report	Self-report
Antenatal care	Self-report	Self-report	Self-report
Gestational age at first ANC[e]	Self-report	Self-report	Self-report
Facility-based births	Self-report	Self-report	Self-report
Postnatal care	Self-report	Self-report	Self-report
HIV counseling and testing	Self-report	Self-report	Self-report
Antiretroviral therapy	Self-report	Self-report	Self-report
Immunizations	Self-report	Self-report	Self-report
Early detection and action for risk factors	Self-report	Self-report	Self-report
Empowerment/gender/self-efficacy	Self-report	Self-report	Self-report

[a]MNHC: maternal, neonatal, and child health.

[b]Not applicable.

[c]CD4: Cluster of differentiation 4.

[d]WHO: World Health Organization.

[e]ANC: antenatal care.

The lack of objective clinical data in the Bangladesh evaluation led to the collection of retrospective clinical data in South Africa [15,16] and an attempt to use maternal and child health record data from the public health system in India. The measurement of physiologic biomarkers, reflecting health outcomes, were included from the start of the implementation in India (see Figure 2). The biomarkers selected were maternal hemoglobin levels and the infant's birth weight. These reflected the country's public health priorities in particular based on the India District Level Health Survey conducted in 2013, which documented that 68.5% pregnant women among the urban poor in Maharashtra are anemic, babies with normal birthweight were 87%, and babies being normal or mildly undernourished at year 1 were 63%.

Given the evolving technology landscape (ie, more affordable technologies and emergence of new technologies), evaluation studies are often not relevant by the time full-scale implementation occurs [17]. There is also an ongoing debate on the appropriateness of randomized controlled trials as the gold standard for mHealth evaluation [5,18,19]. In addition, evaluating behavior change adds complexity, as it is difficult to attribute improvements in health behaviors to the intervention alone. To aid in the study design, and afterwards evaluation and reporting, we used the following resources, respectively:

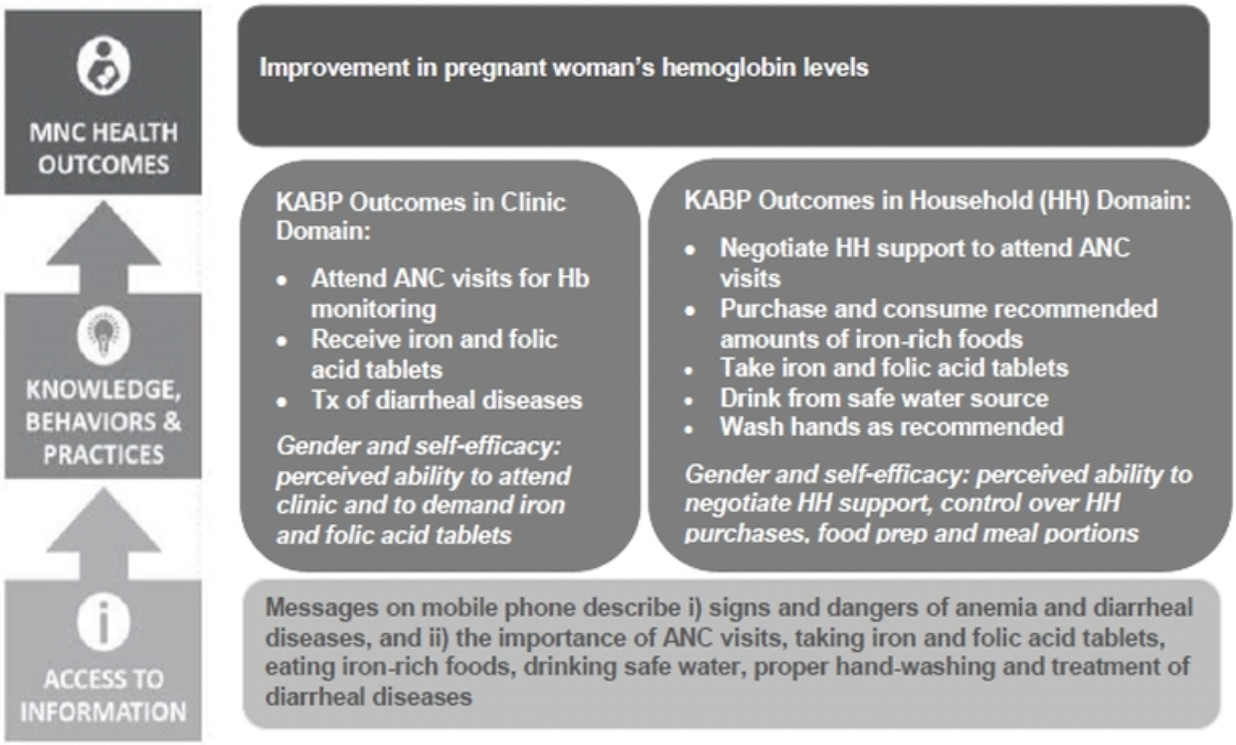
- Whittaker et al's detailed process on rigorous mHealth experimental evaluation approaches and types of measures that should be included [17].

- Roess' publication that outlines process evaluation and implementation science approaches that can be used to measure reach, fidelity of use, and the *dose* of an intervention [20].
- The mHealth evidence reporting and assessment checklist, which is often used to guide reporting on the effectiveness of digital health [21,22].

We ultimately decided to use a pseudo-randomized controlled trial design to assess the impact of mMitra on the desired outcomes [23]. The study design was chosen for its experimental nature, and it built on the lessons learnt from the prior MAMA implementations and evaluations. Importantly, the study design followed women, by trimester, allowing for a dose-response assessment.

All women interested in participating in the study and implementation provided written informed consent. The study was approved by the Foundation for Research in Health System's institutional review board (under protocol no. HHS00009235) and registered with ISRCTN (registration no. 88968111). Specially designed questionnaires were administered to women in the control and intervention groups at 3 time points: at enrollment in the study while the women were pregnant (baseline), after delivery, and when their baby reached 1 year of age. The surveys were conducted in-person with consent in Hindi and Marathi with the enumerator using an Android-based data collection platform. At the same in-person encounters, data were abstracted, with consent, from the woman's Maternal and Child Health (MCH) card. The MCH card is issued to women by hospitals where they register for prenatal care. It serves as their health record, and they are instructed to bring the card with them on every visit for staff to update. Women are incentivized to keep their cards as it determines their eligibility to receive government compensation after delivery. Of the women interviewed post delivery 1040/1113 (93.44%) in intervention and 381/402 (94.8%) in control group showed their MCH card to the enumerators.

Figure 2. Pathway to change for hemoglobin levels in mMitra. ANC: antenatal care; Hb: hemoglobin; HH: household; KABP: knowledge, attitudes, behavior, and practices; MNC: maternal, newborn, and child.



Lessons Learned

Several key lessons were learned through this process. Major insights were incorporated into the study design based on the previous evidence and approach to research of other mobile messaging programs, the literature related to mHealth research, and the cumulative experience in research across the initial 2 MAMA country programs. This led to the adoption of a prospective pseudo-randomized controlled trial that proactively measured dose response based on the duration of exposure to messages by trimester and integrated objective clinical data with self-reported information.

As we had assumed that women would be exposed to external sources of information that would influence their behavior and practices, we prioritized real-world study designs over pure experimental designs. This pragmatic approach allowed us to have a comparison group and ensure that we would appropriately attribute any potential impact of the intervention on behavior and health outcomes, if any. We were also able to ensure that the research aligned with India's health priorities by ensuring the study's health outcomes matched the country's priorities.

The triangulation of the data helped advance linkages between the intervention and clinical outcomes. Through this process, we learned that even though the health record (MCH card)

seemed like a reliable data source because of government incentives, it did not preclude missing and poor-quality data. This was particularly true for maternal health data but less so for child health data. Only 24% of MCH cards in the control group and 34% of cards in the intervention group had entries for maternal hemoglobin—a key maternal health outcome of interest; whereas 65% of cards in both groups had anthropometric child health data that could be used to evaluate key child health outcomes. However, the data from the MCH cards were insufficient, and findings detected from the subsample of MCH cards with complete data were not statistically significant. This impacted our ability to use the data from the MCH cards to assess clinical health outcomes, but it provided an authentic, real-world perspective. It is acknowledged that the issue of missing data may vary by data element as well as by data source as well as use case and context.

From the outset of the study, we kept track of the duration of exposure women had to the messages. This was important to track to determine if a dose response was present. However, we also recognized that greater confidence could be placed in the message exposure findings by having both self-reported and electronic data sources. Subsequently, this brought attention to a future consideration—the inclusion of *data exhaust* generated by the messaging platform and linking this type of message data with research data for a direct one-to-one correlation between messages and effects. Data exhaust are trails of data left by users. Examples of data exhaust include the success rate of phone calls being delivered and accessed, the number of missed calls by message type and timing, and the numbers of call-backs accessed. Data exhaust can be used to aid with determining message access and user trends; it can be useful for ensuring completeness and accuracy of research data without additional human effort for data quality assessments. They are also generated in *real-time*, which can help shift programming toward more timely decision making.

With the increase in the digitization and use of decision-support and community-case management tools by frontline health workers, there are opportunities to further leverage existing data collection processes to validate and supplement self-reported and health record data. Furthermore, future assessments on the nature of the missing data may provide useful insights into how it can be addressed through statistical methods or through efforts made upstream to the data collection (eg, frontline health workers). As research moves toward real-world evidence and data generated through digital systems, it is important for the initial design to abide by data ownership and governance regulations, including mechanisms for obtaining consent for the use of routine health data to be used for research purposes as well as promoting data completeness and data quality. A rigorous focus on these aspects will allow evaluations to harness data more effectively and better lend themselves to high-quality research efforts that generate sound evidence.

Conclusions

The MAMA research experience, especially that of mMitra in India, provides important considerations for how to approach health outcomes research in evaluating the impact of an mHealth intervention. Having a clear understanding of the context and local health priorities is important to ensure that the research agenda is in alignment with local and national priorities and, thus, can have greater relevance to policy makers, program implementers, and beneficiaries. The inclusion of biomarkers and use of clinical records were an attempt to complement self-reported information on health outcomes. These opportunities are increasing as more health systems increase adoption of electronic medical records and shared health records. However, as our experience highlights, real-world data has its limitations—particularly in the areas of data completeness and quality. There is also an opportunity to use the data exhaust and meta-data generated from digital tools as an additional data source. Further study is recommended to assess the impact of missing data and the quality and viability of these types of records for future use in health outcomes–focused mHealth research. As more technology is adopted by health systems, the ability to move toward real-world evidence in the evaluation of digital health interventions and health interventions alike will increase—but great care and adoption of ethical practices will be required in the way that data are captured, structured, managed, analyzed, and shared.

Acknowledgments

The authors would like to thank Aparna Hegde of ARMMAN India for her support in providing access to the mMitra messages during the study and answering questions about the service throughout the research and writing process. They would also like to thank BabyCenter for information on the process of developing, reviewing, and testing the mMitra mobile messaging content. This study was funded by Johnson & Johnson. The funders contributed to critical review of the manuscript and decision to submit the paper for publication.

AG is employed by with Saati Health, Mumbai, India, although contributions to this work though were done in his personal capacity. SC is currently the Ethical Legal Social Issues lead in the *All of Us* Research Program at the National Institutes of Health (NIH), Bethesda, MD, United States. This work does not represent the views of of the *All of Us* Research Program or the NIH.

Authors' Contributions

PNM provided overall oversight and contributed to manuscript writing. NM provided scientific and study oversight, helped design and analyze the study, and assisted in writing the initial draft of the manuscript. SC and NNK contributed to writing final versions of the manuscript. MPP performed all data analyses and contributed to data interpretation. AG and JP contributed to overall study

design, data interpretation, and research concept. All authors critically reviewed the manuscript and approved the final version for publication.

Conflicts of Interest

JP is an employee of Johnson & Johnson, one of the funders of mMitra. No other competing interests are declared.

References

1. Mechael PN, Batavia H, Kaonga N, Searle S, Kwan A, Goldberger A, et al. Global Problems, Global Solutions. 2010. Barriers and Gaps Affecting mHealth in Low and Middle Income Countries: Policy White Paper URL: http://www.globalproblems-globalsolutions-files.org/pdfs/mHealth_Barriers_White_Paper.pdf [WebCite Cache ID 780S2oBl5]
2. Mechael P, Nemser B, Cosmaciuc R, Cole-Lewis H, Ohemeng-Dapaah S, Dusabe S, et al. Capitalizing on the characteristics of mhealth to evaluate its impact. J Health Commun 2012;17(Suppl 1):62-66. [doi: 10.1080/10810730.2012.679847] [Medline: 22548600]
3. Piette JD, Lun KC, Moura Jr LA, Fraser HS, Mechael PN, Powell J, et al. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? Bull World Health Organ 2012 May 1;90(5):365-372 [FREE Full text] [doi: 10.2471/BLT.11.099069] [Medline: 22589570]
4. Tamrat T, Kachnowski S. Special delivery: an analysis of mhealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012 Jul;16(5):1092-1101. [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
5. Gurman TA, Rubin SE, Roess AA. Effectiveness of mhealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17(Suppl 1):82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
6. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
7. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
8. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
9. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mhealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
10. Fotso JC, Robinson AL, Noordam C, Crawford J. Fostering the use of quasi-experimental designs for evaluating public health interventions: insights from an mhealth project in Malawi. Afr Popul Stud 2015 Jun 4;29(1):1607 [FREE Full text] [doi: 10.11564/29-1-713]
11. Lund S, Hemed M, Nielsen BB, Said A, Said K, Makungu MH, et al. Mobile phones as a health communication tool to improve skilled attendance at delivery in Zanzibar: a cluster-randomised controlled trial. BJOG 2012 Sep;119(10):1256-1264 [FREE Full text] [doi: 10.1111/j.1471-0528.2012.03413.x] [Medline: 22805598]
12. Lund S, Nielsen BB, Hemed M, Boas IM, Said A, Said K, et al. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. BMC Pregnancy Childbirth 2014 Jan 17;14:29 [FREE Full text] [doi: 10.1186/1471-2393-14-29] [Medline: 24438517]
13. ARMANN India. mMitra URL: http://armman.org/mmitra/ [accessed 2019-04-29] [WebCite Cache ID 780QqTzAX]
14. mHealth Knowledge. 2015. MAMA: Mobile Alliance for Maternal Action: Research Agenda URL: https://www.mhealthknowledge.org/sites/default/files/MAMA%20Research%20Agenda_FINAL_March%202015_1.pdf [WebCite Cache ID 780Tgezy1]
15. mHealth Knowledge. 2013. MAMA: Mobile Alliance for Maternal Action: Formative Research Report URL: http://www.mhealthknowledge.org/sites/default/files/MAMABangladeshFormativeResearchReport.pdf%20.%20Archived%20at:%20http://www.webcitation.org/780TsytRD [accessed 2019-04-29] [WebCite Cache ID 780TsytRD]
16. Coleman J, Bohlin KC, Thorson A, Black V, Mechael P, Mangxaba J, et al. Effectiveness of an SMS-based maternal mhealth intervention to improve clinical outcomes of HIV-positive pregnant women. AIDS Care 2017 Dec;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
17. 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]
18. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mhealth. J Health Commun 2012;17(Suppl 1):5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
19. Mookherji S, Mehl G, Kaonga N, Mechael P. Unmet need: improving mhealth evaluation rigor to build the evidence base. J Health Commun 2015;20(10):1224-1229. [doi: 10.1080/10810730.2015.1018624] [Medline: 26042660]

20. Roess A. The promise, growth, and reality of mobile health - another data-free zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]
21. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
22. Agarwal S, Lefevre AE, Labrique AB. A call to digital health practitioners: new guidelines can help improve the quality of digital health evidence. JMIR Mhealth Uhealth 2017 Oct 6;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]
23. Cochrane Consumer Network. Levels of Evidence URL: https://consumers.cochrane.org/levels-evidence [accessed 2019-04-29] [WebCite Cache ID 780RUATw9]

Abbreviations

ARMMAN: Advancing Reduction in Mortality and Morbidity of Mothers, Children and Neonates
MAMA: Mobile Alliance for Maternal Action
MCH: maternal and child health
mHealth: mobile health
MNCH: maternal, neonatal, and child health

Edited by G Eysenbach; submitted 09.05.19; peer-reviewed by C Fincham, K Agarwal; comments to author 14.06.19; revised version received 21.06.19; accepted 27.06.19; published 21.08.19.

Please cite as:
Mechael P, Kaonga NN, Chandrasekharan S, Prakash MP, Peter J, Ganju A, Murthy N
The Elusive Path Toward Measuring Health Outcomes: Lessons Learned From a Pseudo-Randomized Controlled Trial of a Large-Scale Mobile Health Initiative
JMIR Mhealth Uhealth 2019;7(8):e14668
URL: http://mhealth.jmir.org/2019/8/e14668/
doi:10.2196/14668
PMID:31436165

Original Paper

An App-Based Intervention for Caregivers to Prevent Unintentional Injury Among Preschoolers: Cluster Randomized Controlled Trial

Peishan Ning[1], MD; Peixia Cheng[1], BA; David C Schwebel[2], PhD; Yang Yang[3], PhD; Renhe Yu[1], BA; Jing Deng[1], PhD; Shukun Li[4], MD; Guoqing Hu[1], PhD

[1]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

[2]Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Department of Biostatistics, College of Public Health and Health Professions, Emerging Pathogen Institute, University of Florida, Gainesville, FL, United States

[4]Information and Network Center, Central South University, Changsha, China

Corresponding Author:
Guoqing Hu, PhD
Department of Epidemiology and Health Statistics
Xiangya School of Public Health
Central South University
110 Xiangya Road
Changsha, 410078
China
Phone: 86 731 84805414
Fax: 86 731 84805414
Email: huguoqing009@gmail.com

Abstract

Background: App-based interventions have the potential to reduce child injury in countries with limited prevention resources, but their effectiveness has not been rigorously examined.

Objective: This study aimed to assess the effectiveness of an app-based intervention for caregivers of preschoolers to prevent unintentional injury among Chinese preschoolers.

Methods: A 6-month cluster randomized controlled trial was conducted from December 2017 to June 2018. Recruitment was conducted through preschools, which were randomly allocated to either the control group (ie, app-based parenting education excluding unintentional injury prevention) or the intervention group (ie, app-based parenting education including unintentional injury prevention). A total of 2920 caregivers of preschoolers aged 3-6 years from 20 preschools in Changsha, China, were recruited offline through the schools. The primary outcome was unintentional injury incidences among preschoolers in the past 3 months; this measure was assessed through an online caregiver-report at the baseline visit and at 3-month and 6-month follow-up visits. Secondary outcome measures included caregivers' self-reported attitudes and behaviors concerning child supervision during the last week. Generalized estimating equations (GEEs) were used to assess the effectiveness of the app-based intervention on responses at 3 and 6 months after adjusting for sociodemographic variables, baseline level of the outcome variable, and engagement with interventions in the assigned group. All analyses were intention-to-treat. A per-protocol sensitivity analysis was also conducted.

Results: In total, 1980 of the 2920 caregivers completed the study. The mean age of participants was 32.0 years (SD 5.5) and 68.99% (1366/1980) of them were female. During the 6-month follow-up visit, unintentional injury incidence did not change significantly in either group: incidence in the intervention group went from 8.76% (94/1073) to 8.11% (87/1073), *P*=.59; incidence in the control group went from 9.4% (85/907) to 7.5% (69/907), *P*=.15. The changes did not differ between the groups (odds ratio [OR] 1.14, 95% CI 0.80-1.62). Changes in the average score in attitude concerning unintentional injury prevention were also similar between the groups (B .05, 95% CI -0.03 to 0.13). Changes in unintentional injury prevention behaviors were greater in the intervention group than in the control group after the intervention (B .87, 95% CI 0.33-1.42). Analyses of individual injury prevention behaviors showed that the intervention reduced three risky behaviors: unsafe feeding of children (OR 0.73, 95% CI 0.60-0.89); incorrectly placing children in cars (OR 0.73, 95% CI 0.57-0.93); and allowing children to ride bicycles, electric bicycles, or motorcycles unsupervised (OR 0.80, 95% CI 0.64-0.99). The intervention also improved scores on three safety-focused

behaviors: testing water temperature before giving children a bath (OR 1.26, 95% CI 1.05-1.52); properly storing sharp objects (OR 1.24, 95% CI 1.01-1.52); and safely storing medicines, detergents, and pesticides (OR 1.24, 95% CI 1.02-1.51).

Conclusions: The app-based intervention did not reduce unintentional injury incidence among preschoolers but significantly improved caregivers' safety behaviors. This app-based intervention approach to improve caregiver behaviors surrounding child injury risk offers promise to be modified and ultimately disseminated broadly.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOR-17010438; http://www.chictr.org.cn/showproj.aspx?proj=17376 (Archived by WebCite at http://www.webcitation.org/75jt17X84)

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5790-1

(JMIR Mhealth Uhealth 2019;7(8):e13519) doi:10.2196/13519

KEYWORDS

unintentional injury; preschoolers; cluster randomized controlled trial; app; mobile health; intervention

Introduction

Unintentional injuries are a major public health threat to children worldwide. In 2017, it is estimated that over 191,000 children under 5 years of age died from unintentional injuries, with 79% of deaths occurring in low- and middle-income countries (LMICs) [1]. Moreover, nonfatal childhood injuries lead to substantial economic burden and long-term adverse consequences, including physical disability, cognitive or social impairment, and lower educational achievement [2,3].

Previous studies [4-6] indicate that lack of safe parenting behavior, inadequate caregiver perception of risks for child injury, and low adoption of safety equipment usage contribute to the occurrence of child unintentional injury. Parenting interventions to promote child safety have proven effective in high-income countries (HICs) [7] but are not commonly implemented in LMICs like China [8,9].

Mobile health (mHealth)-based interventions offer an opportunity to deliver parenting interventions broadly and cost-effectively, as the usage of mobile phones with expanded and advanced functions (ie, smartphones) rapidly becomes commonplace worldwide. A recent review by Omaki et al [10] provides evidence for the effectiveness of technology-based interventions to improve unintentional injury-prevention behaviors. To date, however, only a few mobile phone app interventions are available to assist parents in preventing unintentional child injuries, and most are not based in theory or are insufficiently tested in rigorous randomized trials [11]. Tests of these interventions were conducted in HICs [12-15] and generally involved assessments of knowledge, perception, and behavioral outcomes with relatively small sample sizes [12,14] (eg, the sample sizes in studies by Gielen et al [12] and Burgess et al [14] were 498 and 742, respectively). None have used actual injury events as the primary outcome indicator [12-15].

This cluster randomized controlled trial (RCT) study aims to evaluate the 6-month effectiveness of a theory-driven, app-based mobile phone intervention for caregivers in preventing unintentional injury among Chinese preschoolers. The study also evaluated whether the intervention improved caregivers' attitudes concerning injury and their behaviors to promote child safety.

Methods

Study Design

A single-blinded, cluster RCT with 1:1 allocation ratio and a follow-up period of 6 months was conducted from December 2017 to June 2018 in Changsha, China. We chose cluster randomization to avoid contamination within the same preschools. The protocol (see Multimedia Appendix 1) was approved by the Ethics Committee of Xiangya School of Public Health, Central South University, Changsha, China (approval number: XYGW-2017-02) and has been published elsewhere [16]. All participants provided informed consent online. This report follows the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement: extension to cluster randomized trials (see Multimedia Appendix 2) [17].

Participant Recruitment

The recruitment of study participants was based on preschool children aged 3-6 years. We limited the study to preschools with at least 100 students to increase recruitment efficiency. Eligible preschools were randomly selected and contacted by researchers with an official invitation letter along with information about the project. In total, 28 eligible preschools were contacted and 20 agreed to participate (71% participation rate).

All primary caregivers who owned mobile phones and had a preschooler who was 3-6 years old and who was enrolled in a participating preschool were eligible for the study. Primary caregivers were defined as parents, grandparents, other family members, friends of the family, or babysitters and nannies who served as primary caregivers of preschoolers [18]. Teachers from the schools were excluded. One teacher was recruited at each participating preschool to inform eligible caregivers about the study via existing school-family communication channels, including social media platforms (ie, WeChat and QQ), school apps, printed handouts, and oral notification.

Caregivers who agreed to participate received introductory materials about the project. Upon downloading the app, which was developed by the research team and named Bao Hu San (ie, protective umbrella), caregivers viewed and completed online informed consent. Consenting participants then completed an online baseline survey addressing demographic characteristics of caregivers and their children, attitudes toward child injury

prevention, supervision behaviors in the last week, and information about any unintentional injury that had occurred to their child during the prior 3 months.

Sample Size

To obtain adequate power, we calculated the sample size based on the following criteria: baseline unintentional injury incidence of 23% among preschoolers in the past 3 months, effect size (ie, incidence rate ratio) of 0.75 between intervention and control groups, cluster size of 140 children per preschool, and intraclass correlation (ICC) of .005 [19]. Under these criteria, a total of 2626 participants in 20 schools—10 schools per arm—would achieve power of 80% at a .05 significance level, assuming 10% loss to follow-up.

Randomization Scheme

To avoid potential confounding from the type of preschool, we stratified randomization by type of preschools to reach five public and five private schools per arm. Randomization was performed by an independent (ie, masked) researcher using SAS 9.2 software (SAS Institute).

Control Group

The control group completed Bao Hu San, an app-based parenting education program developed by the research team. The app trained parents concerning pediatric disease risks and parenting skills but excluded explicit information about unintentional child injury prevention.

Intervention Group

The intervention group received all content that the control group received using the same app, Bao Hu San, but were also exposed to additional researcher-developed components that focused specifically on unintentional child injury prevention.

App Design

The app components consisted of four active modules: (1) content learning, including lessons to teach caregivers basic knowledge about parenting skills through short written statements with pictures, cartoon vignettes, video testimonials, and interactive games; (2) interaction, containing three submodules to support communication among users (ie, study participants) and between users and professionals; (3) survey and feedback, namely the questionnaire module to collect data online; and (4) personal modules, allowing participants to select the color of the interface in their app according to their preferences (eg, pink, yellow, and blue) (see Figure 1).

The four active modules were available for both the intervention and control groups and differed in two ways. First, the content learning module provided unintentional injury prevention knowledge for the intervention group, in addition to providing pediatric disease risk and parenting skills content available to both groups. Second, communication in the interaction module with injury prevention professionals was only available to the intervention group.

These injury prevention modules were developed based on the principles of the Theory of Planned Behavior (TPB), the Haddon Matrix, and the Framework for the Rational Analysis of Mobile Education (FRAME) model [20-22]. The TPB interventions were designed to target one or more determinants of behavior: attitude, subjective norms, or perceptions of behavioral control. According to health behavior change theory [20], changes in attitudes, subjective norms, and/or perceptions of behavioral control should lead to changes in behavioral intentions, offer adequate perceived control over the behavior, and ultimately lead to behavior change.

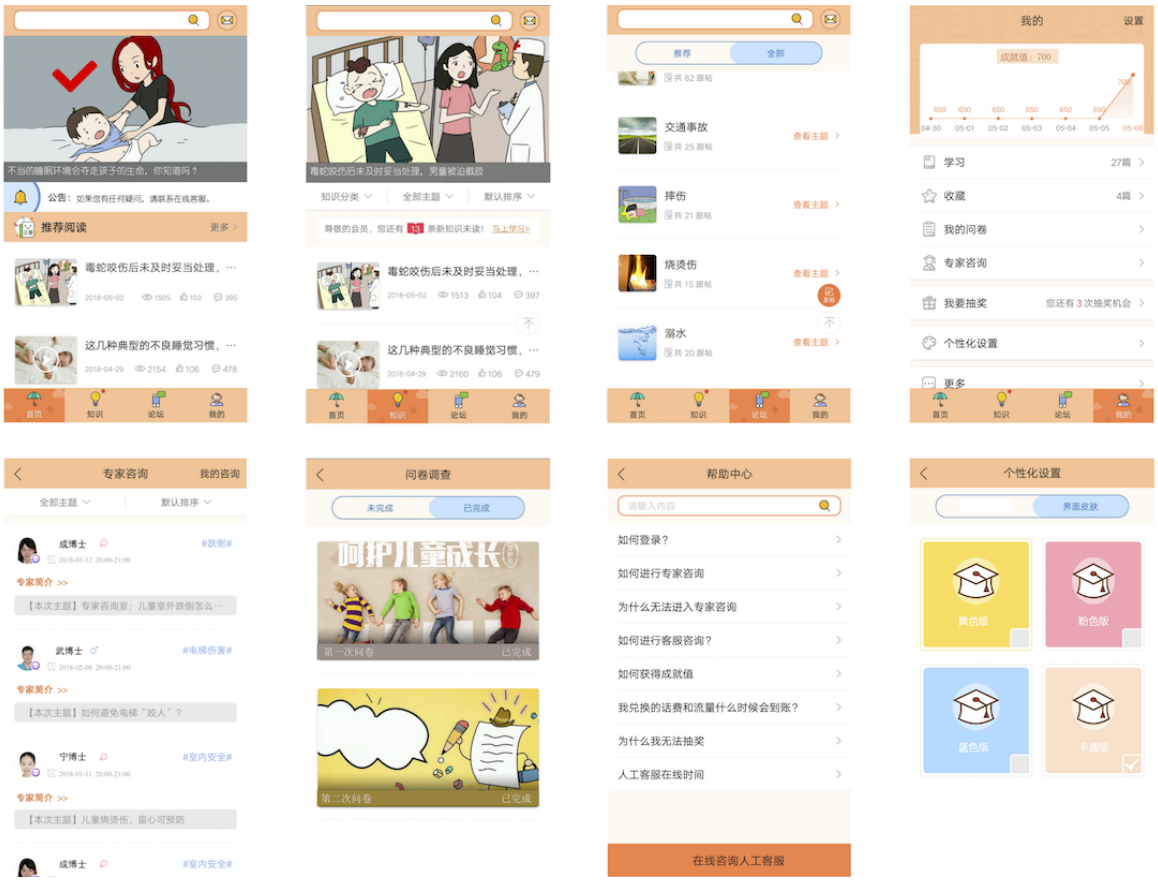
The Haddon Matrix is a classical theoretical framework to describe the occurrence of an injury event from three phases (ie, pre-event, event, and post-event) and four factors (ie, host, agent or vehicle, physical environment, and social environment) [21].

As a comprehensive model to develop mobile learning apps, the FRAME model allows researchers to consider all relevant components of users' learning in the development of the app [22]. Among the components that receive focus are usability of the device and app, capacities of the learners, and social interaction between users. We followed the FRAME model when developing the app and its multiple components; we implemented the FRAME model through results from a needs assessment that consisted of a series of focus groups plus online surveys among key stakeholders, including preschool teachers and caregivers of preschoolers [23].

We implemented several strategies to encourage participants in both the intervention and control groups to actively use the app: (1) awarding of app-based virtual currency rewards, (2) awarding of additional rewards by lottery for participants who continuously logged in to the app for 7 consecutive days, (3) offering monthly rewards to the caregivers of the preschool classes with the most frequent log-ins, and (4) sending of regular reminders about app engagement from preschool teachers. Detailed descriptions about the intervention are available elsewhere [16].

Routine parenting educational activities offered by the preschool or other institutions (eg, children's hospitals and communities), whether related to unintentional injury prevention or not, were permitted in both groups. All preschools involved in the study received honoraria at months 1, 3, and 6.

Figure 1. Home page of the app intervention. The app home page as translated into English appears in Figure MA3-1 in Multimedia Appendix 3. Eight images within the figure were derived from the app Bao Hu San, which was developed by the research team for unintentional injury prevention among preschool students and was tested in this trial.



Outcome Measurements

Primary Outcome

The primary outcome was unintentional injury incidence among preschoolers in the prior 3 months, as collected at the baseline visit and at 3-month and 6-month follow-up visits through online surveys. Unintentional injuries were defined according to the International Classification of Diseases, 10th Revision [24]. Similar to previous studies [25], injury events were defined as incidents meeting any of the following criteria: (1) child receives medical treatment by a doctor or other medical professional following an injury; (2) child receives first aid by a family member, teacher, or other nonmedical staff following an injury (eg, medication, massage, or hot compress); and/or (3) child is restricted from school or other activities, or is kept in bed or rest for more than a half day following an injury.

Unintentional injury incidence in the prior 3 months was calculated as follows:

> (Number of preschoolers experiencing an unintentional injury in the prior 3 months) / (Total number of children supervised by recruited caregivers) × 100%

Secondary Outcomes

Secondary outcomes, as outlined a priori in the study protocol [16], included the following: (1) caregivers' attitudes toward unintentional injury prevention and safety behaviors among preschoolers; (2) economic losses due to unintentional injury, including direct economic costs (eg, medical treatment expenses) and indirect economic costs (eg, caregiver's economic loss from being off work); and (3) the incremental cost-effectiveness ratio (ICER) for the app-based unintentional injury intervention, calculated as the cost difference between the intervention group and the control group divided by the difference in the number of unintentional injury events among children in the two arms. We did not analyze economic losses because some respondents could not remember the amount of medical expenses and indirect costs due to child unintentional injury, and others viewed medical expenses as sensitive private information they were not willing to report. We also excluded the ICER from our final analysis because changes in unintentional injury incidence were not statistically significant.

Both caregiver attitudes toward child injury prevention and frequency of engaging in injury prevention behaviors served as secondary outcomes. Each was measured via self-report using a 4-point scale: attitudes were reported as *completely agree*, *partly agree*, *not sure*, or *not at all agree*; behavior was reported as *0 times in the past week*, *1-2 times in the past week*, *3-5 times in the past week*, or *≥6 times in the past week*. The total attitude score ranged from 2 to 8 points. The four categories were then quantified as 4, 3, 2, and 1 for attitudes and risky behaviors and as 1, 2, 3, and 4 for safe behaviors, so that higher scores in both cases reflected greater awareness of, or health-promoting behavior toward, injury prevention. The total behavior score ranged from 15 to 60 points.

Textbox 1. Items assessing self-reported caregiver attitudes concerning child injury and health behaviors to prevent unintentional child injury.

Two items assessing attitudes:

1. Preventability of preschooler unintentional injury
2. Self-efficacy to keep child safe from unintentional injuries

15 items (nine for risky behaviors and six for safe behaviors):

Risky behaviors:

1. Leaving child alone in the home
2. Leaving child alone in the bathroom while bathing
3. Criticizing child when they are eating or drinking, creating a choking or suffocation risk
4. Giving child whole or large pieces of food that create choking risk
5. Placing child in the front seat while riding in a car
6. Not using child restraints while riding in a car
7. Letting child ride a bicycle, electric bicycle, or motorcycle unsupervised
8. Letting child take an escalator alone
9. Letting child contact unfamiliar or aggressive animals

Safe behaviors:

1. Holding child's hand while crossing the street
2. Testing water temperature before giving child a bath
3. Placing hot substances and lighters where children cannot reach
4. Placing sharp objects where children cannot reach
5. Storing medicines, detergents, and pesticides where children cannot reach
6. Wearing safety equipment when child rides a bicycle, electric bicycle, or motorcycle

The caregiver's overall attitudes and behaviors concerning unintentional injury prevention were calculated as the sum of two and 15 items, respectively (see Textbox 1). All items were adapted from previous studies and have demonstrated psychometric reliability and validity [26,27]. Prior to use in our study, we conducted a brief validation study of the instrument after translating it into Chinese. Among a sample of 100 caregivers not enrolled in the larger study, test-retest reliability over a 1-week interval for the sum score of the two items assessing attitudes and the 15 items assessing behaviors was r_s=0.69 and r_s=0.64, respectively.

Post Hoc Outcomes

We define the incidence of each risky or safe behavior listed in Textbox 1 as follows:

> (Number of caregivers reporting a specific risky or safe behavior in the past week) / (Total number of participating caregivers) × 100%

This allows us to focus on the presence or absence of each of the risky and safe behaviors. In total, 15 item scores were calculated, corresponding to the 15 risky and safe behaviors listed in Textbox 1. We also considered engagement-level data. These data were recorded automatically through electronic strategies embedded in the app and included the number of log-ins, the length of time using the app at each log-in, the number of knowledge segments (eg, short written statements with pictures, cartoon vignettes, video testimonials, and interactive games) studied and bookmarked, and the number of posted comments.

Statistical Analysis

The chi-square test and the Wilcoxon rank-sum test were used to examine differences in demographic characteristics between the two trial groups. The generalized estimating equation (GEE) was used to assess the effectiveness of the app-based intervention on responses at 3 and 6 months after adjusting for sociodemographic variables, being taught about injury prevention in the past 3 months, frequency of using parenting apps, baseline level of the outcome variable, and the level of engagement with the app in the assigned group. A logistic link function was used for dichotomous outcomes and identity link was used for continuous outcomes (see the Statistical Models Section in Multimedia Appendix 3). All analyses were intention-to-treat. Missing values were imputed using the expectation-maximization algorithm before the GEE analysis. To test the robustness of the results, a per-protocol sensitivity analysis was conducted.

For the evaluation of parenting-related attitudes concerning, and behaviors to prevent unintentional child injury, we first examined mean differences in total scores between the two groups. Incidence differences in individual behaviors were examined if the total scores differed significantly between groups.

Group assignment was masked during data analysis. All statistical analyses were performed using SAS 9.2 (SAS Institute). All statistical tests were 2-sided tests at the .05 significance level.

Results

Baseline Characteristics of Participants

In total, 1980 of the consented 2920 caregivers (67.81%) completed the surveys at baseline and at both 3- and 6-month follow-up visits. Of the 1980 who completed all surveys, 1073 (54.19%) were in the intervention group and 907 (45.81%) were in the control group (see Figure 2). Caregivers in the intervention group were significantly younger (32.9 vs 33.6 years, P=.01), more often male than female (371/1073, 34.58% vs 243/907, 26.8%, P<.001), had different household income levels (P=.004), and were more likely to have been taught about injury prevention in the past 3 months (677/1073, 63.09% vs 491/907, 54.1%, P<.001) compared to caregivers in the control group (see Table 1). Other between-group demographic differences were not significant. The imbalanced baseline characteristics were adjusted for in all subsequent analyses of intervention effectiveness.

Figure 2. Flow diagram of the study.

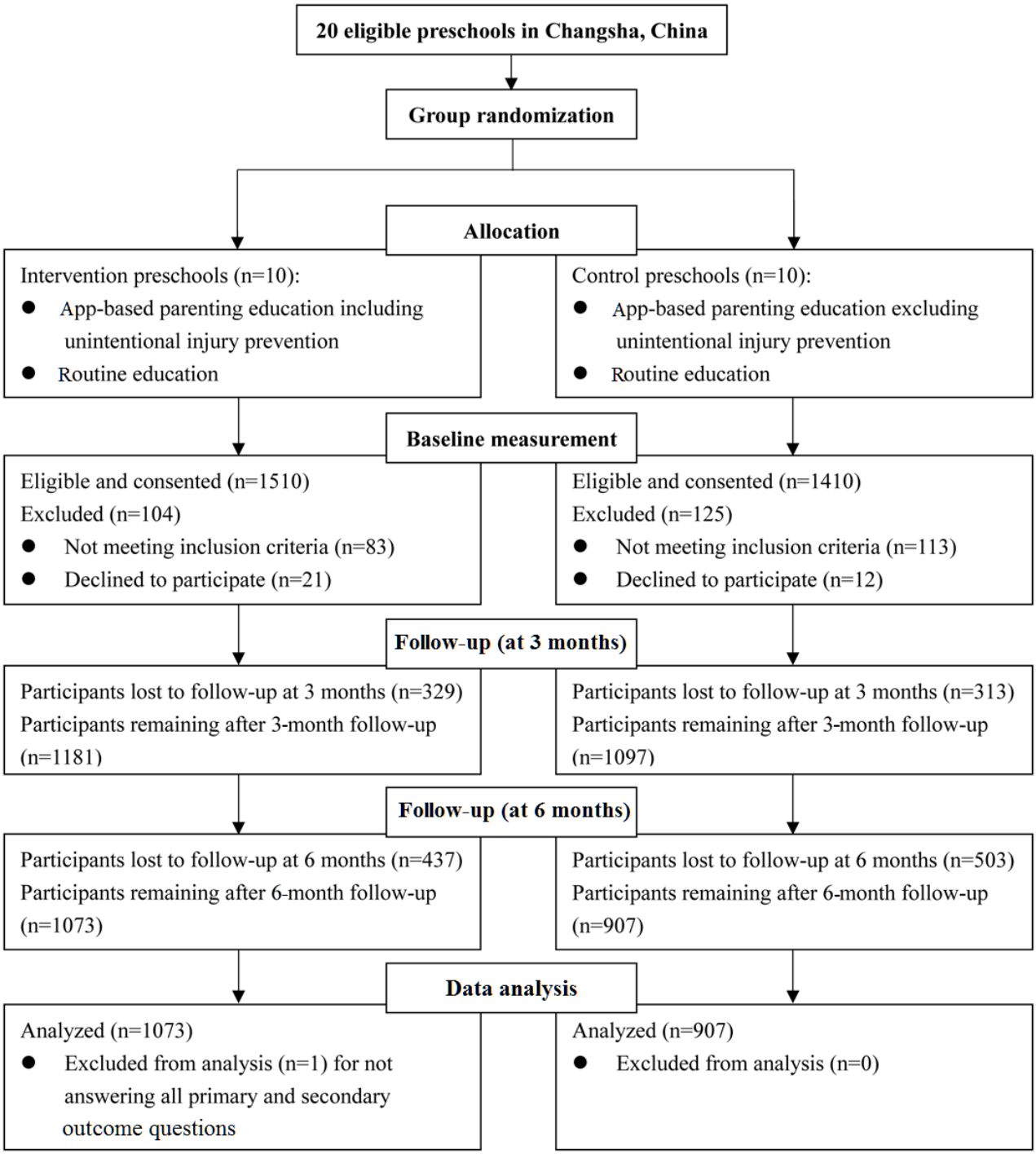


Table 1. Baseline characteristics of study participants who completed baseline surveys and both 3-month and 6-month follow-up surveys.

Characteristics	Total (N=1980)	Intervention group (N=1073)	Control group (N=907)	*P* value
Adult age in years, mean (SD)	32.0 (5.5)	32.9 (5.0)	33.6 (6.0)	.01
Adult gender, n (%)				
Male	614 (31.0)	371 (34.6)	243 (26.8)	<.001
Female	1366 (69.0)	702 (65.4)	664 (73.2)	
Child age in years, mean (SD)	4.5 (0.9)	4.5 (1.0)	4.5 (0.9)	.68
Child gender, n (%)				
Male	1013 (51.2)	536 (50.0)	477 (52.6)	.24
Female	967 (48.8)	537 (50.0)	430 (47.4)	
Adult education, n (%)				
Junior high school or below	118 (6.0)	64 (6.0)	54 (6.0)	.93
High school	475 (24.0)	261 (24.3)	214 (23.6)	
College and above	1387 (70.1)	748 (69.7)	639 (70.5)	
Household income per capita per month in Yuan, n (%)				
<1500	51 (2.6)	37 (3.4)	14 (1.5)	.004
1500-3499	366 (18.5)	213 (19.9)	153 (16.9)	
3500-5499	1339 (67.6)	634 (59.1)	705 (77.7)	
≥5500	224 (11.3)	189 (17.6)	35 (3.9)	
Frequency of using parenting apps, n (%)				
More than once a day	170 (8.6)	96 (8.9)	74 (8.2)	.53
Every two or three days	224 (11.3)	127 (11.8)	97 (10.7)	
Once a week	253 (12.8)	143 (13.3)	110 (12.1)	
Every two weeks or less	1333 (67.3)	707 (65.9)	626 (69.0)	
Taught about injury prevention in the past 3 months, n (%)				
Yes	1168 (59.0)	667 (63.1)	491 (54.1)	<.001
No	812 (41.0)	396 (36.9)	416 (45.9)	

Participants who did not complete the study (n=940) were similar to those who completed the study on several baseline characteristics: mean child age, *P*=.54; child gender, *P*=.83; household income, *P*=.50; and frequency of using parenting apps, *P*=.65. However, the caregivers who withdrew were significantly younger, more often female, had lower education levels, and were previously taught about injury prevention less often than the completers (see Table MA3-1 in Multimedia Appendix 3).

Intervention Engagement

Over the 6-month study period, the mean number of log-ins for all participating caregivers was 37.7 (SD 66.7) times, with an average of 39.3 (SD 42.5) knowledge segments learned, 1.8 (SD 7.3) knowledge segments bookmarked, and 29.9 (SD 67.1) comments posted. Overall log-in time duration using the app in the study period averaged 150.5 minutes (SD 239.7) (see Table MA3-2 in Multimedia Appendix 3). The intervention group on average had a higher number of knowledge segments studied (45.0 vs 32.6, *P*=.001) and a longer overall duration of app usage (161.2 vs 137.8 minutes, *P*=.03) than the control group, but no differences were found between the groups for other engagement indicators.

Unintentional Injury Incidence

After the 6-month intervention, unintentional injury incidence in the past 3 months decreased from 8.76% (94/1073) at baseline to 8.11% (87/1073) in the intervention group (*P*=.59), and decreased from 9.4% (85/907) to 7.5% (69/907) in the control group (*P*=.15) (see Table 2). The changes were not statistically significant between groups after adjusting for covariates: sociodemographic variables, outcome measures at baseline, and engagement to the intervention (odds ratio [OR] 1.14, 95% CI 0.80-1.62).

Table 2. Results for primary and secondary outcomes based on generalized estimating equations (GEEs).

Outcome measure	Intervention group (N=1073)	Control group (N=907)	Adjusted OR[a] or B[b] (95% CI)[c]	*P* value
Unintentional injury incidence, % (95% CI)				
Baseline	8.8 (7.0-10.6)	9.4 (7.4-11.4)	__[d]	—
3-month	7.7 (6.1-9.3)	7.1 (5.3-8.9)	1.16 (0.82 to 1.63)[a]	.41
6-month	8.1 (6.5-9.7)	7.5 (5.7-9.3)	1.14 (0.80 to 1.62)[a]	.47
Score of attitudes toward unintentional child injury prevention, mean (95% CI)				
Baseline	6.6 (6.6-6.7)	6.6 (6.5-6.6)	—	—
3-month	6.6 (6.6-6.7)	6.6 (6.6-6.7)	0.03 (-0.06 to 0.13)[b]	.48
6-month	6.8 (6.7-6.8)	6.7 (6.6-6.7)	-0.05 (-0.14 to 0.04)[b]	.31
Score of behavior to prevent unintentional child injury, mean (95% CI)				
Baseline	47.0 (46.7-47.4)	47.2 (46.8-47.6)	—	—
3-month	48.7 (48.4-49.1)	48.4 (48.0-48.8)	0.40 (-0.14 to 0.94)[b]	.15
6-month	48.9 (48.5-49.3)	48.0 (47.6-48.5)	0.87 (0.33 to 1.42)[b]	.002

[a]OR: odds ratio.

[b]B: regression coefficient.

[c]Odds ratio and regression coefficient are presented for the intervention effect from the generalized estimating equation (GEE) after adjusting for sociodemographic variables (ie, caregiver's age, gender, education level, household income, frequency of using parenting apps, and recent learning about child injury prevention; and child's age and gender), outcome variables at baseline, and engagement with the interventions in the assigned group (ie, number of log-ins, length of time using the app at each log-in, number of knowledge segments studied, number of knowledge segments bookmarked, and number of posted comments).

[d]Reference group.

Attitudes Concerning Child Unintentional Injury and Behaviors to Prevent Injuries

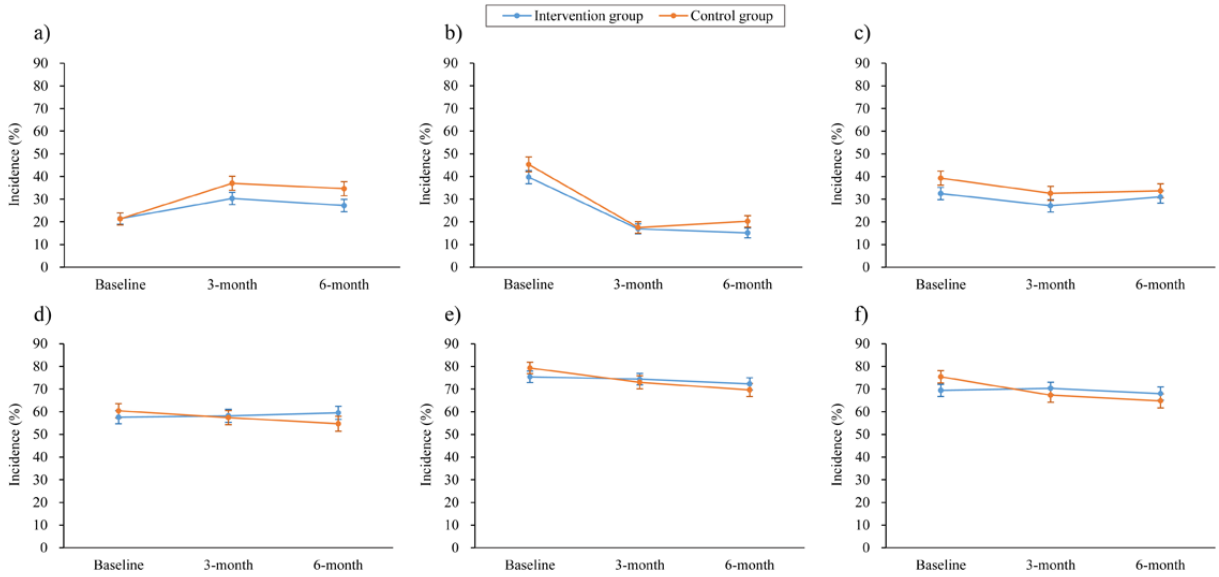
The mean caregiver total attitude score changed from 6.6 (95% CI 6.6-6.7) at baseline to 6.8 (95% CI 6.7-6.8) at the 6-month follow-up for the intervention group, and from 6.6 (95% CI 6.5-6.6) to 6.7 (95% CI 6.6-6.7) for the control group (see Table 2). Mean changes in the total score were not significant between the two groups after adjusting for the covariates (B .05, 95% CI -0.03 to 0.13).

The mean changes in total caregiver behavior scores to prevent unintentional child injury did differ between the control and intervention groups (*P*=.002) after the 6-month intervention. After adjusting for sociodemographics, baseline outcomes, and engagement factors, behavior scores increased from 47.0 (95% CI 46.7-47.4) to 48.9 (95% CI 48.5-49.3) for the intervention group and from 47.2 (95% CI 46.8-47.6) to 48.0 (95% CI 47.6-48.5) for the control group (B .87, 95% CI 0.33-1.42) (see Table 2).

When we examined individual injury prevention behavior scores, we detected significant differences between the groups in the changes of the scores from baseline to 6-month follow-up for two risky behaviors: giving children whole or large pieces of food (OR 0.73, 95% CI 0.60-0.89) and placing children in the front seat while riding a car (OR 0.73, 95% CI 0.57-0.93). We also detected significant between-group differences in score changes for three safe behaviors: testing the water temperature before giving children a bath (OR 1.26, 95% CI 1.05-1.52); placing sharp objects where children cannot reach them (OR 1.24, 95% CI 1.01-1.52); and storing medicines, detergents, and pesticides where children cannot reach them (OR 1.24, 95% CI 1.02-1.51) (see Figure 3). Another risky behavior was reduced at the 3-month follow-up visit: letting child ride a bicycle, electric bicycle, or motorcycle unsupervised (OR 0.80, 95% CI 0.65-0.99), but this behavior did not demonstrate significant change at the 6-month follow-up visit (see Table MA3-3 in Multimedia Appendix 3).

Results from the sensitivity analysis using per-protocol analysis were similar to the presented intention-to-treat analysis (see Table MA3-4 in Multimedia Appendix 3).

Figure 3. Incidence of caregiver parenting behaviors to prevent unintentional child injury in the past week. a) Giving child a whole or large piece of food that creates a choking risk; b) Placing child in the front seat while riding in a car; c) Letting child ride a bicycle, electric bicycle, or motorcycle unsupervised; d) Testing water temperature before giving child a bath; e) Placing sharp objects where children cannot reach; f) Storing medicines, detergents, and pesticides where children cannot reach. Only those parenting behaviors with significant changes between intervention and control groups are included here.



Discussion

Principal Findings

To our knowledge, this study represents the first cluster RCT testing the effectiveness of a theory-driven app-based intervention to prevent unintentional child injury through improved parenting of preschoolers in China. We found that following 6 months of using the app, caregivers in the intervention group did not report significantly different rates of unintentional injury incidence among their preschoolers compared to caregivers in the control group, nor did caregivers' attitudes toward unintentional child injury prevention differ across the groups. We did discover differences in caregivers' self-reported injury prevention behaviors, with caregivers in the intervention group reporting more increases in injury prevention behaviors after using the app for 6 months compared to those in the control group.

Interpretation

Despite findings that suggest parent behaviors changed after engaging in the intervention, results did not support our hypothesis that the app would reduce unintentional child injury incidence. The results we found concerning changes in parenting behavior agree with previous publications in HICs [12-15]. The nonsignificant results for unintentional child injury incidence may be a result of a few factors. First, isolated parenting behaviors do not prevent all child injuries. When children are supervised by others, such as teachers, parent behavior changes will likely have minimal influence. Further, the social and physical environments the child engages within are likely to be influenced minimally by caregivers but are likely to influence child injury risk to some degree [20,28]. A second factor that may have influenced our nonsignificant results is the fact that child injuries leading to serious restriction in activity or professional medical care are comparatively low. As is common in the field [1], injuries were relatively infrequent among our sample: incidence over the past 3 months was 9.14% (ie, at baseline, 181 out of 1980 participants experienced an injury in the past 3 months). Although we recruited a large sample size and examined injuries over a 6-month period, we may not have accumulated sufficient events to statistically detect the intervention effect for a minor effect size (OR 1.16 in this study).

Third, the effectiveness of any app-based interventions depends on the users' engagement [29]. Participants in the intervention group spent an average of 2.7 hours using the app, during which they studied an average of 45 knowledge segments and posted an average of 31.5 comments (see Table MA3-2 in Multimedia Appendix 3). There is no standard to dictate what extent of app engagement might yield knowledge transmission or behavior change. However, results of participant engagement from our study are consistent with another study reporting null results [30], and engagement was significantly lower than in studies reporting significant effects (eg, 13.1 hours of engagement in an effective collaborative care intervention for violence risk behaviors, substance use, and posttraumatic stress and depressive symptoms among injured adolescents [31]).

One other finding is noteworthy: to our surprise, we did not detect significant changes in parenting attitude scores between the groups after the 6-month intervention. This finding may reflect a ceiling effect: caregivers in both groups already had high attitude scores at baseline and, therefore, a significant change was difficult to achieve. For example, the proportion of caregivers reporting low self-efficacy to keep children safe from unintentional injuries was only 3.73% (40/1073) in the intervention group and 3.8% (34/907) in the control group at baseline (see Figure MA3-2 in Multimedia Appendix 3).

Implications

Our findings have two implications. First, in combination with previous findings [12-15], our results suggest that a mobile phone-based app that is grounded in behavior change theory might ultimately be effective to improve child injury prevention behaviors among caregivers of preschoolers in China. Specifically, in this study we yielded change in self-reported behavior, although we did not yield statistically significant change in injury incidence or caregiver attitudes about safety. These results imply there may be a need for further app development and for empirical research to discover what strategies might be most effective to yield child safety-related behavior change among caregivers and ultimately reduce injury risk among children. For example, refinement of the app might include ways to engage caregivers more intensely and effectively, perhaps through a more elaborate system of points, levels, or rewards. Others have reported increased engagement through attractive multimedia content and features with diversified content and delivery forms (eg, video, audio, games, and progress bars) [32]; increased engagement might lead more effectively to desired behavior change outcomes [29].

Second, app-based interventions might not function well in isolation. Instead, they might be more beneficial if they were integrated into a multi-faceted intervention program that maximizes the benefit of mobile health technology in safety education. For example, our app may be integrated within the Basic Public Health Service Program [33] run by the Chinese government, and then re-evaluated for its value in unintentional child injury prevention at a larger scale.

Limitations

This study had several limitations. First, our results relied on self-reported data and used a recall time period of the past 3 months for the child injury reports. Such data may be influenced by various biases [34,35]. Second, a few demographic characteristics differed between the participants who withdrew from the study compared to those who completed. Differential withdrawal rates may have led to unexpected biases in our findings. Third, the inclusion criteria of recruiting children only from preschools with more than 100 students may restrict the generalizability of our findings. Such relatively large preschools represent 87.8% (44/361) of all preschools in Changsha, but families that choose smaller preschools for their children may be different in some characteristics relevant to the effectiveness of the intervention.

In addition, we permitted study participants in both groups to attend parenting-related education and learning activities, which might have contaminated our data in some way. However, we felt it unethical to prohibit such educational activities, and randomization suggests that potential biases in our findings due to such contamination would be minimal. Finally, the actual incidence of child unintentional injury in our study was lower than that estimated when we calculated sample size requirements for the study. The withdrawal rate of participants (940/2920, 32.19%) in the study was also higher than expected. Both of these factors led to a sample size that was likely inadequate to detect small incidence or behavior changes. Future studies should consider strategies to retain participants; these might include a more engaging app, more frequent communications and reminders to users, financial incentives for completion of study tasks, and community involvement in the study design and implementation [36-38].

Conclusions

The app-based intervention for caregivers did not significantly reduce unintentional injury incidence among Chinese preschoolers but did substantially improve caregivers' behaviors relevant to prevention of unintentional injury. We recommend further refinement and assessment of the app-based intervention to improve its effectiveness in reducing child injury risk in China and other LMICs.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (grant number: 81573260). The sponsors had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The authors thank Xiang Wei, Bo Chen, Jieyi He, Wangxin Xiao, Yuyan Gao, Liheng Tan, Yanhong Fu, Deyue Gao, Cifu Xie, Ruotong Li, Jie Li, and Wanhui Wang from the Xiangya School of Public Health, Central South University, Changsha, China, for their assistance with the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Effectiveness of an app-based intervention for unintentional injury among caregivers of preschoolers: Protocol for a cluster randomized controlled trial.
[PDF File (Adobe PDF File), 2MB - mhealth_v7i8e13519_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 519KB - mhealth_v7i8e13519_app2.pdf]

Multimedia Appendix 3
Statistical models, appendix figures, and appendix tables.
[PDF File (Adobe PDF File), 1MB - mhealth_v7i8e13519_app3.pdf]

References

1. Institute for Health Metrics and Evaluation. GBD compare URL: https://vizhub.healthdata.org/gbd-compare/ [WebCite Cache ID 75jsL1grD]
2. Prasad MR, Swank PR, Ewing-Cobbs L. Long-term school outcomes of children and adolescents with traumatic brain injury. J Head Trauma Rehabil 2017;32(1):E24-E32 [FREE Full text] [doi: 10.1097/HTR.0000000000000218] [Medline: 26828713]
3. Sariaslan A, Sharp DJ, D'Onofrio BM, Larsson H, Fazel S. Long-term outcomes associated with traumatic brain injury in childhood and adolescence: A nationwide Swedish cohort study of a wide range of medical and social outcomes. PLoS Med 2016 Dec;13(8):e1002103 [FREE Full text] [doi: 10.1371/journal.pmed.1002103] [Medline: 27552147]
4. Khatlani K, Alonge O, Rahman A, Hoque DME, Bhuiyan AA, Agrawal P, et al. Caregiver supervision practices and risk of childhood unintentional injury mortality in Bangladesh. Int J Environ Res Public Health 2017 Dec 11;14(5):515 [FREE Full text] [doi: 10.3390/ijerph14050515] [Medline: 28492502]
5. Hogan CM, Weaver NL, Cioni C, Fry J, Hamilton A, Thompson S. Parental perceptions, risks, and incidence of pediatric unintentional injuries. J Emerg Nurs 2018 May;44(3):267-273. [doi: 10.1016/j.jen.2017.07.017] [Medline: 28988836]
6. Garzon DL. Contributing factors to preschool unintentional injury. J Pediatr Nurs 2005 Dec;20(6):441-447. [doi: 10.1016/j.pedn.2005.03.014] [Medline: 16298285]
7. Kendrick D, Mulvaney CA, Ye L, Stevens T, Mytton JA, Stewart-Brown S. Parenting interventions for the prevention of unintentional injuries in childhood. Cochrane Database Syst Rev 2013 Mar 28(3):CD006020. [doi: 10.1002/14651858.CD006020.pub3] [Medline: 23543542]
8. Ning P, Schwebel DC, Hu G. Healthy China 2030: A missed opportunity for injury control. Inj Prev 2017 Dec;23(6):363 [FREE Full text] [doi: 10.1136/injuryprev-2017-042314] [Medline: 28232400]
9. Hu G, Baker TD, Baker SP. Injury control in China: Priorities and actions. Lancet 2009 Jan 17;373(9659):214. [doi: 10.1016/S0140-6736(09)60068-4] [Medline: 19150701]
10. Omaki E, Rizzutti N, Shields W, Zhu J, McDonald E, Stevens MW, et al. A systematic review of technology-based interventions for unintentional injury prevention education and behaviour change. Inj Prev 2017 Dec;23(2):138-146. [doi: 10.1136/injuryprev-2015-041740] [Medline: 26787740]
11. Nation M, Crusto C, Wandersman A, Kumpfer KL, Seybolt D, Morrissey-Kane E, et al. What works in prevention: Principles of effective prevention programs. Am Psychol 2003;58(6-7):449-456. [Medline: 12971191]
12. Gielen AC, Bishai DM, Omaki E, Shields WC, McDonald EM, Rizzutti NC, et al. Results of an RCT in two pediatric emergency departments to evaluate the efficacy of an mHealth educational app on car seat use. Am J Prev Med 2018 Dec;54(6):746-755. [doi: 10.1016/j.amepre.2018.01.042] [Medline: 29656914]
13. Omaki E, Shields WC, McDonald E, Aitken ME, Bishai D, Case J, et al. Evaluating a smartphone application to improve child passenger safety and fire safety knowledge and behaviour. Inj Prev 2017 Dec;23(1):58. [doi: 10.1136/injuryprev-2016-042161] [Medline: 27597399]
14. Burgess J, Watt K, Kimble RM, Cameron CM. Combining technology and research to prevent scald injuries (the Cool Runnings Intervention): Randomized controlled trial. J Med Internet Res 2018 Dec 10;20(10):e10361 [FREE Full text] [doi: 10.2196/10361] [Medline: 30305263]
15. McKenzie LB, Roberts KJ, Clark R, McAdams R, Abdel-Rasoul M, Klein EG, et al. A randomized controlled trial to evaluate the Make Safe Happen® app: A mobile technology-based safety behavior change intervention for increasing parents' safety knowledge and actions. Inj Epidemiol 2018 Mar 12;5(1):5 [FREE Full text] [doi: 10.1186/s40621-018-0133-3] [Medline: 29527644]
16. Ning P, Chen B, Cheng P, Yang Y, Schwebel DC, Yu R, et al. Effectiveness of an app-based intervention for unintentional injury among caregivers of preschoolers: Protocol for a cluster randomized controlled trial. BMC Public Health 2018 Dec 11;18(1):865 [FREE Full text] [doi: 10.1186/s12889-018-5790-1] [Medline: 29996813]
17. 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]
18. Community Law. 2004. Guardianship of children URL: http://communitylaw.org.nz/community-law-manual/chapter-26-parents-guardians-and-caregivers/guardianship-of-children-chapter-26/ [WebCite Cache ID 75jsVemQd]
19. Coupland C, DiGuiseppi C. The design and use of cluster randomised controlled trials in evaluating injury prevention interventions: Part 2. Design effect, sample size calculations and methods for analysis. Inj Prev 2010 Apr;16(2):132-136. [doi: 10.1136/ip.2009.023127] [Medline: 20363822]
20. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

21. Barnett DJ, Balicer RD, Blodgett D, Fews AL, Parker CL, Links JM. The application of the Haddon matrix to public health readiness and response planning. Environ Health Perspect 2005 May;113(5):561-566 [FREE Full text] [doi: 10.1289/ehp.7491] [Medline: 15866764]
22. Yin C, Song Y, Tabata Y, Ogata H, Hwang GJ. Developing and implementing a framework of participatory simulation for mobile learning using scaffolding. J Educ Techno Soc 2013;16(3):137-150 ERIC Number: EJ1016553.
23. Ning P, Gao D, Cheng P, Schwebel DC, Wei X, Tan L, et al. Needs analysis for a parenting app to prevent unintentional injury in newborn babies and toddlers: Focus group and survey study among Chinese caregivers. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e11957 [FREE Full text] [doi: 10.2196/11957] [Medline: 31038465]
24. World Health Organization. 2016. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version for 2016 URL: https://icd.who.int/browse10/2016/en#/W00-X59 [WebCite Cache ID 75jsp1hcS]
25. Wu Y, Zhang W, Zhang L, Schwebel DC, Ning P, Cheng X, et al. Non-fatal injuries treated outside a hospital in Hunan, China: Results from a household interview survey. Eur J Public Health 2017 Dec 01;27(2):331-334 [FREE Full text] [doi: 10.1093/eurpub/ckw114] [Medline: 27497437]
26. Morrongiello BA, House K. Measuring parent attributes and supervision behaviors relevant to child injury risk: Examining the usefulness of questionnaire measures. Inj Prev 2004 Apr;10(2):114-118 [FREE Full text] [doi: 10.1136/ip.2003.003459] [Medline: 15066978]
27. Will KE, Lorek EJ, Sabo CS, Kidd DG. Measuring injury risk perceptions: Feasibility of a Risk Estimation Scale. Am J Health Behav 2009;33(6):639-649. [Medline: 19320613]
28. Anderson ES, Winett RA, Wojcik JR. Self-regulation, self-efficacy, outcome expectations, and social support: Social cognitive theory and nutrition behavior. Ann Behav Med 2007;34(3):304-312. [doi: 10.1007/bf02874555] [Medline: 18020940]
29. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Dec;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
30. Elbers NA, Akkermans AJ, Cuijpers P, Bruinvels DJ. Effectiveness of a Web-based intervention for injured claimants: A randomized controlled trial. Trials 2013 Jul 20;14:227 [FREE Full text] [doi: 10.1186/1745-6215-14-227] [Medline: 23870540]
31. Zatzick D, Russo J, Lord SP, Varley C, Wang J, Berliner L, et al. Collaborative care intervention targeting violence risk behaviors, substance use, and posttraumatic stress and depressive symptoms in injured adolescents: A randomized clinical trial. JAMA Pediatr 2014 Jun;168(6):532-539. [doi: 10.1001/jamapediatrics.2013.4784] [Medline: 24733515]
32. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: An analysis of app functionality and user reviews. J Med Internet Res 2018 Dec 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
33. Central People's Government of the People's Republic of China. National basic public health service project management information platform URL: http://www.nbphsp.org.cn/ [WebCite Cache ID 75jv4yRsJ]
34. Moshiro C, Heuch I, Astrøm AN, Setel P, Kvåle G. Effect of recall on estimation of non-fatal injury rates: A community based study in Tanzania. Inj Prev 2005 Feb;11(1):48-52 [FREE Full text] [doi: 10.1136/ip.2004.005645] [Medline: 15691990]
35. Mock C, Acheampong F, Adjei S, Koepsell T. The effect of recall on estimation of incidence rates for injury in Ghana. Int J Epidemiol 1999 Aug;28(4):750-755. [doi: 10.1093/ije/28.4.750] [Medline: 10480706]
36. Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol 2007 Aug;60(8):757-765 [FREE Full text] [doi: 10.1016/j.jclinepi.2006.11.023] [Medline: 17606170]
37. Robinson KA, Dinglas VD, Sukrithan V, Yalamanchilli R, Mendez-Tellez PA, Dennison-Himmelfarb C, et al. Updated systematic review identifies substantial number of retention strategies: Using more strategies retains more study participants. J Clin Epidemiol 2015 Dec;68(12):1481-1487 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.013] [Medline: 26186981]
38. Gourash WF, Ebel F, Lancaster K, Adeniji A, Koozer Iacono L, Eagleton JK, LABS Consortium Retention Writing Group. Longitudinal Assessment of Bariatric Surgery (LABS): Retention strategy and results at 24 months. Surg Obes Relat Dis 2013;9(4):514-519 [FREE Full text] [doi: 10.1016/j.soard.2013.02.012] [Medline: 23747313]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FRAME: Framework for the Rational Analysis of Mobile Education
GEE: generalized estimating equation
HIC: high-income country
ICC: intraclass correlation
ICER: incremental cost-effectiveness ratio
LMICs: low- and middle-income countries

mHealth: mobile health
OR: odds ratio
RCT: randomized controlled trial
TPB: Theory of Planned Behavior

Edited by G Eysenbach; submitted 28.01.19; peer-reviewed by MS Aslam, R McAdams, Q Yuan; comments to author 22.03.19; revised version received 08.05.19; accepted 16.06.19; published 09.08.19.

Please cite as:
Ning P, Cheng P, Schwebel DC, Yang Y, Yu R, Deng J, Li S, Hu G
An App-Based Intervention for Caregivers to Prevent Unintentional Injury Among Preschoolers: Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(8):e13519
URL: https://mhealth.jmir.org/2019/8/e13519/
doi:10.2196/13519
PMID:31400105

Original Paper

Counseling With Guided Use of a Mobile Well-Being App for Students Experiencing Anxiety or Depression: Clinical Outcomes of a Feasibility Trial Embedded in a Student Counseling Service

Emma Broglia[1,2*], BSc, MRes, PhD; Abigail Millings[2*], BSc, PhD; Michael Barkham[2*], BEd, MA, MSc, PhD

[1]Research Department, British Association for Counselling and Psychotherapy, Leicestershire, United Kingdom

[2]Department of Psychology, University of Sheffield, Sheffield, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Emma Broglia, BSc, MRes, PhD
Research Department
British Association for Counselling and Psychotherapy
15 St John's Business Park
Lutterworth
Leicestershire,
United Kingdom
Phone: 44 1455 883318
Email: emma.broglia@bacp.co.uk

Abstract

Background: Anxiety and depression continue to be prominent experiences of students approaching their university counseling service. These services face unique challenges to ensure that they continue to offer quality support with fewer resources to a growing student population. The convenience and availability of mobile phone apps offer innovative solutions to address therapeutic challenges and expand the reach of traditional support.

Objective: The primary aim of this study was to establish the feasibility of a trial in which guided use of a mobile phone well-being app was introduced into a student counseling service and offered as an adjunct to face-to-face counseling.

Methods: The feasibility trial used a two-arm, parallel nonrandomized design comparing counseling alone (treatment as usual, or TAU) versus counseling supplemented with guided use of a mobile phone well-being app (intervention) for 38 university students experiencing moderate anxiety or depression. Students in both conditions received up to 6 sessions of face-to-face counseling within a 3-month period. Students who approached the counseling service and were accepted for counseling were invited to join the trial. Feasibility factors evaluated include recruitment duration, treatment preference, randomization acceptability, and intervention fidelity. Clinical outcomes and clinical change were assessed with routine clinical outcome measures administered every counseling session and follow-up phases at 3 and 6 months after recruitment.

Results: Both groups demonstrated reduced clinical severity by the end of counseling. This was particularly noticeable for depression, social anxiety, and hostility, whereby clients moved from elevated clinical to low clinical or from low clinical to nonclinical by the end of the intervention. By the 6-month follow-up, TAU clients' (n=18) anxiety had increased whereas intervention clients' (n=20) anxiety continued to decrease, and this group difference was significant (Generalized Anxiety Disorder–7: t_{22}=3.46, P=.002). This group difference was not replicated for levels of depression: students in both groups continued to decrease their levels of depression by a similar amount at the 6-month follow-up (Physical Health Questionnaire–9: t_{22}=1.30, P=.21).

Conclusion: Supplementing face-to-face counseling with guided use of a well-being app is a feasible and acceptable treatment option for university students experiencing moderate anxiety or depression. The feasibility trial was successfully embedded into a university counseling service without denying access to treatment and with minimal disruption to the service. This study provides preliminary evidence for using a well-being app to maintain clinical improvements for anxiety following the completion of counseling. The design of the feasibility trial provides the groundwork for the development of future pilot trials and definitive trials embedded in a student counseling service.

Trial registration: ISRCTN registry ISRCTN55102899; http://www.isrctn.com/ISRCTN55102899

(JMIR Mhealth Uhealth 2019;7(8):e14318) doi:10.2196/14318

KEYWORDS

counseling; students; mental health; mobile app; feasibility studies; outcome measures; depressive symptoms; generalized anxiety; universities

Introduction

Identifying the prevalence of mental ill-health in university students has been a longstanding priority of educational institutions, and the growing concern over student mental health is widespread [1-3]. Research reports collating global accounts of student mental ill-health have found that up to 74% of students experience moderate emotional distress by the second semester, and this, in turn, has been linked to negative outcomes including low academic performance, isolation, financial problems, and time away from education leading to depression [4-7]. In recent years, the World Mental Health Survey of 13,984 full-time students from 19 colleges across 8 countries found that 35% met the threshold for at least one of the following common lifetime mental disorders (according to the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*, or DSM-IV): major depression, mania, generalized anxiety disorder, alcohol abuse, substance misuse, or panic disorder [8]. The survey also found that 75% of the students would not seek help for an emotional or mental health concern, with the most prominent reasons being that students would rather tackle the problem alone or seek help from friends or they feel too embarrassed to seek professional help [9]. An earlier study spanning 21 countries found that only 16% of 6452 students with diagnosable mental health disorders received the minimum adequate treatment, and this percentage decreased proportionally according to the income level of the country [10].

Research into the prevalence of mental ill-health in students has gained national attention, but there is limited recent data. In 2013, the National Union of Students (NUS) surveyed 1336 UK university students and found that the most common factors contributing to mental distress were academic deadlines, academic performance, and work-life balance [11]. Regarding help-seeking, 26% of students did not discuss their mental health concerns and of those who did, the most popular contacts were friends, family, their doctor, or an academic. Only 10% approached the university counseling service compared with 17% to 21% who approached and used advice from their union. Between 2014 and 2015, the Higher Education Funding Council for England commissioned a report to identify the mental health needs of students and found increased declarations for student disability on the basis of mental ill-health [12]. The report identified increased academic staff time being used to address student mental health as well as therapeutic staff managing larger caseloads and seeing more students with complex or comorbid concerns for anxiety, depression, and self-harm.

The challenges to addressing student mental health in UK higher education institutions have been characterized by comparing 113 in-house student counseling services [13]. Increased demand for counseling, drop out between counseling sessions, and growing waiting lists are some of the critical issues that student counseling services face. While not unique to the education sector, these challenges, along with sector-specific challenges of changes to the funding climate and disruptions from academic timetables [14,15], have hindered the development of robust research, rendering the evidence base for student counseling limited and pathways to improve service provision unclear. That said, many higher education institutions are either offering or are interested in offering digital solutions as a way of attempting to do more for less [13] and make their limited resources reach further. The use of technology in student counseling services has steadily grown over time, and the types of technology being used have changed in line with the availability of new technology and evidence. Early applications of technology in therapy altered the format in which counseling was traditionally delivered, offering counseling via telephone, email, and videoconference as a way of extending face-to-face support to clients unable to attend the counseling service [16].

As the application and convenience of technology have advanced, the use of digital technology offers innovative ways to address therapeutic challenges and expand the reach of traditional support. The most commonly used digital mental health tools today are Web- and phone-based apps, both of which are widely used and recommended in national health services, including the National Health Service (NHS) in England [17]. Evidence for these therapeutic technologies is steadily growing, and Web-based interventions for anxiety, depression, and stress can be effective in students [18]. Web-based interventions appear particularly effective for anxiety and depression, leading to significant reductions in symptom severity with small to medium effect sizes that are maintained long term [19]. These findings are consistent with a meta-analysis of 16 clinical trials comparing internet-based cognitive behavioral therapy (iCBT) versus waitlist control groups, demonstrating less worsening of symptoms for intervention groups [20]. Such evidence has been heavily weighted toward studies of the general population rather than students. However, research involving students has emerged in recent years and highlights a shift toward evaluating the role of digital technology in improving student mental health outcomes. One randomized controlled trial (RCT) established the efficacy of an unguided internet intervention for 200 university students diagnosed with social anxiety disorder [21]. Following a 10-week self-guided text-based intervention, moderate to large effects were shown to reduce social phobia (d=0.76, 95% CI 0.47-1.04), depression (d=0.50, 95% CI 0.22-0.78), and fear of positive evaluation (d=0.27, 95% CI 0.01-0.55) and improve quality of life (d=0.41, 95% CI 0.13-0.69).

A meta-analysis of mental health and well-being internet interventions for university students evaluated 48 RCTs and identified small effects on stress (g=0.20, 95% CI 0.02-0.38) and depression (g=0.18, 95% CI 0.08-0.27), with moderate effects on disordered eating (g=0.52, 95% CI 0.22-0.83) [22]. These effects were particularly high for interventions that were

4 to 8 weeks long and based on CBT and when compared to control groups receiving no intervention. These findings show potential for internet interventions in the context of student mental health as in-house support services typically offer short-term support in the first instance. While promising, this meta-analysis also demonstrates the need for research to explore broader applications of internet interventions outside the remit of CBT to be more inclusive of the range of mental health interventions available at universities (eg, counseling) and expand the evidence base to include research designs that go beyond comparisons with no intervention. Moreover, recent years have seen the rapid development of smartphone-based apps, which are arguably more suited to the student lifestyle. Smartphone apps can be effective for treating anxiety [23] and depression [24], but little is known about the perception, application, or potential benefits of augmenting student counseling services with apps.

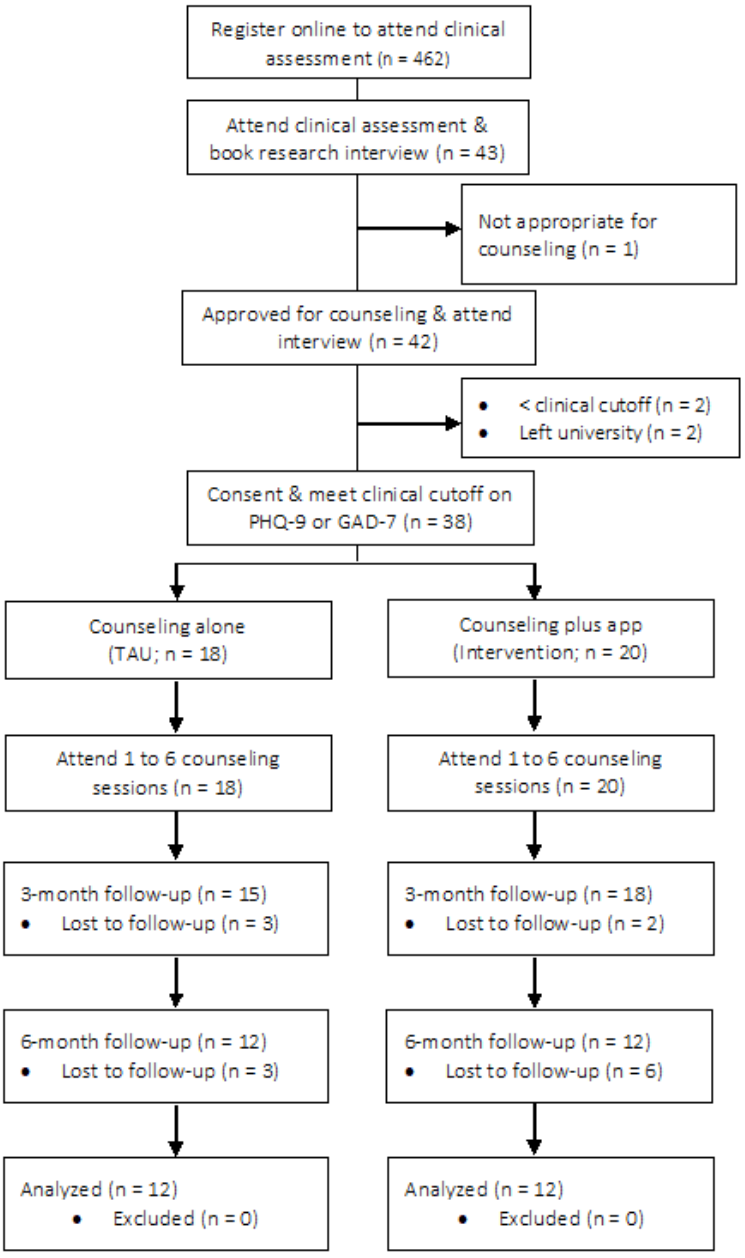
There is a significant need to explore innovative solutions to ensure that counseling services continue to offer quality support to a growing student population while being under continual threat of reduced resources. Augmenting counseling with apps (ie, blended intervention) may provide one such solution, and a feasibility trial is required to examine the feasibility and acceptability of doing so. Recent studies exploring the determinants of blended approaches have shown that therapists perceive the possibilities of adding blended work to their routine practice, blended approaches may help to reduce treatment gaps, there are training opportunities for offering mobile apps in clinical settings, clients can experience apps to be supportive of face-to-face interventions, and blended approaches offer potential to reach more patients than face-to-face alone [25-28]. There is a need to explore the feasibility of offering blended interventions to university students and to do so in a natural setting without disrupting the service or denying access to support. Accordingly, the primary aim of this study was to report on the outcomes of a feasibility trial in which guided use of a mobile phone well-being app was introduced into a student counseling service and offered as an adjunct to face-to-face counseling [29]. This primary aim was assessed by a range of feasibility factors including recruitment, treatment preference, and randomization acceptability. The secondary aim was to examine clinical outcomes and clinical change.

Methods

Trial Design and Setting

This feasibility trial used a two-arm, parallel, nonrandomized design comparing counseling alone (treatment as usual, or TAU) versus counseling supplemented with guided use of a well-being app and discussion of app activities (intervention) for university students experiencing anxiety or depression (Figure 1). The feasibility trial was registered on the ISRCTN registry [ISRCTN55102899] in 2016 under the acronym CASELOAD (Counseling Plus Apps for Students Experiencing Levels of Anxiety or Depression) and has been published elsewhere [29]. The study received ethical approval from the University of Sheffield Department of Psychology Research Ethics Committee on January 5 2016 (ref: 006171) and underwent independent scientific review. The trial was embedded in the university counseling service with high stakeholder engagement from the head of service and therapeutic team.

Figure 1. Participant flow diagram summarizing the number of clients that were recruited for the feasibility trial, were eligible, and participated at 3-month and 6-month follow-up phases. PHQ-9: Patient Health Questionnaire, 9-item; GAD-7: Generalized Anxiety Disorder, 7-item; TAU: treatment as usual.



Participants and Procedure

Participants were 38 help-seeking university students (aged 18 years and over) who had been accepted for counseling and met moderate clinical criteria on one of two standardized outcome measures for anxiety or depression. Inclusion criteria also included (1) newly registered for counseling, (2) undergraduate or postgraduate, and (3) having access to a personal smartphone with ability to install a publicly available app (iOS or Android). Participants were excluded if they met any of the following criteria: (1) present with a high risk to self or others, (2) already receiving therapeutic support, or (3) complex mental health problems beyond anxiety and/or depression. In line with routine practice, students who approached the counseling service were assessed by a therapist to determine their appropriateness for counseling. Students approved for counseling were invited to attend a 20-minute research interview to determine their eligibility. Students who attended the research interview were asked to provide written informed consent and were assessed for eligibility through completion of the standardized outcome measures. Eligible participants were allocated to either the TAU condition or intervention according to the clinical judgement of the therapist who provided the initial assessment. Further details on the rationale of the allocation procedure can be found in the trial protocol [29]. The progression from registration through the feasibility trial, with reasons for exclusions, is presented in Figure 1.

Demographics

Demographic information (eg, age, gender, and course information) was provided by clients when they completed the online registration form to be considered for counseling. Following online registration, clients attended the counseling service for the initial assessment to determine whether counseling would be an appropriate treatment option. Clients who were eligible for counseling and met the trial eligibility criteria attended a research interview and provided written consent for their demographic information to be added to the data collected during the trial (eg, clinical data).

Measures

The time frame for administering measures has been detailed in the trial protocol [29] in the form of a Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) diagram.

Generalized Anxiety Disorder–7 and Patient Health Questionnaire–9

Anxiety and depression were measured with the 7-item Generalized Anxiety Disorder (GAD-7) [30] and the 9-item Patient Health Questionnaire (PHQ-9) [31]. These measures were in addition to the clinical measures used by the counseling service to allow comparisons with external counseling services. The GAD-7 and PHQ-9 were administered at baseline, 3 months, and 6 months following recruitment. Items refer to the previous 2 weeks and are scored on a 4-point Likert scale (0 = not at all, 3 = nearly every day) whereby higher scores indicate higher severity. The PHQ-9 is a reliable measure of depression severity that has been validated against other measures of depression [31-33]. The GAD-7, which is widely used alongside the PHQ-9, is a reliable screening tool for generalized anxiety with good construct, criterion, factorial, and procedural validity.

Clinical Outcomes in Routine Evaluation Outcome Measure–10

The 10-item Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-10) [34] was administered at the initial clinical assessment (preintervention) and at every counseling session to measure changes in general psychological functioning. Items refer to the previous week and are scored on a 5-point Likert scale (0 = not at all, 4 = most or all of the time), whereby higher scores indicate higher symptom severity. The CORE-10 is a shortened version of the 34-item Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) [35], which has been used extensively in mental health services in the United Kingdom. The 10-item version has been validated against the CORE-OM, is sensitive to change, and can be used to determine whether a client meets membership of a clinical population (score of ≥10).

Counseling Center Assessment of Psychological Symptoms–62

The Counseling Center Assessment of Psychological Symptoms (CCAPS-62) [36] is a measure developed specifically for the student college population and was administered with the CORE-10 at the initial clinical assessment (preintervention) to measure changes in student-specific mental health concerns. In line with its intended use, the shortened version CCAPS-34 was used alongside the CORE-10 at every counseling session after the initial assessment. The two versions are used interchangeably, and items refer to the previous 2-week period. Items are scored on a 5-point Likert scale (0 = not at all like me, 4 = extremely like me), whereby higher scores indicate higher symptom severity. The CCAPS-62 monitors changes in the following areas: depression, generalized anxiety, social anxiety, academic distress, eating concerns, hostility, substance use, family distress, and suicide ideation. The CCAPS-34 also monitors these areas except for family distress, which does not appear, and substance use, which is replaced with alcohol use. The CCAPS-62 has been validated against the CORE-10 using a UK student sample, which replicated the psychometric factor structure and internal reliability of the CCAPS [37].

Clinical Interventions

All participants received face-to-face counseling in line with standard practice, which included a wait period of 3 to 5 working days for the initial clinical assessment and 8 to 10 days between ongoing therapy sessions. Sessions were 50 minutes in length, and the frequency of sessions was determined through client-therapist discussions as well as the usual disruptions from the academic timetable (eg, student time-off during Easter, summer, and course placements). Two clinical interventions were available through the trial: counseling alone (TAU) and counseling supplemented with guided use of a well-being app (intervention).

Counseling Supplemented With Well-Being App (Intervention)

In addition to the standard level of care, counseling sessions in the intervention were supplemented with feedback on client use of a well-being app. Clients were encouraged to use the app independently between counseling sessions with the intention to review app exercises with their therapist during each face-to-face counseling session. Therapists were provided with tablet computers to allow clients to access their app account and review their progress. Through this process, therapists reviewed clients' app activity and together they decided which activities could be beneficial for clients to use between sessions.

App features included (1) daily behavior monitoring (eg, for mood, sleep, exercise, alcohol consumption, medication use, and time spent outside); (2) reflective thinking exercises with guided CBT, mindfulness, and positive visualization; (3) guided relaxation with breathing and meditation; (4) peer-led support through anonymous online communities and private groups; and (5) tracking short-term and long-term goals. Between counseling sessions, the app provided daily prompts to encourage participants to log their mood and behavior. However, any additional activity required clients to go into the app and select an exercise—for example, the exercise that was suggested during their last counseling session. Therapists did not have access to clients' app activity except during face-to-face counseling sessions if clients decided to show therapists their activity. Therapists were encouraged to prompt the decision to review app activity; however, the decision was ultimately that of the clients.

Selection of Well-Being App

There are many smartphone apps that offer tools for improving well-being through a range of common features typically based on CBT and mindfulness. The quality of mental health apps was assessed in a recent systematic review, and the top-ranking apps, starting with the highest quality, were (1) HealthyMinds, (2) AnxietyCoach, (3) Moodkit, (4) Pacifica, and (5) Self-help for Anxiety Management (SAM) [38]. This quality assessment incorporated 16 recommendations, and these top 5 apps have the following properties: based on CBT; address anxiety and low mood; permit the reporting of thoughts, feelings, and behaviors; offer reminders; offer inbuilt activities; and contain a visual log to monitor progress. To aid the decision of selecting a well-being app for the purpose of this trial, the following criteria were applied: (1) applicable to university students, (2) demonstrates potential to be integrated with face-to-face counseling, (3) available on iOS and Android platforms, and (4) offers a range of features overlapping with other well-being apps.

Based on these criteria, the Pacifica app [39] was selected and evaluated by a volunteer sample of students and in-house therapists before it was implemented in the feasibility trial. It is important to note that while this study used a specific app, the app and its functions are representative of well-designed apps, and the feasibility trial is not intended to be an evaluation of Pacifica per se. Annual app subscriptions were purchased and provided to participants as unique gift codes. All payments were subject to the standard fee for public users and no financial incentives or waivers were provided by the Pacifica development team.

Therapists Delivering the Interventions

All therapists were accredited by the British Association for Counselling and Psychotherapy (BACP) or the UK Council for Psychotherapy and were employed by the university counseling service. Six therapists (4 intervention, 2 TAU) were assigned to support the trial and delivered either the intervention or TAU condition, based on their preference. Therapists received additional training specific to the intervention they were allocated; this training has been described elsewhere [29]. Therapists were provided with a competence framework [40] for the university and college counseling context together with the most recent service clinical handbook to ensure best practice. Clinical practice was reinforced throughout the trial with regular team meetings and optional daily drop-in sessions for members of the counseling team to query issues with the onsite researcher.

Primary Outcomes (Feasibility)

The yield of the feasibility trial was a series of primary and secondary outcomes relating to a range of components that aimed to inform a definitive trial. The primary outcomes were recruitment duration, treatment preference, and randomization acceptability. The recruitment period for a definitive trial was estimated from this study by exploring the required time needed to reach 40 participants while also considering therapist availability. Treatment preference was determined by asking participants to indicate their preferred treatment out of the two options available: counseling alone versus counseling with guided use of a well-being app. Clients were further asked whether their assigned treatment condition affected their decision to join the trial—for example, if they were not assigned to their preferred condition. Similarly, clients were asked whether being randomized would increase their likelihood of withdrawing from the trial. For brevity, additional criteria listed in the protocol have not been reported [29].

Secondary Outcomes (Clinical Effects)

The secondary outcomes were clinical outcomes, clinical change across counseling sessions, clinical change at 3 months and 6 months following recruitment, and reliable and clinically significant improvement (RCSI) at 3 months and 6 months. Clinical outcomes were calculated as the difference between each counseling session and the difference between baseline, 3-month, and 6-month measures for the following mental health indicators: anxiety (GAD-7), depression (PHQ-9), psychological functioning (CORE-10), student-specific symptoms such as academic distress, social anxiety, substance use, and eating concerns (CCAPS-34 and 62). RCSI was calculated in line with standard methods [41]. According to these methods, clients must have had contact with the counseling service twice and their precounseling scores on the PHQ-9/GAD-7 must have been in the clinical range. To meet the threshold for RCSI, individual scores on both measures needed to meet each of two criteria: extent of change significantly greater than measurement error (reliable change) and postintervention score below the clinical cutoff, indicating nonclinical status (clinical change). For brevity, additional secondary criteria in the protocol have not been reported.

Statistical Analysis

Analyses were predominantly descriptive (eg, mean, standard deviation, minimum, and maximum) to characterize the study population and outline various feasibility measures. As the sample size was not powered to detect significant differences between TAU and intervention groups, data were primarily used to summarize group outcomes to reveal preliminary trends and inform the design of a pilot trial. Independent *t* tests were conducted where appropriate to clarify whether group differences were significant (eg, clinical severity at intake, across conditions).

Results

Recruitment

The recruitment period was from February to June 2016; however, recruitment was not active for the whole period because the service experienced several disruptions. These included disruptions from the academic calendar (eg, Easter break), breaks in staff contracts, and therapists with full caseloads (ie, no counseling slots available). Recruitment aimed to simulate the natural demands from the service to reduce service disruption. Participants were recruited into the trial when therapists had available slots rather than protecting counseling slots for trial participants. Similarly, the time between sessions was dependent on therapist availability as it would in routine practice. Recruitment ended with 38 participants (intervention, n=20; TAU, n=18); the trial entered the academic summer period

when students typically leave university before reaching the recruitment goal of 40 participants. Attention then focused on collecting follow-up measures with existing participants.

Treatment Preference and Randomization Acceptability

Of the TAU participants, 11 preferred their allocated condition because it required less work than the intervention, 6 participants preferred the intervention because it offered additional support, and 1 had no preference. Of participants in the intervention group, 1 participant preferred the TAU condition for requiring less input, 10 preferred the intervention, and 9 had no preference. Despite their treatment preferences, 37 participants reported they would still have joined the trial if they had been randomized to the alternative condition; 1 participant in the intervention group claimed they would likely have withdrawn if allocated to the TAU group as they were specifically interested in using the app alongside counseling.

Intervention Fidelity

Intervention fidelity was assessed using anonymized transcripts of counseling audio recordings from clients in the intervention group. A total of 45 recordings (12 clients across 5 therapists) were available and transcribed for analysis. Transcripts were scored to assess the following criteria: (1) number of times app was discussed, (2) duration of app discussion, (3) whether therapist reviewed client app use, (4) number of app features therapist suggested, and (5) missed opportunities to discuss client app use. All transcripts were rated by author EB, and 24% (11/45) of transcripts were also rated by a blinded independent researcher. Interrater reliability analysis revealed substantial to almost perfect agreement across raters with kappa values in the range of .81 to 1.00 [42]. On average, therapists reviewed client app use, held brief discussions to review client app activity (lasting 2 to 5 minutes per counseling session), and provided advice based on client feedback. Therapists rarely missed an opportunity to review client app use; however, 2 of the 5 therapists were more likely to miss an opportunity and one was less likely to initiate app discussion (ie, relied on client to raise app discussion). There were no clear associations between therapist checklist scores and therapeutic modality (eg, CBT psychotherapist versus humanistic).

Baseline Demographic Data

Forty-two participants were initially recruited by therapists and attended the research consent session to determine eligibility (Figure 1). Two participants, both female, were excluded because one scored <10 on the PHQ-9 and GAD-7 measures and another had been referred for online self-help instead of counseling. A further 2 participants left the university after the initial service assessment and were excluded from the trial. Of the remaining 38 participants, 20 were allocated to the intervention group (10/20, 50%, were women) and 18 were allocated to the TAU group (12/18, 67%, were women; allocation procedure can be found elsewhere [29]). The average age in the TAU group was 23 years (SD 4.11, minimum 19, maximum 32); the intervention group was younger with a mean age of 21 years (SD 3.24, minimum 19, maximum 35). Of the total sample, 79% (30/38) were undergraduate, 8 were postgraduate (taught = 3, research = 5), 30 were students studying in their birth country or from the European Union and 8 students were from an international university and visiting a UK university as part of their course. Participants studied in the following faculties: arts and humanities (6/38, 16%); engineering (9/38, 24%); medicine, health, or dentistry (3/38, 8%); science (10/38, 26%); and social science (10/38, 26%).

Baseline Clinical Scores

Table 1 shows that both the TAU and intervention groups met the eligibility criteria of moderate clinical threshold (≥10) on the PHQ-9 and GAD-7 at baseline. The TAU group scored higher than the intervention group for both PHQ-9 and GAD-7, but independent samples *t* tests revealed that these differences were not significant (PHQ-9: t_{36}=1.53, *P*=.14; GAD-7: t_{36}=0.82, *P*=.42). The group means for CORE-10 and CCAPS distress index met moderately severe clinical threshold, and the TAU group scored higher than the intervention group for the CCAPS distress index; however, this difference was not significant (t_{34}=1.11, *P*=.55). The remaining group differences between the CCAPS-62 subscales were also not significant (depression: t_{34}=1.42, *P*=.11; generalized anxiety: t_{34}=1.58, *P*=.12; social anxiety: t_{34}=1.84, *P*=.08; academic distress: t_{34}=1.11, *P*=.17; eating concerns: t_{34}=1.22, *P*=.23; family distress: t_{34}=1.19, *P*=.55; hostility: t_{34}=1.11, *P*=.85; substance use: t_{34}=1.38, *P*=.13).

Table 1. Baseline scores on clinical measures on the Physical Health Questionnaire–9, Generalized Anxiety Disorder–7, Clinical Outcomes in Routine Evaluation Outcome Measure–10, and Counseling Center Assessment of Psychological Symptoms–Distress Index across intervention and treatment as usual groups from a feasibility trial.

Measures	Baseline clinical scores across groups						*P* value[a]
	Treatment as usual			Intervention			
	N	Mean (SD)	Min-Max	N	Mean (SD)	Min-Max	
PHQ-9[b,c]	18	17.58 (5.99)	10-26	20	15.00 (4.71)	10-24	.14
GAD-7[c,d]	18	13.75 (4.56)	5-19	20	13.92 (4.27)	7-20	.42
CORE-10[e,f]	11	22.64 (6.44)	13-32	17	23.18 (5.41)	15-34	.55
CCAPS-62[g,h]							
Depression	16	2.70 (0.69)	1.31-3.54	20	2.30 (0.92)	0.08-3.54	.11
Generalized anxiety	16	2.70 (0.85)	1.44-4.00	20	2.26 (0.80)	0.11-3.56	.12
Social anxiety	16	3.00 (0.59)	1.86-3.86	20	2.53 (0.90)	0.00-3.86	.08
Academic distress	16	2.79 (0.98)	1.40-4.00	20	2.34 (0.93)	0.60-3.40	.17
Eating concerns	16	1.49 (0.96)	0.00-3.33	20	1.13 (0.82)	0.00-2.78	.23
Family distress	16	1.11 (0.79)	0.17-2.83	20	1.05 (0.87)	0.00-2.83	.55
Hostility	16	1.57 (0.93)	0.29-2.86	20	1.59 (0.95)	0.14-3.57	.85
Substance use	16	0.77 (1.20)	0.00-4.00	20	0.93 (0.94)	0.00-3.50	.13
Distress index	16	2.68 (0.66)	1.42-3.75	20	2.40 (0.82)	0.25-3.80	.55

[a]*P* value from independent samples *t* test comparing treatment as usual and intervention group means.

[b]PHQ-9: Physical Health Questionnaire, 9-item.

[c]PHQ-9 and GAD-7 clinical boundaries: 0-5 = mild, 6-10 = moderate, 11-15 = moderately severe, 16-20/21 = severe.

[d]GAD-7: Generalized Anxiety Disorder, 7-item.

[e]CORE-10: Clinical Outcomes in Routine Evaluation Outcome Measure, 10-item.

[f]CORE-10 clinical boundaries: 0-5 = healthy, 5-10 = low, 10-15 = mild, 15-20 = moderate, 20-25 = moderately severe

[g]CCAPS-62: Counseling Center Assessment of Psychological Symptoms, 62-item.

[h]CCAPS-Distress Index clinical boundaries: 12.1 = low, 2.15 = elevated.

Clinical Change

Participants in the intervention group on average received 5 counseling sessions (mean 4.73 [SD 1.62], minimum 2, maximum 9), and TAU participants received 4 (mean 5.29 [SD 1.73], minimum 2, maximum 10). The average waiting period between counseling sessions was approximately 4 days longer for the TAU group than the intervention group (TAU mean 19.50 [SD 16.43], minimum 4, maximum 75; intervention mean 15.92 [SD 6.91], minimum 7, maximum 31). As shown in Figures 2 and 3 both the TAU and intervention groups decreased their distress scores across counseling sessions for all CCAPS subscales but not for the CORE-10. By session 6, both groups had decreased their scores on social anxiety, alcohol use, and hostility to a large enough extent to leave the clinical group they met at session 1 (ie, below the elevated or low clinical boundary). This was also true for the intervention group mean for academic distress and eating concerns.

Figure 2. Clinical change scores for CORE-10 and CCAPS subscales across therapy sessions 1 to 6 for the TAU and intervention groups of the feasibility trial. CORE-10: Clinical Outcomes in Routine Evaluation Outcome Measure, 10-item; CCAPS-62: Counseling Center Assessment of Psychological Symptoms, 62-item; TAU: treatment as usual.

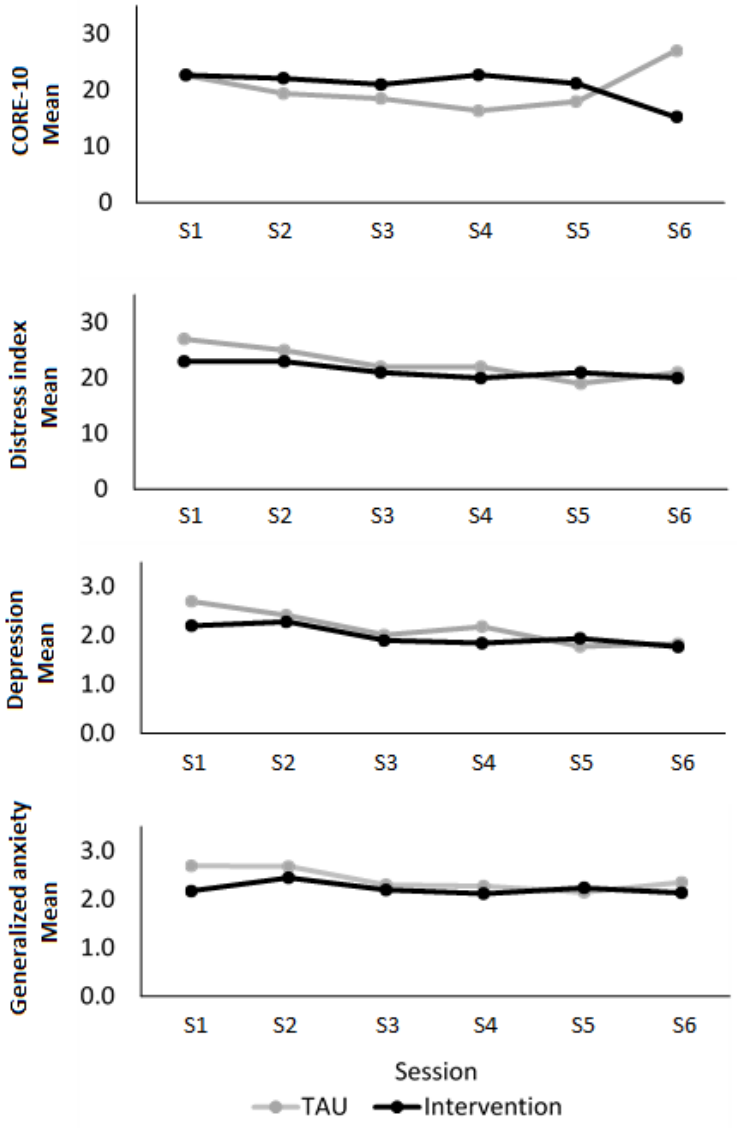
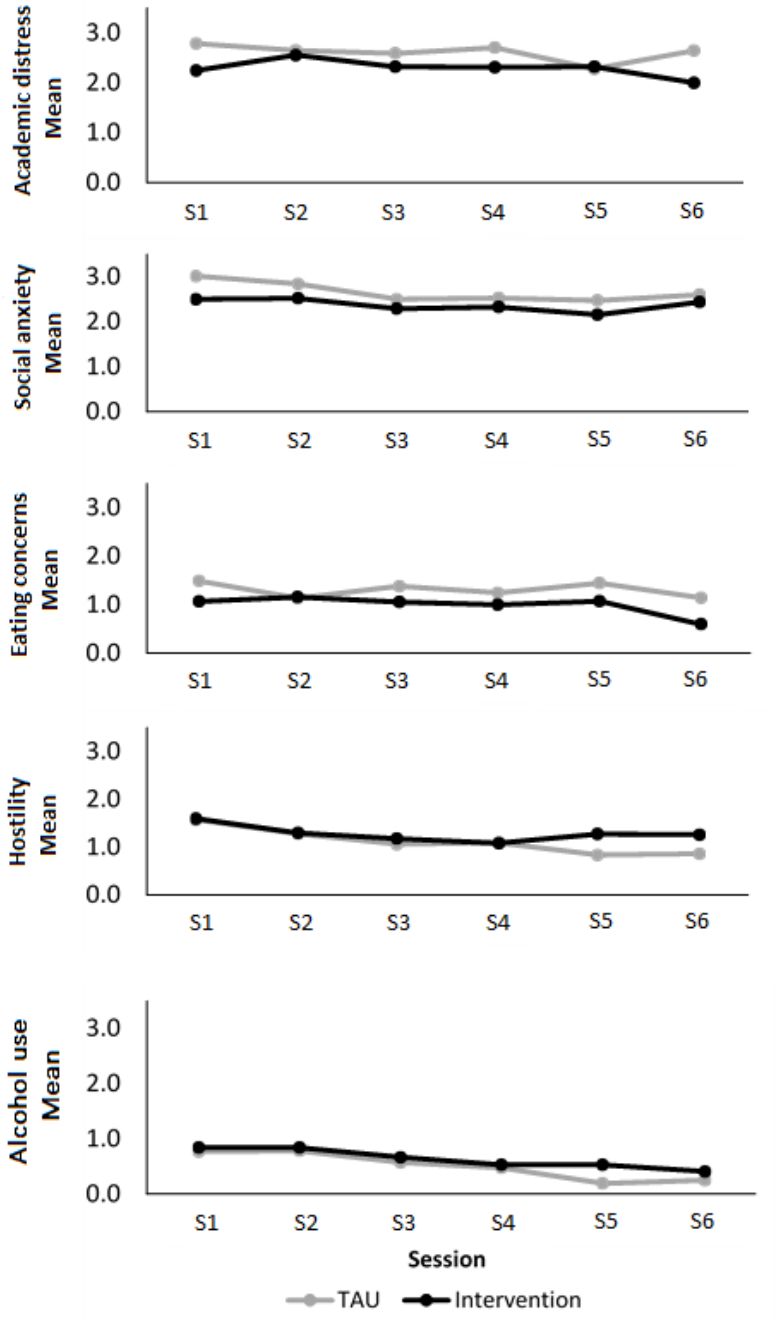


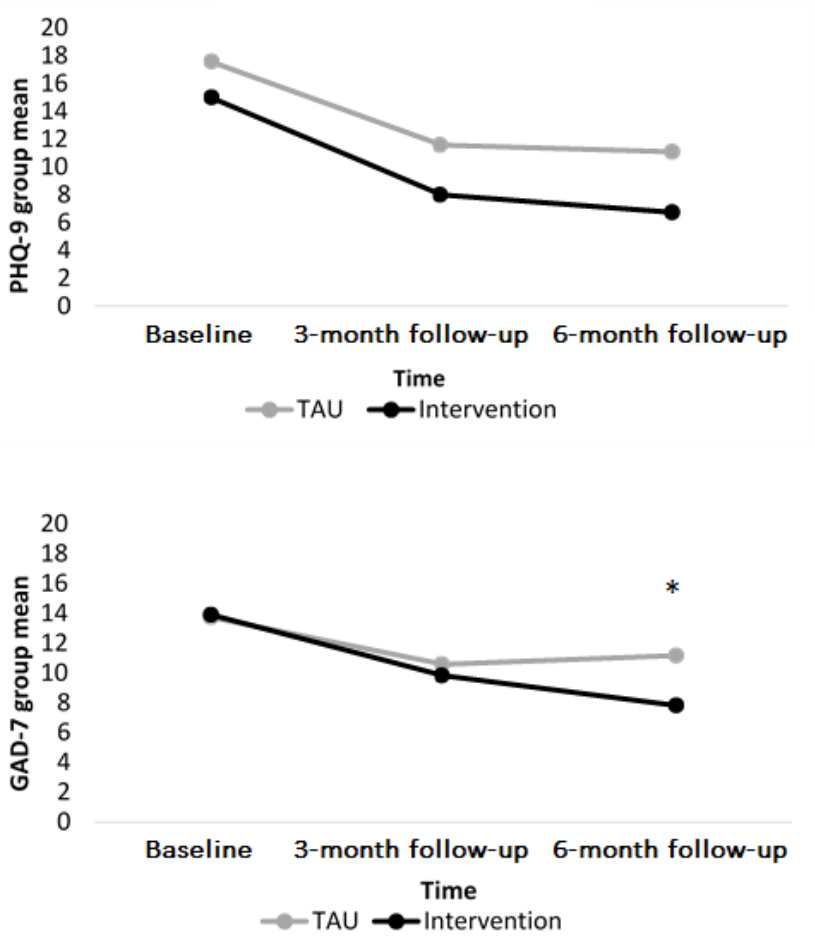
Figure 3. Clinical change scores for remaining CCAPS subscales across therapy sessions 1 to 6 for the TAU and intervention groups of the feasibility trial. CCAPS-62: Counseling Center Assessment of Psychological Symptoms, 62-item; TAU: treatment as usual.



Pre-Post Clinical Change

Clinical change was calculated as the difference between clinical scores for measures administered at baseline, 3 months, and 6 months following recruitment. All participants had completed counseling by the 6-month follow-up period. However, clients in both groups were still receiving treatment during the 3-month follow-up period. Depression and anxiety were measured at baseline, 3 months, and 6 months with the PHQ-9 and GAD-7. Figure 4 shows that PHQ-9 scores of both groups decreased between the baseline and 3-month follow-up period and continued to decrease slightly at the 6-month follow-up. GAD-7 scores of both groups also decreased between the baseline and 3-month follow-up but to a lesser extent than the PHQ-9 scores. By the 6-month follow-up, GAD-7 scores of the TAU clients had increased whereas GAD-7 scores of the intervention clients continued to decrease, and this difference was significant (GAD-7: t_{22}=3.46, P=.002). The difference between PHQ-9 scores at follow-up, however, was not significant (PHQ-9: t_{22}=1.30, P=.21).

Figure 4. PHQ-9 and GAD-7 scores across participants from the intervention and TAU groups at 3-month and 6-month follow-up measures. PHQ-9: Patient Health Questionnaire, 9-item; GAD-7: Generalized Anxiety Disorder, 7-item; TAU: treatment as usual.



Reliable and Clinically Significant Improvement

All clients in the feasibility trial met the criteria to calculate RCSI in line with the established criteria [41], which required clients to obtain a PHQ-9 ≥10 at baseline for depression, have decreased this score by ≥6 points, and have a post-counseling score of <10 (below the clinical threshold). The criteria for RCSI for anxiety required clients to obtain a GAD-7 ≥8 at baseline, have decreased this score by ≥5 points and have a post-counseling score of <8. If client post-counseling scores decreased by ≥6 points on the PHQ-9 or ≥5 points on the GAD-7 but their scores are not below the clinical thresholds, they have made a reliable improvement. Tables 2 and 3 show that at 6 months, 75% (9/12) of TAU clients met RCSI on the PHQ-9 and 17% (2/12) met RCSI on the GAD-7. Fewer intervention clients met RCSI on the PHQ-9 by 6 months compared with TAU clients. By contrast, more intervention clients met RCSI on the GAD-7 by 6 months, and 50% (6/12) met RCSI at 3 months.

Table 2. Individual treatment as usual participant scores on the Physical Health Questionnaire–9 and Generalized Anxiety Disorder–7 at baseline and 3- and 6-month follow-up measures with reliable and clinically significant improvement indicators.

Measure across time	PHQ-9[a]			GAD-7[b]		
	Baseline	3 month FU[c]	6 month FU	Baseline	3 month FU	6 month FU
Individual						
11	22	18	16[d]	16	13	14
12	10	4[e]	3[e]	5	3	4
15	16	10[d]	10[d]	11	9	10
17	13	6[e]	5[e]	10	1[e]	3[e]
19	10	10	8	12	11	11
26	23	20	20	14	15	16
28	23	20	21	18	17	17
31	25	10[d]	14[d]	19	13[d]	15
34	26	14[d]	13[d]	19	15	15
35	18	12[d]	12[d]	17	15	15
36	12	8	5[e]	16	11[d]	10[d]
41	13	7[e]	6[e]	8	4	4
Aggregate						
Mean	17.58	11.58	11.08	13.75	10.58	11.17
SD	5.99	5.38	5.96	4.56	5.28	5.06
Minimum	10	4	3	5	1	3
Maximum	26	20	21	19	17	17
RI[f] count, n (%)	—[g]	7 (58)	9 (75)	—	3 (25)	2 (17)
RCSI[h] count, n (%)	—	3 (25)	4 (33)	—	1 (8)	1 (8)

[a]PHQ-9: Physical Health Questionnaire, 9-item.

[b]GAD-7: Generalized Anxiety Disorder, 7-item.

[c]FU: follow-up.

[d]RI scores decreased by ≥6 points on the PHQ-9 and ≥5 points on the GAD-7.

[e]RCSI scores meet RI criteria and have postcounseling scores below the clinical cut-points.

[f]RI: reliable improvement.

[g]Not applicable, as RI and RCSI scores must be calculated from two data points.

[h]RCSI: reliable and clinically significant improvement.

Table 3. Individual intervention participant scores on the Physical Health Questionnaire–9 and Generalized Anxiety Disorder–7 at baseline and 3- and 6-month follow-up measures with reliable and clinically significant improvement indicators.

Measure across time	PHQ-9[a]			GAD-7[b]		
	Baseline	3 month FU[c]	6 month FU	Baseline	3 month FU	6 month FU
Individual						
02	23	6[d]	7[d]	17	11[e]	8[e]
03	16	9[d]	5[d]	7	3	3
04	15	6[d]	6[d]	20	18	15[e]
05	12	10	8	18	11[e]	9[e]
06	18	9[d]	8[d]	15	10[e]	10[e]
08	10	6	7	8	5	5
10	11	9	9	12	15	5[d]
13	12	7	5[d]	13	8[e]	7[d]
14	24	10[e]	8[d]	19	12[e]	10[d]
22	16	8[d]	5[d]	12	7[d]	6[d]
23	13	8	7[d]	10	6	6
27	10	8	6	16	12	10[e]
Aggregate						
Mean	15.00	8.00	6.75	13.92	9.83	7.83
SD	4.71	1.48	1.36	4.27	4.28	3.21
Minimum	10	6	5	7	3	3
Maximum	24	10	9	20	18	15
RI[f] count, n (%)	—[g]	6 (50)	8 (67)	—	6 (50)	9 (75)
RCSI[h] count, n (%)	—	5 (42)	8 (67)	—	1 (8)	4 (33)

[a]PHQ-9: Physical Health Questionnaire, 9-item.

[b]GAD-7: Generalized Anxiety Disorder, 7-item.

[c]FU: month follow-up.

[d]RCSI scores meet RI criteria and have post-counseling scores below the clinical cut-points.

[e]RI scores decreased by ≥6 points on PHQ-9 and ≥5 points on GAD-7.

[f]RI: reliable improvement.

[g]Not applicable, as RI and RCSI scores must be calculated from two data points.

[h]RCSI: reliable and clinically significant improvement.

Discussion

Principal Findings

This study reports on the feasibility, benefits, and challenges of supplementing counseling with a well-being app for students experiencing anxiety or depression. The trial assessed essential design elements to inform how well the trial was implemented into a service and whether the intervention was acceptable. The aim of the trial was to explore the feasibility and acceptability of supplementing counseling with a well-being app and whether this, in turn, had positive clinical outcomes.

Primary Feasibility Outcomes

The trial was conducted in a naturalistic setting and embedded into a student counseling service with early engagement from the therapeutic team. The design of the trial and subsequent intervention was developed with high stakeholder engagement from across the higher education counseling sector including service users, providers, and experts in the field of trial design. This collaborative approach aimed to minimize service disruption, optimize acceptability, and streamline the recruitment capacity of the service. Recruitment mirrored routine practice as far as possible in that therapist availability to take on new clients was a key factor. Recruitment delays were anticipated because of the reliance on therapist availability, but further

delays were introduced from breaks in staff contracts and therapists managing full caseloads.

A challenge of embedding a trial into practice is implementing the additional research requirements without negatively impacting the service. Randomization was a concern for this reason, and identifying randomization acceptability can help to alleviate disruption. Only one participant reported that they would have withdrawn if they had been randomized to the alternative condition, suggesting that randomization would be acceptable in a fully powered trial. While client preferences did not impact participation, evidence suggests that discussing and measuring client preference is associated with higher treatment satisfaction, engagement, and clinical outcomes [43]. Clients perceiving desirable components in both conditions also highlights a benefit of using active control groups compared with traditional wait-list control groups. This design choice is particularly relevant for the student counseling context as research suggests that comparing an active treatment against a wait-list control group is not an accurate comparison when testing short-term psychotherapy for depression or social anxiety [44].

Secondary Feasibility Outcomes

Clients in both groups decreased their levels of distress and clinical severity on all outcome measures as they progressed through counseling. This gradual reduction across counseling was observed despite clients in both groups entering counseling with moderately severe scores. The symptom profile of clients was multifaceted and particularly elevated for social anxiety, academic distress, depression, and generalized anxiety. These symptoms complement findings on the CCAPS from other UK-derived student counseling samples [37]. The elevated symptoms of social anxiety, depression, and generalized anxiety mimic three prevalent conditions reported in primary care with the additional impact on academic performance. In our counseling sample, the reduction of clinical scores across counseling provides preliminary evidence on the effectiveness of student counseling embedded in higher education.

In addition to the clinical improvement observed during counseling, depression scores for the majority of clients across both groups reliably improved by the end of the trial (ie, at 6-month follow-up). The addition of the well-being app had a larger impact on anxiety than depression as clients using the app reliably decreased their anxiety scores sooner and to a greater extent than clients who only received counseling. The group difference in anxiety is noteworthy because it was achieved despite clients entering the trial with similar levels of anxiety across the groups. This suggests that the addition of the app contributed to the group difference observed at the end of the trial, suggesting that counseling augmented with the app may be more effective than counseling alone at reducing anxiety. Given the aforementioned pressures on student counseling services, this finding has important implications for counseling services and the wider student population. The use of app-augmented counseling may be a way for counseling services embedded in higher education institutions to do more with less.

Limitations and Future Research

Feasibility trials are, by nature, underpowered and explorative because they provide the groundwork for clinical trials and inform key design elements. Research embedded into practice must also be pragmatic to ensure that trials are sufficiently implemented with minimal disruption to the service. These necessary design decisions introduced variability across the trial and potentially diluted the quality of the intervention. The delivery of the intervention needed to be flexible for several reasons. For instance, in-house counseling is variable across therapists and institutions because services are designed to respond to the unique needs of their students. Counseling is also more variable than other forms of therapy (eg, CBT), and at the time the trial was being designed, there was limited direction from clinical manuals. Another necessity for the flexibility of the trial was to ensure that client risk was at the forefront of any design decision as the intervention had not been previously tested. That is, supplementing short-term counseling with guided use of a well-being app in the student context was a new concept, and there was little research or clinical experience that could anticipate risk. It was, therefore, a collective decision not to randomize at this stage, and this decision limits the implications of the subsequent feasibility trial.

Blended face-to-face and app approaches have proven effective in other domains [45], and student counseling seems an appropriate setting in which to develop this approach further. For example, evidence from RCTs has shown internet and app-based interventions to be particularly effective at reducing depressive symptoms [46]. These findings have been demonstrated across a range of technologies and therapeutic approaches including guided internet CBT with text message reminders, internet-delivered mindfulness, and an internet-delivered prevention program comprising self-monitoring and relaxation exercises [47]. Additional lines of research could include the introduction of an app precounseling, which could help to bridge a treatment gap by offering simple guided exercises through the convenience and privacy of a mobile phone. Offering well-being apps while waiting for counseling has the potential to prepare students by providing a means to self-monitor moods, thoughts, and behaviors to take to counseling. Future research would benefit from adopting clear descriptions of blended interventions involving counseling in order to build the evidence base and inform training and guidelines as much of the literature is dominated by blended interventions offering CBT. Regarding the design and implementation of the feasibility trial, future research should consider the need for pragmatic research methodologies to explore outcomes from natural settings that translate to routine practice.

Conclusion

Mobile apps for health and well-being have become ubiquitous and have been developed quicker than research has been able to evaluate their quality [17]. This disconnect has led to a large proportion of apps becoming available without the necessary evidence to validate the quality or efficacy of app content. However, increased awareness of this evidence gap has sparked the development of new quality frameworks, assessment criteria,

and a shared acceptance to co-create new technologies with clinical expertise and embedded evaluations [48]. The future of digital technology is promising for addressing some of the treatment barriers of traditional therapeutic interventions and provides an innovative solution to extend existing services. Offering therapeutic technologies as a low-level preventative measure for subclinical symptoms of depression is particularly promising for the student population and has been shown to be effective. Using apps as an adjunct to counseling also offers a unique solution to addressing some of the attrition issues observed in student counseling services by encouraging self-guided support between face-to-face sessions. This blended approach in our study was shown to be acceptable and feasible and showed potential to maintain clinical improvement on anxiety following the completion of brief counseling.

Acknowledgments

We would like to thank the British Association for Counselling and Psychotherapy (BACP, www.bacp.co.uk) for funding this work as part of a 3-year PhD scholarship awarded to MB and undertaken by EB at the University of Sheffield. All scientific decisions relating to the CASELOAD trial were made independently of the funder. We would like to thank Louise Knowles as Head of Service and Rob Barnsley as Mental Health Support Coordinator at the University of Sheffield Counseling and Psychological Well-Being service for their ongoing collaboration with research activities. We would also like to thank Andy Hill as advisor on the competence framework, Emeritus Professor Michael Campbell as advisor on feasibility trial design, and Dr David Saxon for scientifically reviewing the protocol. Finally, we would like to thank the counselors, administrative team, and students who participated in this study and supported the entire research process.

Conflicts of Interest

EB declares that she is employed by BACP, which previously funded the project as part of her PhD studies undertaken at the University of Sheffield, UK. The funding was awarded to MB and undertaken by EB as part of a 3-year PhD scholarship.

References

1. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J Affect Disord 2015 Mar 01;173:90-96. [doi: 10.1016/j.jad.2014.10.054] [Medline: 25462401]
2. Bruffaerts R, Mortier P, Kiekens G, Auerbach RP, Cuijpers P, Demyttenaere K, et al. Mental health problems in college freshmen: prevalence and academic functioning. J Affect Disord 2018 Jan 01;225:97-103. [doi: 10.1016/j.jad.2017.07.044] [Medline: 28802728]
3. Eisenberg D, Lipson S, Ceglarek P, Kern A, Phillips M. College Student Mental Health: The National Landscape. Promoting Behavioral Health and Reducing Risk among College Students. London: Routledge; 2018.
4. Collins ME, Mowbray CT. Higher education and psychiatric disabilities: national survey of campus disability services. Am J Orthopsychiatry 2005 Apr;75(2):304-315. [doi: 10.1037/0002-9432.75.2.304] [Medline: 15839766]
5. Megivern D, Pellerito S, Mowbray C. Barriers to higher education for individuals with psychiatric disabilities. Psychiatr Rehabil J 2003;26(3):217-231. [Medline: 12653444]
6. Pritchard ME, Wilson GS. Using emotional and social factors to predict student success. J Coll Stud Dev 2003;44(1):18-28. [doi: 10.1353/csd.2003.0008]
7. Storrie K, Ahern K, Tuckett A. A systematic review: students with mental health problems—a growing problem. Int J Nurs Pract 2010 Feb;16(1):1-6. [doi: 10.1111/j.1440-172X.2009.01813.x] [Medline: 20158541]
8. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
9. Ebert DD, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, WHO World Mental Health-International College Student Initiative collaborators. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO World Mental Health International College Student Initiative. Int J Methods Psychiatr Res 2019 Jun;28(2):e1782. [doi: 10.1002/mpr.1782] [Medline: 31069905]
10. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Dec;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
11. Kerr H. National Union of Students mental distress survey. URL: https://www.nus.org.uk/global/campaigns/20130517%20mental%20distress%20survey%20%20overview.pdf [accessed 2019-04-03] [WebCite Cache ID 77MntK2Lw]
12. Byrom N. A summary of the report to HEFCE: understanding provision for students with mental health problems and intensive support needs. URL: https://www.studentminds.org.uk/uploads/3/7/8/4/3784584/summary_of_the_hefce_report.pdf [WebCite Cache ID 77Mo352hr]

13. Broglia E, Millings A, Barkham M. Challenges to addressing student mental health in embedded counselling services: a survey of UK higher and further education institutions. Br J Guid Counc 2017 Sep 29;46(4):441-455. [doi: 10.1080/03069885.2017.1370695]
14. Benton SA, Robertson JM, Tseng W, Newton FB, Benton SL. Changes in counseling center client problems across 13 years. Prof Psychol Res Pr 2003;34(1):66-72. [doi: 10.1037/0735-7028.34.1.66]
15. Mair D. Short-Term Counselling in Higher Education: Context, Theory and Practice. London: Routledge; 2015.
16. Goss S, Anthony K. Developments in the use of technology in counselling and psychotherapy. Br J Guid Counc 2009 Aug;37(3):223-230. [doi: 10.1080/03069880902956967]
17. Bennion MR, Hardy G, Moore RK, Millings A. E-therapies in England for stress, anxiety or depression: what is being used in the NHS? A survey of mental health services. BMJ Open 2017 Jan 23;7(1):e014844 [FREE Full text] [doi: 10.1136/bmjopen-2016-014844] [Medline: 28115336]
18. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
19. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
20. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. Psychol Med 2018 Nov;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
21. Kählke F, Berger T, Schulz A, Baumeister H, Berking M, Cuijpers P, et al. Efficacy and cost-effectiveness of an unguided, internet-based self-help intervention for social anxiety disorder in university students: protocol of a randomized controlled trial. BMC Psychiatry 2019 Jun 25;19(1):197 [FREE Full text] [doi: 10.1186/s12888-019-2125-4] [Medline: 31238907]
22. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759. [doi: 10.1002/mpr.1759] [Medline: 30585363]
23. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
24. 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]
25. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
26. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR Mhealth Uhealth 2018 Jun 22;6(6):e10253 [FREE Full text] [doi: 10.2196/10253] [Medline: 29934285]
27. East ML, Havard BC. Mental health mobile apps: from infusion to diffusion in the mental health social system. JMIR Ment Health 2015;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]
28. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
29. Broglia E, Millings A, Barkham M. Comparing counselling alone versus counselling supplemented with guided use of a well-being app for university students experiencing anxiety or depression (CASELOAD): protocol for a feasibility trial. Pilot Feasibility Stud 2017;3:1-15 [FREE Full text] [doi: 10.1186/s40814-016-0119-2] [Medline: 28127442]
30. 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]
31. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA 1999 Nov 10;282(18):1737-1744. [Medline: 10568646]
32. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
33. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008 Jan;58(546):32-36. [doi: 10.3399/bjgp08X263794] [Medline: 18186994]
34. Barkham M, Bewick B, Mullin T, Gilbody S, Connell J, Cahill J, et al. The CORE-10: a short measure of psychological distress for routine use in the psychological therapies. Couns Psychother Res 2013 Mar;13(1):3-13. [doi: 10.1080/14733145.2012.729069]
35. Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. Br J Psychiatry 2002 Jan;180:51-60. [Medline: 11772852]

36. Locke BD, Buzolitz JS, Lei P, Boswell JF, McAleavey AA, Sevig TD, et al. Development of the Counseling Center Assessment of Psychological Symptoms-62 (CCAPS-62). J Couns Psychol 2011 Jan;58(1):97-109. [doi: 10.1037/a0021282] [Medline: 21133541]
37. Broglia E, Millings A, Barkham M. The Counseling Center Assessment of Psychological Symptoms (CCAPS-62): acceptance, feasibility, and initial psychometric properties in a UK student population. Clin Psychol Psychother 2017 Sep;24(5):1178-1188. [doi: 10.1002/cpp.2070] [Medline: 28124512]
38. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
39. Pacifica: the #1 app for stress, anxiety, & depression. URL: https://www.thinkpacifica.com/ [accessed 2019-04-03] [WebCite Cache ID 77MoLFo7F]
40. Hill A, Roth T. Competences required to deliver effective counseling in further education. URL: https://www.bacp.co.uk/events-and-resources/ethics-and-standards/competences-and-curricula/ [accessed 2019-04-03] [WebCite Cache ID 77MoVxu5s]
41. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
42. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977 Jun;33(2):363-374. [Medline: 884196]
43. Lindhiem O, Bennett CB, Trentacosta CJ, McLear C. Client preferences affect treatment satisfaction, completion, and clinical outcome: a meta-analysis. Clin Psychol Rev 2014 Aug;34(6):506-517 [FREE Full text] [doi: 10.1016/j.cpr.2014.06.002] [Medline: 25189522]
44. Steinert C, Stadter K, Stark R, Leichsenring F. The effects of waiting for treatment: a meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. Clin Psychol Psychother 2017 May;24(3):649-660. [doi: 10.1002/cpp.2032]
45. Vollert B, Beintner I, Musiat P, Gordon G, Görlich D, Nacke B, et al. Using internet-based self-help to bridge waiting time for face-to-face outpatient treatment for bulimia nervosa, binge eating disorder and related disorders: study protocol of a randomized controlled trial. Internet Interv 2019 Apr;16:26-34 [FREE Full text] [doi: 10.1016/j.invent.2018.02.010] [Medline: 30775262]
46. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
47. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of Project UPLIFT: distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
48. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98. [doi: 10.1002/wps.20592] [Medline: 30600619]

Abbreviations

BACP: British Association for Counselling and Psychotherapy
CASELOAD: Counseling plus Apps for Students Experiencing Levels Of Anxiety or Depression
CCAPS-62: Counseling Center Assessment of Psychological Symptoms, 62-items
CORE-10: Clinical Outcomes in Routine Evaluation Outcome Measure, 10-item
CORE-OM: Clinical Outcomes in Routine Evaluation–Outcome Measure
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
GAD-7: Generalized Anxiety Disorder, 7-item
iCBT: internet-based cognitive behavioral therapy
NHS: National Health Service
PHQ-9: Physical Health Questionnaire, 9-item
RCSI: reliable and clinically significant improvement
RCT: randomized controlled trial
SAM: Self-help for Anxiety Management
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
TAU: treatment as usual

Edited by G Eysenbach; submitted 15.04.19; peer-reviewed by F Kählke, M Macalli, M Reynolds; comments to author 25.06.19; revised version received 09.07.19; accepted 19.07.19; published 15.08.19.

Please cite as:
Broglia E, Millings A, Barkham M
Counseling With Guided Use of a Mobile Well-Being App for Students Experiencing Anxiety or Depression: Clinical Outcomes of a Feasibility Trial Embedded in a Student Counseling Service
JMIR Mhealth Uhealth 2019;7(8):e14318
URL: http://mhealth.jmir.org/2019/8/e14318/
doi:10.2196/14318
PMID:31418424

Original Paper

A Mobile Phone App Featuring Cue Exposure Therapy As Aftercare for Alcohol Use Disorders: An Investigator-Blinded Randomized Controlled Trial

Angelina Isabella Mellentin[1,2,3], PhD; Bent Nielsen[1], MD, PhD; Anette Søgaard Nielsen[1,2], PhD; Fei Yu[4], PhD; Anna Mejldal[1], MSc; Dorthe Grüner Nielsen[1], MD; Elsebeth Stenager[2,3], MD, PhD

[1]Unit for Clinical Alcohol Research, Unit for Psychiatric Research, Department of Clinical Research, University of Southern Denmark, Odense Center, Denmark

[2]Brain Research-Inter-Disciplinary Guided Excellence, Department of Clinical Research, University of Southern Denmark, Odense Center, Denmark

[3]Unit for Psychiatric Research, Institute of Regional Health Services Research, University of Southern Denmark, Aabenraa, Denmark

[4]Technology Entrepreneurship and Innovation section, Mads Clausen Institute, University of Southern Denmark, Sønderborg, Denmark

Corresponding Author:
Angelina Isabella Mellentin, PhD
Unit for Clinical Alcohol Research, Unit for Psychiatric Research
Department of Clinical Research
University of Southern Denmark
J.B. Winsløwsvej 18
Odense Center, 5000
Denmark
Phone: 45 50517901
Email: amellentin@health.sdu.dk

Abstract

Background: Cue exposure therapy (CET) is a psychological approach developed to prepare individuals with alcohol use disorder (AUD) for confronting alcohol and associated stimuli in real life. CET has shown promise when treating AUD in group sessions, but it is unknown whether progressing from group sessions to using a mobile phone app is an effective delivery pathway.

Objective: The objectives of this study were to investigate (1) whether CET as aftercare would increase the effectiveness of primary treatment with cognitive behavior therapy, and (2) whether CET delivered through a mobile phone app would be similarly effective to CET via group sessions.

Methods: A total of 164 individuals with AUD were randomized to one of three groups: CET as group aftercare (CET group), CET as fully automated mobile phone app aftercare (CET app), or aftercare as usual. Study outcomes were assessed face-to-face at preaftercare, postaftercare, and again at 6 months after aftercare treatment. Generalized mixed models were used to compare the trajectories of the groups over time on drinking, cravings, and use of urge-specific coping skills (USCS).

Results: In all, 153 of 164 individuals (93%) completed assessments both at posttreatment and 6-month follow-up assessments. No differences in the trajectories of predicted means were found between the experimental groups (CET group and app) compared with aftercare as usual on drinking and craving outcomes over time. Both CET group (predicted mean difference 5.99, SE 2.59, z=2.31, P=.02) and the CET app (predicted mean difference 4.90, SE 2.26, z=2.31, P=.02) showed increased use of USCS compared to aftercare as usual at posttreatment, but this effect was reduced at the 6-month follow-up. No differences were detected between the two experimental CET groups on any outcomes.

Conclusions: CET with USCS delivered as aftercare either via group sessions or a mobile phone app did not increase the effectiveness of primary treatment. This suggests that CET with USCS may not be an effective psychological approach for the aftercare of individuals treated for AUD.

Trial Registration: ClinicalTrials.gov NCT02298751; https://clinicaltrials.gov/ct2/show/NCT02298751

(JMIR Mhealth Uhealth 2019;7(8):e13793) doi:10.2196/13793

KEYWORDS

alcohol use disorder; aftercare; cue exposure therapy; cognitive behavior therapy; randomized controlled trial; mobile phone app

Introduction

Background

In Western societies, individuals with an alcohol use disorder (AUD) are constantly exposed to alcohol and associated cues in their everyday life. Even when successfully treated with evidence-based psychological approaches, such as cognitive behavioral therapy, this pervasive exposure can induce cue-controlled cravings and lead to relapses, with devastating personal, familial, and socioeconomic consequences and increased burden on national health care resources.

Cue exposure therapy (CET) is a behavioristic, psychological approach to treating AUD that aims to reduce cue-induced cravings by repeatedly exposing individuals with AUD to relevant alcohol cues and hindering their habitual drinking response [1-5]. AUD individuals can thus reduce their cue reactivity and be better prepared to navigate in society.

Although CET seems to target one of the main challenges for relapse, a recent meta-analysis of controlled trials of CET for the treatment of AUD found no effect to small overall additional effects when CET was compared with other active control conditions. However, analysis of a priori defined trial covariates indicated that the type of CET and active comparison might be crucial for its effectiveness. CET combined with urge-specific coping skills (USCS) proved to be a better option for treating AUD than conventional CET [6]. During CET with USCS, individuals are initially taught coping skills and are then exposed to their preferred alcohol to generate cue-induced cravings. When the cravings peak in intensity during the exposure, the individual is actively encouraged to apply a USCS to reduce the cravings to a manageable level. This is in contrast to more conventional CET approaches in which cravings are expected to decrease without the use of USCS [7,8]. The meta-analysis also indicated that CET may prove more effective when compared with other active control conditions than cognitive behavioral therapy [6]. This is expected because cognitive behavioral therapy is one of the most effective evidence-based psychological interventions for AUD [9-12]. Furthermore, cognitive behavioral therapy has much in common with CET, especially when combined with USCS (eg, thinking about negative or positive consequences of alcohol consumption), and often integrates CET when treating other psychiatric disorders (eg, [13-15]). Currently, cognitive behavioral therapy and CET are segregated when targeting AUD and other substance use disorders, in which the emphasized difference is the in vivo exposure element featured in CET. No study has compared CET with USCS to cognitive behavioral therapy, and it is therefore not possible to disentangle the effects of type of CET from the type of comparison treatment. An important research question is whether CET combined with USCS increases the effectiveness of cognitive behavioral therapy as patients can practice coping skills while they are exposed to alcohol cues in vivo and thus experiencing cue-induced cravings.

When offering evidence-based primary and add-on aftercare AUD treatments, such as CET with USCS, the duration of the treatment is often shortened and may even be performed in group settings rather than individual sessions [10]. Due to the heavy burden on health care resources, this in itself may be an advantage of implementing evidence-based treatment strategies. Moreover, we may be experiencing another paradigm change in treatment delivery pathways, progressing from individual and group sessions to eHealth interventions through mobile devices [16-20], which could also lower the costs of treatment. Group therapies are effective for many psychological approaches [21], but little is known about the effectiveness of psychological interventions delivered through eHealth interventions, such as computers, tablets, and mobile phones [17,21-24]—and especially when delivered through a mobile phone app [23,25-27]. Preliminary evidence indicates that CET is equally effective in group settings as in individual sessions [28-30], but we do not know if a mobile phone app featuring CET with USCS would be as effective as CET delivered in group sessions.

Objectives

The objectives of this study are two-fold: (1) to investigate whether CET with USCS (based on a published treatment manual) delivered as aftercare would increase the effectiveness of cognitive behavioral therapy compared with aftercare as usual and (2) to investigate whether CET with USCS delivered through a mobile phone app would be noninferior to CET with USCS via group sessions. In light of available evidence, it was hypothesized that the experimental CET aftercare groups would achieve superior outcomes (such as alcohol consumption, urges or cravings, and coping skills) compared with active controls treated with aftercare as usual. It was an explorative research question whether a CET mobile phone app was noninferior to the group therapy.

Methods

The trial was registered in ClinicalTrials.gov (ID: NCT02298751) on November 6, 2014, and was conducted based on the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) statement [31,32]. The CONSORT flow diagram is presented in Multimedia Appendix 1. The study methods are presented in more detail elsewhere [33,34].

Study Design and Setting

The CET aftercare study was conducted as a single-site, investigator-blinded, parallel randomized controlled trial (RCT) in an outpatient alcohol treatment clinic in Funen, Denmark. Most individuals with AUD are offered outpatient treatment when seeking treatment in Denmark. The public outpatient treatment is paid via taxes and is open for self-referral, and patients can remain anonymous during treatment [35]. Only alcohol problems are treated at the outpatient clinic; individuals with mainly illegal drug abuse are treated elsewhere. Individuals with alcohol and/or illegal drug abuse in combination with major psychiatric disorders (schizophrenia, bipolar disorders) are also treated elsewhere.

Primary Treatment at the Outpatient Alcohol Clinic

The primary treatment lasts 3 months and consists of both pharmacological and psychological treatment. At treatment entry, the individual is offered detoxification, if this is needed, and other pharmacological treatment (eg, disulfiram,

acamprosate, naltrexone) as appropriate. The psychological treatment consists of cognitive behavioral therapy provided during 1-hour individual or group sessions and usually consists of eight sessions. The treatment course is planned together with the individual, and the therapy typically incorporates psychoeducation, functional analysis of drinking situations, development of coping strategies (eg, waiting out until the urges pass, thinking about the negative consequences of drinking, thinking about positive consequences of sobriety, and intake of alternative food and beverage), problem solving, and homework between sessions. The psychological treatments are delivered by therapists who are nurses and social workers educated within the treatment range. Supervision is frequent, and psychiatrists regularly monitor the treatment course [36].

Recruitment

Study participants were recruited from June 1, 2015, to June 1, 2017. Shortly before concluding the standard 3-month primary treatment at the outpatient clinic, the individuals were briefly informed about the aftercare project. The information was given in the second-last session of the treatment, during which the patients were given written information about the aftercare study and were asked by the therapist if they would be willing to meet with a research assistant immediately after the end of treatment to learn more about the aftercare project. Willing individuals then received additional oral and written project information from the research assistant. After informed consent was obtained, the preaftercare interview was carried out, and individuals fulfilling the eligibility criteria were randomized to one of the three aftercare treatment groups described subsequently.

Eligibility Criteria

To be eligible for inclusion in the study, individuals had to provide informed written consent, be aged between 18 and 80 years, and have completed primary treatment. Individuals who did not speak Danish, had an acute psychotic disorder, severe cognitive impairment, or terminal somatic illness were not eligible to participate.

Randomization

Randomization occurred using computerized urn randomization. To ensure adequate allocation concealment, the random allocation sequence was generated by a statistician so that the research team was not involved in generating the sequence. In contrast to simple or block randomization, urn randomization is dynamic, and the probability of treatment assignment changes depending on the degree of treatment imbalance throughout the trial. Increased probability of randomization to the groups with the least number of participants increases systematically. This ensures allocation balance throughout the study and random allocation of covariates.

Experimental and Control Aftercare Groups

Individuals who fulfilled the eligibility criteria were randomized to one of three aftercare treatment groups: (1) CET as group aftercare (CET group), (2) CET as a mobile phone app aftercare (CET app), or (3) aftercare as usual.

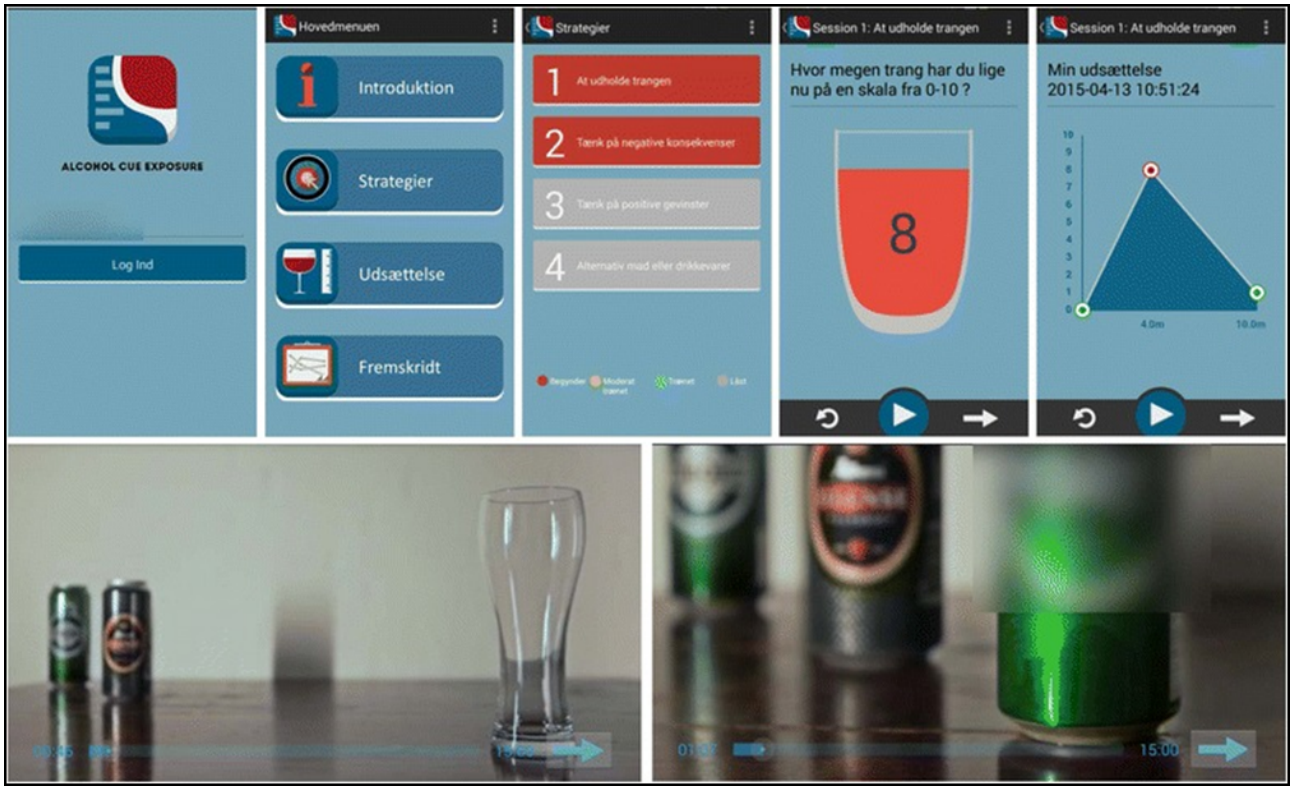
Cue Exposure Therapy Aftercare Delivered as Group Therapy

The CET aftercare in groups was conducted according to Monty and coworkers' treatment manual for CET with USCS [28], which emphasizes the importance of individuals being confronted with alcohol to reduce cue-induced cravings. During each CET session, the individual is introduced or reintroduced to effective USCS and afterward required to practice the learned strategies while exposed to alcohol in vivo. The recommended coping strategies are (1) waiting it out, (2) thinking about the negative consequences of drinking, (3) thinking about the positive consequences of sobriety, and (4) alternative food and beverage intake. Participants were required to turn up for therapy every other week for 8 weeks (four sessions of 120 minutes each, with a maximum of eight patients in each group). Similar to cognitive behavioral therapy, the aftercare treatment was delivered by therapists (nurses and social workers trained for the purpose), and frequent supervision was conducted by a psychiatrist specialized in psychotherapy throughout the treatment phase of the study. Fidelity to the aftercare treatment manual was ensured by assessing and analyzing a 10% random sample of audio-recorded group treatment sessions.

Cue Exposure Therapy Aftercare Via a Mobile Phone App

Based on the same treatment manual as the group CET approach, the CET aftercare intervention was transformed into a fully automated mobile phone app (see Figure 1).

Figure 1. Illustration of the mobile phone app for cue exposure therapy.



The app was individually adaptive in terms of the featured coping strategies and the alcohol exposure material. Exposure to alcohol was simulated by watching one of eight alcohol videos on the mobile phone (eg, beer, red or white wine, mixed alcoholic drinks, hard liquor), which allowed individuals to select their preferred beverage as the exposure material. The alcohol exposure videos imitated sessions with a therapist, and the alcohol presented in the videos became increasingly appetitive during the exposure session to induce cue-controlled cravings. The app contained a direct phone number to a CET therapist in case of uncontrollable cravings, and the app was only accessible during the opening hours of the alcohol outpatient clinic (Monday to Friday from 9 am to 6 pm) so that the patient could meet the therapist and get help to control the cravings. Individuals could practice exposure once a day, four times a week (a maximum of 32 sessions of approximately 15 minutes each), and they were supposed to receive a reminder every week. Due to technical issues, the software failed to send text messages to remind patients to use the app on a regular basis.

Aftercare as Usual

Aftercare as usual consisted of one individual follow-up session during which individuals were asked how they were doing and, if needed, were offered a brush-up of the coping strategies taught during primary treatment with cognitive behavioral therapy. The aftercare as usual session was offered 8 weeks after discharge from primary treatment, and the session lasted 60 minutes.

Measures

Primary Treatment Measures

Demographic data were recorded for all patients when they entered primary alcohol treatment. The AUD assessment using the *International Classification of Diseases, Tenth Revision* Diagnostic Criteria for Research [37,38]) and the Addiction Severity Index [39,40] formed part of the routine clinical investigation and was carried out for all individuals at both the start and completion of the primary treatment. The Addiction Severity Index measured domains related to addictive behavior, including an alcohol domain indicating the severity and remission state of the AUD [39,41].

Cue Exposure Therapy Aftercare Outcome Measures

After primary treatment and before randomization, individuals fulfilling the eligibility criteria were further assessed on measures aimed at assessing primary and secondary outcomes to determine the effects of CET aftercare treatment. Data were collected at preaftercare (ie, after primary treatment but before entering aftercare), after 2 months (at postaftercare treatment), and again at 6 months (completion of aftercare treatment).

Primary Study Outcomes: Alcohol Consumption

Alcohol consumption measures were derived from Timeline Follow Back, which involves using a calendar to identify alcohol consumption patterns in the last 30 days [42-44]. The following variables were derived from this measure: sensible drinking (drinking ≤14 drinks per week for women and ≤21 drinks per week for men, as well as ≤5 drinks per day for both genders as recommended by the Danish Health Authorities [45]), abstinence, drinking days, and days with excessive drinking

(drinking >5 drinks per day, where one drink contains 12 grams of ethanol) in the past 30 days.

Secondary Study Outcomes: Cravings and Urge-Specific Coping Skills

The Visual Analog Scale (VAS) comprises single items used to measure the degree of alcohol cravings on scales ranging from 0 to 10, with 0 representing no cravings and 10 representing extreme cravings. The scale is presented visually on a ruler; individuals were requested to report the mean level and the peak level of cravings experienced during the last 30 days [46-48].

The 22-item Urge-Specific Coping Skills Questionnaire (USCSQ) assesses 11 coping strategies that the patient may currently be using and the effectiveness of each of these when experiencing cravings or urges to drink. The strategies include those taught in primary cognitive behavioral therapy treatment and/or the CET intervention groups. Items are rated on scales ranging from 0 (never) to 10 (always) [7].

Although the routine assessments were conducted by the outpatient clinic staff, the primary and secondary study outcomes were assessed by research assistants. Attempts were made to retain participants in posttreatment assessments, even if they withdrew from aftercare.

Real-Time Measures of Cue-Induced Cravings

Participants in the two experimental CET arms were also asked to rate the degree of real-time cue-induced cravings experienced. Urges were measured on scales ranging from 0 (no urges) to 10 (severe urges) at three different time points: (1) at baseline (before exposure), (2) when the urge was expected to peak (during exposure), and (3) at the endpoint (after exposure) [7]. These real-time cue-induced urges were assessed by therapists for CET group and by software for CET app.

Data Analysis

Sample characteristics were described for sociodemographic, primary, and secondary measures at baseline using frequencies for categorical variables, mean and standard deviation (SD) for normally distributed variables, and median and interquartile range (IQR) for nonnormally distributed variables.

Real-Time Cue-Induced Cravings

The real-time cue-induced craving measures were analyzed using Wilcoxon rank test (Mann-Whitney) to determine whether the experimental groups differed from each other. Proxy measures were calculated for the intensity of the urge induced by the selected exposure video and the effectiveness of the selected USCS in reducing the urge. The first measure was calculated by subtracting the baseline measure from the peak measure. The effectiveness of the USCS was calculated by subtracting the endpoint measure from the peak measure.

Aftercare Treatment Outcomes

Generalized linear mixed models were used to examine the trajectories of the primary and secondary outcomes by group allocation over time. Specifically, we wanted to examine how alcohol consumption, cravings, and USCS were influenced by the variable time point (preaftercare, postaftercare, and 6-month follow-up) and the interaction between time point and group (CET group, CET app, and aftercare as usual).

The alcohol consumption outcomes for sensible drinking and abstinence were coded dichotomously and analyzed with mixed-effects logistic regression models. Drinking days and days with excessive drinking met the assumptions for mixed-effects linear regression models and were analyzed accordingly. Similarly, cravings and USCS were analyzed with mixed-effects linear regression models.

Fixed effects consisted of time point and group×time point interaction. Because patients were randomized into three groups, this interaction tests for the existence of a treatment effect over time. All models included a subject-specific random intercept and allowed for a subject-specific random slope over time. Assuming the dropout mechanism is missing at random, these models deal efficiently with missing values due to dropout using the maximum likelihood estimator (missing were 7%, n=11 in both the first and second follow-ups). Therefore, with the mixed-effects model approach, all available data were used. If there was an overall significant interaction effect between group and any one time point, we examined whether the change over time differed between the groups and the time point using contrasts.

Two analyses were performed for each outcome measure: (1) intention-to-treat analysis (ie, irrespective of whether individuals had completed the interventions or were reinterviewed), and (2) completer (on-treatment) analyses for AUD individuals who had completed the respective intervention (completed all four CET with USCS sessions at least once or completed the aftercare as usual). The significance level was set at 5%, and two-sided analyses were conducted. All analyses were conducted using Stata version 15.

Sample Size

The sample size was calculated based on sensible drinking, which represents abstinence and not drinking excessively at any occasion or over time (ie, ≤5 standard drinks on one occasion, or drinking ≤14 drinks for women or ≤21 drinks for men per week). The power calculation was estimated from quality assurance and research data from the outpatient alcohol treatment clinic participating in this study. With the current treatment regimen, 65% of the patients have sensible drinking habits 6 months after treatment (information stemming from local continuous monitoring of the treatment quality). To detect an effect by comparing the three groups, a sample of 100 patients in each group was needed for 90% power of detecting a difference corresponding to an improvement of 18% using a 5% level of statistical significance.

Statement of Ethics

The study protocol was approved by the Regional Scientific Ethical Committees for Southern Denmark (Project-ID S-20140176) and was conducted according to the World Medical Association Declaration of Helsinki.

Results

Sample Characteristics

During the inclusion period, 323 individuals with AUD who fulfilled the eligibility criteria concluded primary treatment and were offered participation in the aftercare treatment study. Of these, 159 individuals declined to participate, and 164 (51%) were enrolled in the study and completed preaftercare assessment (see Multimedia Appendix 1). A total of 153 of 164 (93%) individuals completed the postaftercare assessment and the 6-month follow-up: 94% (51/54) in CET group, 91% (49/54) in CET app, and 95% (53/56) in aftercare as usual.

As shown in Table 1, approximately 70% of the sample was relatively well-educated having completed either vocational training, a bachelor degree at vocational academies or university colleges (≤4 years), or a university degree or other higher education (>4 years) after finishing elementary school or high school. Approximately 50% of the sample were employed, and 10% were students receiving grants, state loans, and employment income. Approximately 35% were pensioned (mainly due to retirement), and the rest of the sample was temporarily out of employment, on sickness benefit, unemployment benefit, or cash assistance.

The preaftercare alcohol consumption measures indicated that the sample was successfully treated during the primary treatment course: 80% (132/164) achieved sensible drinking and 70% (115/164) achieved total abstinence. Among individuals who reported being nonabstinent (n=46), the median number of drinking days was 3 (IQR 4), and days with excessive drinking was median 1.5 (IQR 2). In addition, the level of cravings ranged from low to moderate on the VAS. Before initiation of the aftercare interventions, the use and perceived effectiveness of the USCS were high, particularly for “thinking about the negative consequences of drinking” and “thinking about the positive consequences of sobriety.” This could be expected given that the primary psychological treatment was cognitive behavioral therapy.

Aftercare Treatment Retention

During the aftercare intervention, 81% (44/54) of the individuals attended at least one CET group session, 78% (42/54) attended two CET group sessions, 74% (40/54) attended three CET group sessions, and 67% (36/54) participated in all four group sessions. Similarly, 78% (42/54) completed at least one CET app session, 72% (39/54) completed two CET app sessions, 53% (29/54) completed CET app sessions, and 41% (22/54) completed at least four CET app sessions. In total, 73% (41/56) completed aftercare as usual.

Real-Time Cue-Induced Cravings

In line with the low level of cravings as measured by the VAS, a low intensity of real-time cue-induced cravings was observed during CET with USCS treatment. From the beginning of the alcohol exposure session until the urge intensity was supposed to peak, the CET group reported a median increase of 2.25 (IQR 2.46); when applying the USCS from the peak urge until the session ended, the median reduction was 2.25 (IQR 2.88). The CET app group showed a median increase of 0.19 (IQR 2.08) from the onset of exposure to the peak measure, followed by a median reduction of 0.37 (IQR 2.0).

Aftercare Treatment Outcomes

As shown in Table 2, there were significant effects of time on all primary outcomes. Although the decrease in predicted mean sensible drinking from preaftercare to postaftercare was not significant (predicted mean change [PMC] −0.06, SE 0.04, z=−1.62, P=.11), the decrease in sensible drinking from preaftercare to 6-month follow-up was significant (PMC −0.17, SE 0.05, z=−3.07, P=.002) as was the decrease in abstinence from both preaftercare to postaftercare (PMC −0.12, SE 0.04, z=−3.15, P=.002) and to 6-month follow-up (PMC −0.27, SE 0.06, z=−4.92, P<.001). In line with these findings, fwe found significant increases in drinking days from preaftercare to postaftercare (PMC 1.18, SE 0.43, z=2.74, P=.006) and to 6-month follow-up (PMC 3.08, SE 0.61, z=5.03, P<.001), as well as in days with excessive drinking from preaftercare to postaftercare (PMC 1.03, SE 0.33, z=3.08, P=.002) and to 6-month follow-up (PM change 2.13, SE 0.46, z=4.60, P<.001).

Taken together, these changes showed that the total sample increased their drinking following treatment conclusion and following aftercare, and even began to relapse.

Table 3 and Figures 2 and 3 illustrate the intention-to-treat analyses for the trajectories of primary and secondary outcomes by group allocation over time.

Table 1. Sample characteristics before aftercare treatment with cue exposure treatment (CET) or aftercare as usual (N=164).

Sample characteristics	CET app (n=54)	CET group (n=54)	Aftercare as usual (n=56)
Demographics			
Age (years), mean (SD)	46 (14)	48 (13)	45 (12)
Age (years), range	19-68	18-80	23-69
Male, n (%)	39 (72)	45 (83)	43 (77)
Further education, n (%)			
None	16 (30)	11 (21)	10 (19)
≤4 years	34 (63)	37 (71)	39 (74)
>4 years	4 (7)	4 (8)	4 (8)
Income, n (%)			
Employed	29 (54)	24 (44)	27 (48)
Temporarily unemployed	12 (22)	10 (19)	18 (32)
Student	3 (6)	6 (10)	5 (9)
Pensioned	12 (22)	16 (30)	5 (9)
Disulfiram, n (%)	39 (21)	28 (15)	34 (19)
Alcohol consumption			
Sensible drinking, n (%)	40 (74)	48 (89)	40 (71)
Abstinence, n (%)	39 (72)	46 (80)	36 (64)
Drinking days, median (IQR[a])	0 (2)	0 (0)	0 (2)
Days with excessive drinking, median (IQR)	0 (1)	0 (0)	0 (1)
Alcohol cravings (Visual Analog Scale), median (IQR)			
Highest urge	3 (7)	2 (6)	4 (8)
Mean urge	2 (3)	1 (4)	3 (5)
Urge-specific coping skills (USCSQ[b]), median (IQR)			
Use			
Waiting it out	3 (8)	4 (9)	5 (9)
Thinking about negative consequences	5 (8)	7 (7)	8 (6)
Thinking about the positive consequences	8 (8)	8 (8)	8 (5)
Alternative food and beverage intake	0 (0)	0 (0)	0 (1)
Effectiveness			
Waiting it out	5 (10)	4.5 (9)	5 (9)
Thinking about the negative consequences	8 (8)	8 (5)	8 (6)
Thinking about the positive consequences	9 (5)	8 (5)	8 (5)
Alternative food and beverage intake	0 (0)	0 (0)	0 (1)

[a]IQR: interquartile range.

[b]USCSQ: urge-specific copings skills questionnaire.

Table 2. Predicted means of primary and secondary outcomes over time.

Time point	Overall, PM[a] (SE)	Overall change from pre-treatment, PMC[b] (SE)	Time	
			z	*P* value
Primary outcomes				
Sensible drinking				
Pretreatment	0.77 (0.03)	—[c]	—	—
Posttreatment	0.71 (0.04)	−0.06 (0.04)	−1.62	.11
6-month follow-up	0.61 (0.06)	−0.17 (0.05)	−3.07	.002
Abstinence				
Pretreatment	0.71 (0.03)	—	—	—
Posttreatment	0.60 (0.05)	−0.12 (0.04)	−3.15	.002
6-month follow-up	0.44 (0.06)	0.27 (0.06)	−4.92	<.001
Drinking days				
Pretreatment	1.65 (0.34)	—	—	—
Posttreatment	2.83 (0.46)	1.18 (0.43)	2.74	.006
6-month follow-up	4.73 (0.63)	3.08 (0.61)	5.03	<.001
Days with excessive drinking				
Pretreatment	0.66 (0.13)	—	—	—
Posttreatment	1.69 (0.35)	1.03 (0.33)	3.08	.002
6-month follow-up	2.80 (0.48)	2.13 (0.46)	4.60	<.001
Secondary outcomes				
VAS[d] mean				
Pretreatment	2.40 (0.18)	—	—	—
Posttreatment	2.45 (0.20)	0.05 (0.20)	0.23	.82
6-month follow-up	2.56 (0.21)	0.15 (0.22)	0.68	.50
VAS peak				
Pretreatment	3.75 (0.27)	—	—	—
Posttreatment	4.23 (0.28)	0.48 (0.26)	1.88	.06
6-month follow-up	3.85 (0.28)	0.10 (0.29)	0.33	.74
USCS[e] use				
Pretreatment	20.02 (0.93)	—	—	—
Posttreatment	19.47 (0.95)	−0.55 (1.01)	−0.54	.59
6-month follow-up	15.66 (1.02)	−4.36 (1.02)	−4.27	<.001
USCS effectiveness				
Pretreatment	22.04 (0.95)	—	—	—
Posttreatment	21.05 (1.03)	−0.99 (1.10)	−0.90	.37
6-month follow-up	16.41 (1.06)	−5.63 (1.22)	−4.60	<.001

[a]PM: predicted mean.

[b]PMC: predicted mean change.

[c]Not applicable.

[d]VAS: Visual Analog Scale (alcohol cravings).

[e]USCS: urge-specific coping skills.

Table 3. Predicted mean differences of primary and secondary outcomes over time and by group allocation.

Time point	CET[a] app/AAU[b], PMD[c] (SE)	Time × group		CET group/AAU, PMD (SE)	Time × group		CET app/CET group, PMD (SE)	Time × group	
		z	*P* value		*z*	*P* value		*z*	*P* value
Primary outcomes									
Sensible drinking									
Posttreatment	0.06 (0.09)	0.63	.53	0.03 (0.10)	0.31	.76	0.02 (0.09)	0.25	.78
6-month follow-up	0.02 (0.12)	0.17	.86	−0.07 (0.13)	−0.57	.57	0.09 (0.12)	0.78	.44
Abstinence									
Posttreatment	0.04 (0.08)	0.45	.65	0.08 (0.09)	0.97	.33	−0.05 (0.08)	−0.58	.56
6-month follow-up	0.07 (0.10)	0.69	.49	−0.01 (0.09)	−0.11	.91	0.08 (0.11)	0.69	.49
Drinking days									
Posttreatment	−0.23 (1.10)	−0.21	.83	−0.53 (1.02)	−0.52	.61	0.30 (1.05)	0.28	.78
6-month follow-up	−0.30 (1.56)	−0.19	.85	1.61 (1.44)	1.12	.26	−1.91 (1.50)	−1.27	.20
Days with excessive drinking									
Posttreatment	−0.82 (0.76)	−1.08	.28	−0.78 (0.89)	−0.87	.39	−0.05 (0.78)	−0.06	.95
6-month follow-up	−0.89 (1.19)	−0.74	.46	0.13 (1.12)	0.12	.91	−1.02 (1.07)	−0.95	.34
Secondary outcomes									
VAS[d] mean									
Posttreatment	0.19 (0.51)	0.37	.71	0.45 (0.47)	0.95	.34	−0.26 (0.50)	−0.52	.60
6-month follow-up	0.19 (0.58)	0.33	.74	−0.23 (0.52)	−0.44	.66	0.42 (0.56)	0.75	.45
VAS peak									
Posttreatment	−0.18 (0.69)	−0.26	.80	0.77 (0.58)	1.34	.18	−0.94 (0.62)	−1.54	.12
6-month follow-up	−0.38 (0.74)	−0.51	.61	−0.93 (0.70)	−1.32	.19	0.55 (0.72)	0.76	.44
USCS[e] use									
Posttreatment	4.90 (2.26)	2.17	.03	5.99 (2.59)	2.31	.02	−1.09 (2.52)	−0.43	.67
6-month follow-up	1.22 (2.29)	0.53	.60	0.29 (2.65)	0.11	.91	0.93 (2.54)	0.37	.71
USCS effectiveness									
Posttreatment	4.35 (2.48)	1.75	.08	5.05 (2.93)	1.72	.09	−0.70 (2.65)	−0.26	.79
6-month follow-up	0.47 (2.76)	0.17	.86	−0.06 (3.16)	−0.02	.99	0.53 (3.03)	0.18	.86

[a]CET: cue exposure therapy.

[b]AAU: aftercare as usual.

[c]PMD: predicted mean difference.

[d]VAS: Visual Analog Scale (alcohol cravings).

[e]USCS: urge-specific coping skills.

For the primary outcomes, Table 3 and Figure 2 show that the trajectories and the intention-to-treat analysis with generalized linear mixed models detected no interactions between time point and group. That is, there were no predicted mean differences either when the two experimental CET groups were compared with aftercare as usual or when CET app was compared with CET group over time.

For the secondary outcomes, Table 3 and Figure 3 show that no interactions were detected on alcohol cravings (mean urge and highest urge) between the groups over time. However, an interaction was detected on USCS revealing that both CET app (PMD 4.90, SE 2.26, z=2.17, P=.03) and CET group (PMD 5.99, SE 2.59, z=2.31, P=.02) applied the coping strategies more than the active controls at postaftercare, but this effect was lower at the 6-month follow-up.

Figure 2. Primary outcomes by group allocation over time among individuals receiving cue exposure therapy (CET) as group aftercare or as a mobile phone app, or aftercare as usual.

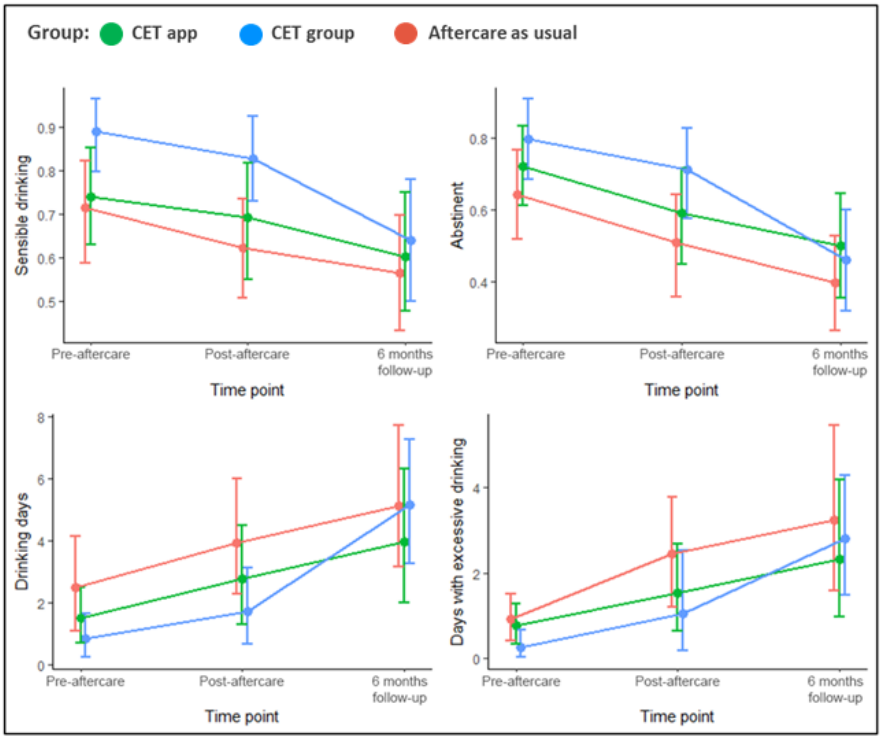
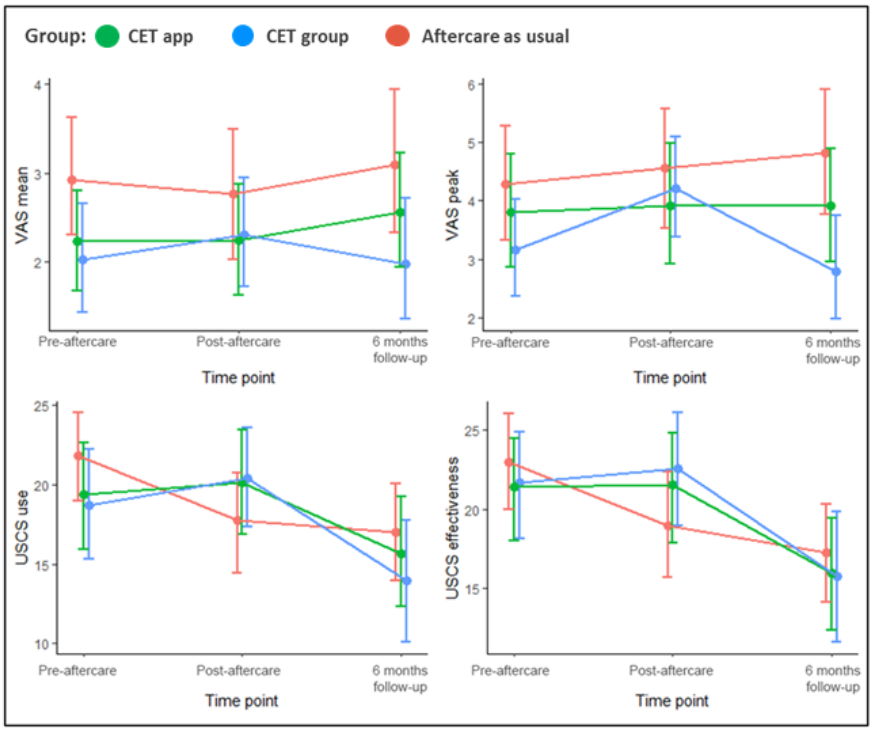


Figure 3. Secondary outcomes by group allocation over time among individuals receiving cue exposure therapy (CET) as group aftercare or as a mobile phone app, or aftercare as usual. USCS: urge-specific coping skills (a lower score reflects less use or lower perceived effectiveness); VAS: visual analog scale (a lower score reflects fewer alcohol cravings).



The completer analyses only included individuals who completed CET with USCS (ie, completed all four sessions at least once) or completed aftercare as usual. This revealed the same pattern as the intention-to-treat analysis where a main effect for time was detected in the predicted mean of all primary outcomes, but no interactions were found between group and time points. This indicated that the entire sample increased their alcohol consumption after ending primary treatment. Similarly, no significant interactions were detected on cravings measures, and both CET app (PMD 8.35, SE 3.14, z=2.66, P=.008) and CET group (PMD 8.53, SE 3.20, z=2.66, P=.008) used the USCS more than aftercare as usual. The completer analysis also

showed that the experimental groups found the USCS more effective than did the active controls (CET app: PMD 8.46, SE 3.01, z=2.74, P=0.006; CET group: PMD 7.73, SE 3.51, z=2.26, P=.02).

Discussion

Principal Findings

The objectives of this study were to investigate whether CET based on a treatment manual and combined with USCS delivered as alcohol treatment aftercare would increase the effectiveness of cognitive behavioral therapy, and whether delivery by means of a mobile phone app would demonstrate similar effectiveness as group sessions. Contrary to our a priori hypothesis, we did not find support for superior effectiveness of CET compared with aftercare as usual on alcohol consumption and urge-related outcomes over time. The experimental CET groups used USCS more than the controls at posttreatment, but this effect was reduced at 6-month follow-up. We found no differences in outcomes between the experimental groups receiving CET through group sessions or a mobile phone app.

Our findings are similar to the previous meta-analysis indicating that individuals with AUD exposed to CET showed no effect to small overall additional effects on drinking-related outcomes [6]. However, the meta-analysis also suggested that CET combined with USCS may achieve more favorable outcomes than conventional CET [28,29]. Monti et al [28] investigated the effectiveness of CET with USCS compared with relaxation and meditation therapy for increasing abstinence among 40 inpatients with AUD on otherwise standard treatment. The 22 individuals who received CET achieved superior results on several alcohol consumption outcomes, including abstinence, fewer drinking days, and drinks per drinking day at 6-month follow-up [28]. Rohsenow et al [29] conducted a larger RCT with 100 individuals with AUD, of whom 59 received CET with USCS as an add-on to standard inpatient treatment. Again, the active control condition was relaxation and meditation therapy, and the treatment goal was abstinence. Patients receiving CET with USCS reported fewer days with excessive drinking at 6- and 12-month follow-ups [29]. In further support of these findings, Monti et al [30] reported similar positive drinking outcomes among inpatients at 6- and 12-month follow-ups when the intervention was delivered as aftercare combined with other psychological and urge-reducing pharmacotherapy treatment. Similar to our findings, these previous studies did not find differences between the experimental and nonexperimental groups on completion of treatment and at 3-month follow-up. However, their findings at 6 and 12 months are in sharp contrast to our results.

It has been demonstrated that CET achieves more favorable outcomes when compared with active control conditions (eg, relaxation and meditation therapy) other than cognitive behavioral therapy [6]. Of four controlled trials (using either randomized or sequential allocation) comparing conventional CET to cognitive behavioral therapy, three reported an equal effect of CET on alcohol consumption outcomes [49-51], and the fourth study reported a superior effect for CET [52]. Furthermore, Kavanagh and coworkers [53] performed an RCT on individuals with AUD allocated to receive either cognitive behavioral therapy alone or cognitive behavioral therapy with CET as an add-on intervention. Results indicated that the addition of conventional CET to cognitive behavioral therapy did not improve alcohol outcomes at posttreatment or at 6- and 12-month follow-ups [53]. Although conventional CET has been compared to cognitive behavioral therapy in prior studies [49-52], and CET has been implemented as an add-on intervention to cognitive behavioral therapy [53], no studies have compared CET combined with USCS to cognitive behavioral therapy. Individuals having CET with USCS can practice coping skills while they are exposed to alcohol cues in vivo, which contrasts with most cognitive behavioral therapy approaches. However, the results from this study suggest that when CET plus USCS is added to cognitive behavioral therapy, it does not enhance the effectiveness of cognitive behavioral therapy in preventing relapses in outpatients on completion of treatment. The combination of CET with coping skills thus appears to be a less important feature for the effectiveness at 6-month follow-up.

Cue exposure therapy with or without USCS is assumed to work by reducing cue-induced cravings (eg, [1-5]). Although prior studies combining CET with USCS have found positive alcohol consumption outcomes at 6- and 12-month follow-ups, it has proven difficult to demonstrate a decrease in the degree of cravings [28,29]. Further, very few of the previous studies investigating a conventional CET approach applied psychometric outcome measures of cravings, or they only applied them at pretreatment to assess whether they predicted alcohol consumption outcomes at later follow-ups [50,52,54,55]. Our sample can generally be defined as non- or low-urge reactors, which is evident from the low craving level reported, and the experimental groups did not achieve a higher degree of cue-reactivity reduction than the control group. This is consistent with prior findings and with the notion that a clinically relevant degree of cravings is required at pretreatment; otherwise, cravings cannot be expected to decrease and may obscure the probability of detecting a change [56]. However, this does not fully explain why this RCT and prior studies did not find more significant results. The fact that it is difficult to document a decrease in cravings relative to alcohol consumption challenges the theoretical assumption that cue-induced cravings cause addictive behavior or, perhaps more likely, the way we currently measure cravings in clinical studies. Cue-induced cravings most often represent automatic and implicit cognitive processes [57,58], and current self-report craving measures may be insufficient to capture such unconscious processes unless cravings reach a certain threshold and are experienced consciously.

It was found that CET delivered through group sessions resulted in greater use of USCS compared with the control group, as reported previously [28,29]. However, this was only the case at posttreatment, and the effect attenuated at 6-month follow-up, suggesting that the advantages may be short-lived. Further, as our overall sample reported low levels of alcohol cravings, they may not have used the USCS or perceived it to be ineffective in reducing the cravings.

The results indicate that the mobile phone app featuring CET with USCS may be equivalent in effectiveness to CET delivered in group sessions for improving alcohol consumption, craving, and USCS outcomes. However, if CET through any delivery pathway does not increase the effectiveness of the already well-documented approach of cognitive behavioral therapy, there is no need to implement CET as an add-on or aftercare. Indeed, this is the first study to show that CET with USCS may not be an effective psychological approach. Prior studies did not implement the experimental approaches as aftercare or in extension to primary treatment with cognitive behavioral therapy, but instead as add-on to primary treatment with other interventions (eg, community meetings, alcohol and health education, vocational counseling, 12-step meetings). It is plausible that we found no effect of CET due to the short follow-up period. The effect of aftercare may require longer than 6 months, and more so when primary treatment results in such a well-treated sample as that in this study, with almost no alcohol intake or cravings. Prior studies have applied CET with USCS among inpatients with a higher degree of addiction severity and using abstinence as the only treatment goal, suggesting that the approach may be more effective in specific patient populations. More research is thus warranted to determine whether CET with USCS is an effective approach. If it does prove effective under certain conditions, we have shown that the method can easily be implemented into a mobile phone app with apparently the same effect as group sessions. In addition to reducing treatment costs, this may have other advantages such as greater access to psychological treatment for treatment-seeking and non-treatment-seeking AUD populations.

Strengths and Limitations

This study is the largest RCT conducted in the area and was based on the CONSORT statement for RCT studies, which was not customary in previous studies. Further strengths are that we eliminated selection bias at entry to aftercare treatment by conducting the study as an RCT, and we based the experimental interventions on a published treatment manual, which optimized the replicability of the study and future implementation of the app as an evidence-based treatment. A high postaftercare follow-up rate of 93% was achieved, which heightens the power and generalizability of the study and reduces the risk of bias. Finally, the findings from this RCT can guide the development of evidence-based eHealth interventions, which is important given the current widespread availability of questionable AUD treatments in app stores.

A number of limitations should also be mentioned. Although the RCT is the largest conducted to date, it may still lack power to detect potential effects and hence commit type II errors, particularly in view of the well-treated sample. Indeed, the power calculation led us to aim for 300 participants, which could not be achieved due to surprisingly few patients entering primary treatment during the recruitment period and many patients declining to participate in the aftercare study. Second, it was challenging to get the individuals with AUD to use the app, and technical issues meant that the software failed to send regular text messages to remind patients to use the app. These limitations could be easily overcome in the future. Finally, no objective measures of alcohol consumption and cravings were applied to validate self-reported outcomes.

Conclusion

Cue exposure therapy combined with urge-specific coping skills delivered as aftercare either in a group session or by a mobile phone app did not increase the effectiveness of cognitive behavioral therapy in this study. This is the first study to show that CET with USCS may not be an effective psychological approach for aftercare of individuals treated for AUD.

Future Directions

There is a need for more large, high-quality RCT studies to assess the effects of CET combined with USCS for individuals with AUD. Further research is especially warranted to investigate (1) appropriate measures of alcohol consumption and craving, including objective measures; (2) the effectiveness when targeting different urge-reactor profiles; (3) the long-term effectiveness of CET, especially when delivered through a mobile device; and (4) whether the app (and also CET in general) is better suited to treat subsamples of individuals with AUD (eg, severe AUD, urge reactors, younger individuals, or those with higher education, living in rural areas, or with busy life schedules).

Acknowledgments

The authors would like to acknowledge the staff at the Alcohol Treatment Clinic, Odense (in particular therapists Gitte Skovgaard, Vibeke Reuschenbach Torp, and Gitte Trier; supervisor and psychiatrist Bodil Andersen; and head of clinic Elisabeth Jessen) and the staff at the Unit for Clinical Alcohol Research, Unit for Psychiatric Research, Department of Clinical Research, University of Southern Denmark (research assistant Birgit Jensen and student assistant Sibel Yilmaz) for their hard work that enabled the implementation and execution of the study. We would also like to acknowledge the staff at the Technology Entrepreneurship and Innovation section, Mads Clausen Institute, University of Southern Denmark, Soenderborg (professor Arne Bilberg, graphic designer Ye Wu, and cand scient tech Minghao Wang) for their assistance in designing and developing the mobile phone app. The study was funded by unconditional grants from the Lundbeck Foundation, Tryg Foundation, and the Region of Southern Denmark.

Authors' Contributions

Authors AIM, BN, ASN, and ES designed the study. AIM and FY designed and developed the mobile phone app under supervision of BN, ASN, and ES. Author DN played a crucial role in the implementation of the study in practice and data acquisition. AIM

and AM conducted the statistical analyses. AIM wrote the first draft of the manuscript, and all authors contributed to 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), 414KB - mhealth_v7i8e13793_app1.pdf]

References

1. Hodgson RJ, Rankin HJ. Modification of excessive drinking by cue exposure. Behav Res Ther 1976;14(4):305-307. [Medline: 962785]
2. Conklin CA, Tiffany ST. Applying extinction research and theory to cue-exposure addiction treatments. Addiction 2002 Feb;97(2):155-167. [Medline: 11860387]
3. Marlatt GA. Cue exposure and relapse prevention in the treatment of addictive behaviors. Addict Behav 1990;15(4):395-399. [Medline: 2248112]
4. Niaura R. Cognitive social learning and related perspectives on drug craving. Addiction 2000 Aug;95(Suppl 2):S155-S163. [Medline: 11002910]
5. Drummond D. What does cue-reactivity have to offer clinical research? Addiction 2000 Aug;95(Suppl 2):S129-S144. [Medline: 11002908]
6. Mellentin AI, Skøt L, Nielsen B, Schippers GM, Nielsen AS, Stenager E, et al. Cue exposure therapy for the treatment of alcohol use disorders: a meta-analytic review. Clin Psychol Rev 2017 Nov;57:195-207 [FREE Full text] [doi: 10.1016/j.cpr.2017.07.006] [Medline: 28781153]
7. Monti P, Kadden RM, Rohsenow DJ, Cooney NL, Abrams DB. Treating Alcohol Dependence: A Coping Skills Training Guide. New York: Guilford Press; 2002.
8. Monti P, Rohsenow DJ. Coping skills training and cue exposure treatment. In: Hester RK, Miller WR, editors. Handbook of Alcoholism Treatment Approaches. Boston, MA: Allyn and Bacon; 2003:213-236.
9. Berglund M, Thelander S, Salaspuro M, Franck J, Andréasson S, Ojehagen A. Treatment of alcohol abuse: an evidence-based review. Alcohol Clin Exp Res 2003 Oct;27(10):1645-1656. [doi: 10.1097/01.ALC.0000090144.99832.19] [Medline: 14574236]
10. National Institute for Health and Clinical Excellence. Alcohol-Use Disorders: Diagnosis, Assessment and Management of Harmful Drinking and Alcohol Dependence. London: RCPsych Publications; Feb 23, 2011.
11. Magill M, Ray LA. Cognitive-behavioral treatment with adult alcohol and illicit drug users: a meta-analysis of randomized controlled trials. J Stud Alcohol Drugs 2009;70(4):516-527 [FREE Full text] [doi: 10.15288/jsad.2009.70.516] [Medline: 19515291]
12. Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction 2002;97(3):265-277. [Medline: 11964100]
13. Barkowski S, Schwartze D, Strauss B, Burlingame GM, Barth J, Rosendahl J. Efficacy of group psychotherapy for social anxiety disorder: a meta-analysis of randomized-controlled trials. J Anxiety Disord 2016;39:44-64. [doi: 10.1016/j.janxdis.2016.02.005] [Medline: 26953823]
14. Cuijpers P, Gentili C, Banos RM, Garcia-Campayo J, Botella C, Cristea IA. Relative effects of cognitive and behavioral therapies on generalized anxiety disorder, social anxiety disorder and panic disorder: A meta-analysis. J Anxiety Disord 2016 Oct;43:79-89. [doi: 10.1016/j.janxdis.2016.09.003] [Medline: 27637075]
15. Tolin DF. Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. Clin Psychol Rev 2010 Aug;30(6):710-720. [doi: 10.1016/j.cpr.2010.05.003] [Medline: 20547435]
16. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol-Res Pr 2011 Dec;42(6):505-512. [doi: 10.1037/a0024485]
17. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
18. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
19. Aboujaoude E, Salame W, Naim L. Telemental health: a status update. World Psychiatry 2015 Jun;14(2):223-230 [FREE Full text] [doi: 10.1002/wps.20218] [Medline: 26043340]
20. Rehm J, Anderson P, Manthey J, Shield KD, Struzzo P, Wojnar M, et al. Alcohol use disorders in primary health care: what do we know and where do we go? Alcohol Alcohol 2016 Jul;51(4):422-427. [doi: 10.1093/alcalc/agv127] [Medline: 26574600]

21. Burlingame GM, Strauss B, Joyce A. Change mechanisms and effectiveness of small group treatments. In: Bergin and Garfield's Handbook of Psychotherapy and Behavior. Hoboken, NJ: John Wiley & Sons; 2013:640-689.
22. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
23. Dedert EA, McDuffie JR, Stein R, McNiel JM, Kosinski AS, Freiermuth CE, et al. Electronic interventions for alcohol misuse and alcohol use disorders: a systematic review. Ann Intern Med 2015 Aug 04;163(3):205-214 [FREE Full text] [doi: 10.7326/M15-0285] [Medline: 26237752]
24. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016 Sep;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
25. 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]
26. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
27. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
28. Monti PM, Rohsenow DJ, Rubonis AV, Niaura RS, Sirota AD, Colby SM, et al. Cue exposure with coping skills treatment for male alcoholics: a preliminary investigation. J Consult Clin Psychol 1993 Dec;61(6):1011-1019. [Medline: 7906700]
29. Rohsenow DJ, Monti PM, Rubonis AV, Gulliver SB, Colby SM, Binkoff JA, et al. Cue exposure with coping skills training and communication skills training for alcohol dependence: 6- and 12-month outcomes. Addiction 2001 Aug;96(8):1161-1174. [doi: 10.1080/09652140120060752] [Medline: 11487422]
30. Monti PM, Rohsenow DJ, Swift RM, Gulliver SB, Colby SM, Mueller TI, et al. Naltrexone and cue exposure with coping and communication skills training for alcoholics: treatment process and 1-year outcomes. Alcohol Clin Exp Res 2001 Nov;25(11):1634-1647. [Medline: 11707638]
31. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
32. 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]
33. Mellentin AI, Nielsen B, Nielsen AS, Yu F, Stenager E. A randomized controlled study of exposure therapy as aftercare for alcohol use disorder: study protocol. BMC Psychiatry 2016 Apr 21;16:112 [FREE Full text] [doi: 10.1186/s12888-016-0795-8] [Medline: 27098817]
34. Mellentin AI, Stenager E, Nielsen B, Nielsen AS, Yu F. A smarter pathway for delivering cue exposure therapy? The design and development of a smartphone app targeting alcohol use disorder. JMIR Mhealth Uhealth 2017 Jan 30;5(1):e5 [FREE Full text] [doi: 10.2196/mhealth.6500] [Medline: 28137701]
35. Schwarz A, Nielsen B, Nielsen AS. Changes in profile of patients seeking alcohol treatment and treatment outcomes following policy changes. Z Gesundh Wiss 2018;26(1):59-67 [FREE Full text] [doi: 10.1007/s10389-017-0841-0] [Medline: 29416960]
36. Nielsen AS, Nielsen B. Implementation of a clinical pathway may improve alcohol treatment outcome. Addict Sci Clin Pract 2015 Mar 07;10:7 [FREE Full text] [doi: 10.1186/s13722-015-0031-8] [Medline: 25928550]
37. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization; 1992.
38. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva: World Health Organization; 1993.
39. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. The fifth edition of the Addiction Severity Index. J Subst Abuse Treat 1992;9(3):199-213. [Medline: 1334156]
40. Kokkevi A, Hartgers C. EuropASI: European adaptation of a multidimensional assessment instrument for drug and alcohol dependence. Eur Addict Res 1995;1(4):208-210. [doi: 10.1159/000259089)]
41. McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. J Nerv Ment Dis 1980 Jan;168(1):26-33. [doi: 10.1097/00005053-198001000-00006] [Medline: 7351540]
42. Sobell MB, Maisto SA, Sobell LC, Cooper AM, Cooper T, Sanders B. Developing a Prototype for Evaluating Alcohol Treatment Effectiveness. Fort Lauderdale, FL: NOVA Southeastern University, College of Psychology; 1980.
43. Sobell LC. Timeline follow-back. In: Sobell MB, editor. Measuring Alcohol Consumption. Tahowa, NJ: Humana Press; 1992:41-72.

44. Sobell LC, Brown J, Leo GI, Sobell MB. The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. Drug Alcohol Depend 1996 Sep;42(1):49-54. [Medline: 8889403]
45. Health and Lifestyle. Copenhagen: Danish Health Authority; 2014. Our 7 recommendations about alcohol URL: https://www.sst.dk/en/health-andlifestyle/alcohol
46. Drobes DJ, Thomas SE. Assessing craving for alcohol. Alcohol Res Health 1999;23(3):179-186 [FREE Full text]
47. Lee JW, Brown ES, Perantie DC, Bobadilla L. A comparison of single-item visual analog scales with a multiitem Likert-type scale for assessment of cocaine craving in persons with bipolar disorder. Addict Disord Treat 2002;1(4):140-142. [doi: 10.1097/00132576-200211000-00005]
48. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. Res Nurs Health 1990 Aug;13(4):227-236. [Medline: 2197679]
49. Heather N, Brodie J, Wale S, Wilkinson G, Luce A, Webb E, et al. A randomized controlled trial of Moderation-Oriented Cue Exposure. J Stud Alcohol 2000 Jul;61(4):561-570. [Medline: 10928726]
50. Dawe S, Rees VW, Mattick R, Sitharthan T, Heather N. Efficacy of moderation-oriented cue exposure for problem drinkers: a randomized controlled trial. J Consult Clin Psychol 2002 Aug;70(4):1045-1050. [Medline: 12182268]
51. Loeber S, Croissant B, Heinz A, Mann K, Flor H. Cue exposure in the treatment of alcohol dependence: effects on drinking outcome, craving and self-efficacy. Br J Clin Psychol 2006 Nov;45(Pt 4):515-529. [doi: 10.1348/014466505X82586] [Medline: 17076961]
52. Sitharthan T, Sitharthan G, Hough MJ, Kavanagh DJ. Cue exposure in moderation drinking: a comparison with cognitive-behavior therapy. J Consult Clin Psychol 1997;65(5):878-882. [Medline: 9337506]
53. Kavanagh DJ, Sitharthan G, Young RM, Sitharthan T, Saunders JB, Shockley N, et al. Addition of cue exposure to cognitive-behaviour therapy for alcohol misuse: a randomized trial with dysphoric drinkers. Addiction 2006 Aug;101(8):1106-1116. [doi: 10.1111/j.1360-0443.2006.01488.x] [Medline: 16869840]
54. Heather N, Tebbutt J, Greeley J. Alcohol cue exposure directed at a goal of moderate drinking. J Behav Ther Exp Psychiatry 1993;24(2):187-195. [Medline: 7903320]
55. Drummond DC, Glautier S. A controlled trial of cue exposure treatment in alcohol dependence. J Consult Clin Psychol 1994;62(4):809-817. [Medline: 7962885]
56. Monti PM, Rohsenow DJ, Rubonis AV, Niaura RS, Sirota AD, Colby SM, et al. Alcohol cue reactivity: effects of detoxification and extended exposure. J Stud Alcohol 1993;54(2):235-245. [Medline: 8384678]
57. Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: 'liking', 'wanting', and learning. Curr Opin Pharmacol 2009;9(1):65-73 [FREE Full text] [doi: 10.1016/j.coph.2008.12.014] [Medline: 19162544]
58. Everitt BJ, Robbins TW. Drug addiction: updating actions to habits to compulsions ten years on. Annu Rev Psychol 2016;67:23-50. [doi: 10.1146/annurev-psych-122414-033457] [Medline: 26253543]

Abbreviations

AUD: alcohol use disorder
CET: cue exposure therapy
CONSORT: Consolidated Standards of Reporting Trials
IQR: interquartile range
PM: predicted mean
PMC: predicted mean change
PMD: predicted mean difference
RCT: randomized controlled trial
USCS: urge-specific coping skills
VAS: Visual Analog Scale

Edited by G Eysenbach; submitted 22.02.19; peer-reviewed by M Blankers, M Bendtsen, MW Huang; comments to author 20.03.19; revised version received 17.06.19; accepted 07.07.19; published 16.08.19.

Please cite as:
Mellentin AI, Nielsen B, Nielsen AS, Yu F, Mejldal A, Nielsen DG, Stenager E
A Mobile Phone App Featuring Cue Exposure Therapy As Aftercare for Alcohol Use Disorders: An Investigator-Blinded Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(8):e13793
URL: http://mhealth.jmir.org/2019/8/e13793/
doi:10.2196/13793
PMID:31420960

Original Paper

An Experimental Investigation of Human Presence and Mobile Technologies on College Students' Sun Protection Intentions: Between-Subjects Study

Zhaomeng Niu[1], PhD; David C Jeong[2], PhD; Elliot J Coups[1,3], PhD; Jerod L Stapleton[1,3], PhD

[1]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

[2]University of Southern California, Los Angeles, CA, United States

[3]Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States

Corresponding Author:
Zhaomeng Niu, PhD
Rutgers Cancer Institute of New Jersey
195 Little Albany Street
New Brunswick, NJ,
United States
Phone: 1 5093393076
Email: zhaomeng.niu@rutgers.edu

Abstract

Background: Health promotion and education programs are increasingly being adapted and developed for delivery through digital technologies. With this shift toward digital health approaches, it is important to identify design strategies in health education and promotion programs that enhance participant engagement and promote behavior change.

Objective: This study aimed to examine the impact of an experiment testing various mobile health (mHealth) skin cancer prevention messages on sun protection intentions and message perceptions among American college students.

Methods: A sample of 134 college students aged 18 years or older participated in a 2×2×2 between-subjects experimental study, designed to examine the individual and combinatory effects of multiple dimensions (human presence, screen size, and interactivity) of digital technologies. The primary study outcome was intention to use sun protection; secondary outcomes included attitudes toward the information, two dimensions of trust, and information processing.

Results: Generally, intention to use sun protection was positively associated with the presence of human characters in the health educational messages (P<.001), delivering educational health messages on a large screen (ie, iPad; P<.001), and higher interactivity (P<.001). Only human presence produced more favorable attitudes (P=.02). Affective trust was positively associated with human presence (P=.006) and large screen size (P<.001), whereas cognitive trust was positively associated with human presence (P<.001) and small screen size (P=.007). Moreover, large screen size led to more heuristic processing (P=.03), whereas small screen size led to more systematic processing (P=.04).

Conclusions: This experimental study demonstrates that the impact of mHealth skin cancer prevention messages differs based on platform and delivery design features. Effects on behavioral intentions, attitudes, and trust were found for conditions with human presence, highlighting the importance of including this feature in mHealth programs. Results from this experimental study can be used to optimize the design of mHealth educational interventions that promote sun protection.

(JMIR Mhealth Uhealth 2019;7(8):e13720) doi:10.2196/13720

KEYWORDS

skin cancer; mHealth; education; sunscreens

Introduction

Skin cancer is the most common cancer in the United States [1], and ultraviolet (UV) exposure has been found to be significantly positively associated with skin cancer. Indoor tanning and outdoor tanning caused by UV exposure are highly prevalent among college students [2]. Previous research has demonstrated that the knowledge level of skin cancer and sun prevention behaviors is low [3], which may lead to harmful outcomes such as sunburns and sun damage. There is a widely recognized need for health interventions to reduce skin cancer

risk behaviors in young adult populations [4]. More than 85% of college students regularly use a smart device [5], and most of them have positive attitudes and high acceptance of mobile learning [6,7]. Thus, mobile digital technologies represent a potentially impactful approach to educate college students about the harms of tanning and importance of sun protection. However, there is a dearth of research about how to best design mobile health (mHealth) educational interventions using both technological features along with social elements in messages. This study used an experimental design to examine the impact of manipulating design aspects of a website promoting sun protection on recipients' message engagement and intentions.

College students today use digital technologies for both social (networking) and academic (learning) purposes [8], warranting the application of social elements of technologies when designing educational messages. Social cognitive theory posits that individuals could learn knowledge and behaviors by following a role model within a particular social context [9,10]. Following a model may motivate individuals to perform behaviors they are less familiar with or those that were more recently learned [11]. For instance, the social presence of humans in food advertising, as opposed to objects, has been found to have a positive influence on eating behaviors and food choices through body image and norms [12,13]. Although the importance of applying social elements in health intervention is widely recognized [14,15], few studies have been designed to specifically examine the impact of human presence in health interventions. This study aimed to explore the influences of presence of human characters in health educational messages on behavioral intentions and information processing.

The impact of messages on intentions and processing also depends on the way that information is presented on mobile devices [16]. One critical form–based factor of mobile devices that has impacted viewers' media experiences is screen size [17,18]. Different screen sizes of mobile devices may impact information processing and different dimensions of trust [17-20]. Heuristic processing described the type of message process that is based on judgmental cues and requires less cognitive effort to process information and make decisions. Systematic processing represents the use of more cognitive effort, such as knowledge and attention, when processing messages [21]. Understanding these mechanisms on different mobile devices can help understand how information process affects the outcomes of persuasive communications through the mobile platform. Affective trust is an emotion-driven trust based on personal bonds or feelings and is often measuring trust from an emotional perspective with adjectives that describe feelings such as *likable* and *warm*. Cognitive trust is a logic-driven trust and is related to judgments of reliability of information [22]. In a previous study, large-screen smartphones lead to higher amount of heuristic processing and affective trust, whereas small-screen smartphones resulted in more comprehensive, systematic information processing and cognitive trust in advertising [20].

Another unique feature of digital technologies is the interactivity function of computers or mobile devices. Modality interactivity refers to the incorporation of interactive tools that afford users greater activity onto medium interfaces [23] and has been demonstrated as effective for users [24], particularly in the domain of health communication [25-27]. For instance, modality interactivity has been found to positively affect attitudes toward health websites [27,28] and the persuasive effect health messages [26]. Furthermore, interactivity has been positively associated with behavioral intentions such as intention to recommend a fitness center [25], as well as intention of actual physical exercise [29].

This study was designed to examine how manipulations in the design and presentation of sun protection health messages impact college students' sun protection intentions (primary outcome) as well as their attitudes, trust of messages, and information processing (secondary outcomes). Specifically, we examined (1) how 3 experimental message manipulations (ie, human presence in health messages, screen size, and interactivity) influenced the primary outcome and secondary outcomes; and (2) potential interactions in the impact of the experimental messages on both primary and secondary outcomes.

Methods

Recruitment and Participants

Participants from a large US university in the American Pacific Northwest were recruited to sign up through a research participation system called Sona system and received course credit as incentive. Study approval was obtained from the university's institutional review board before data collection.

Procedures

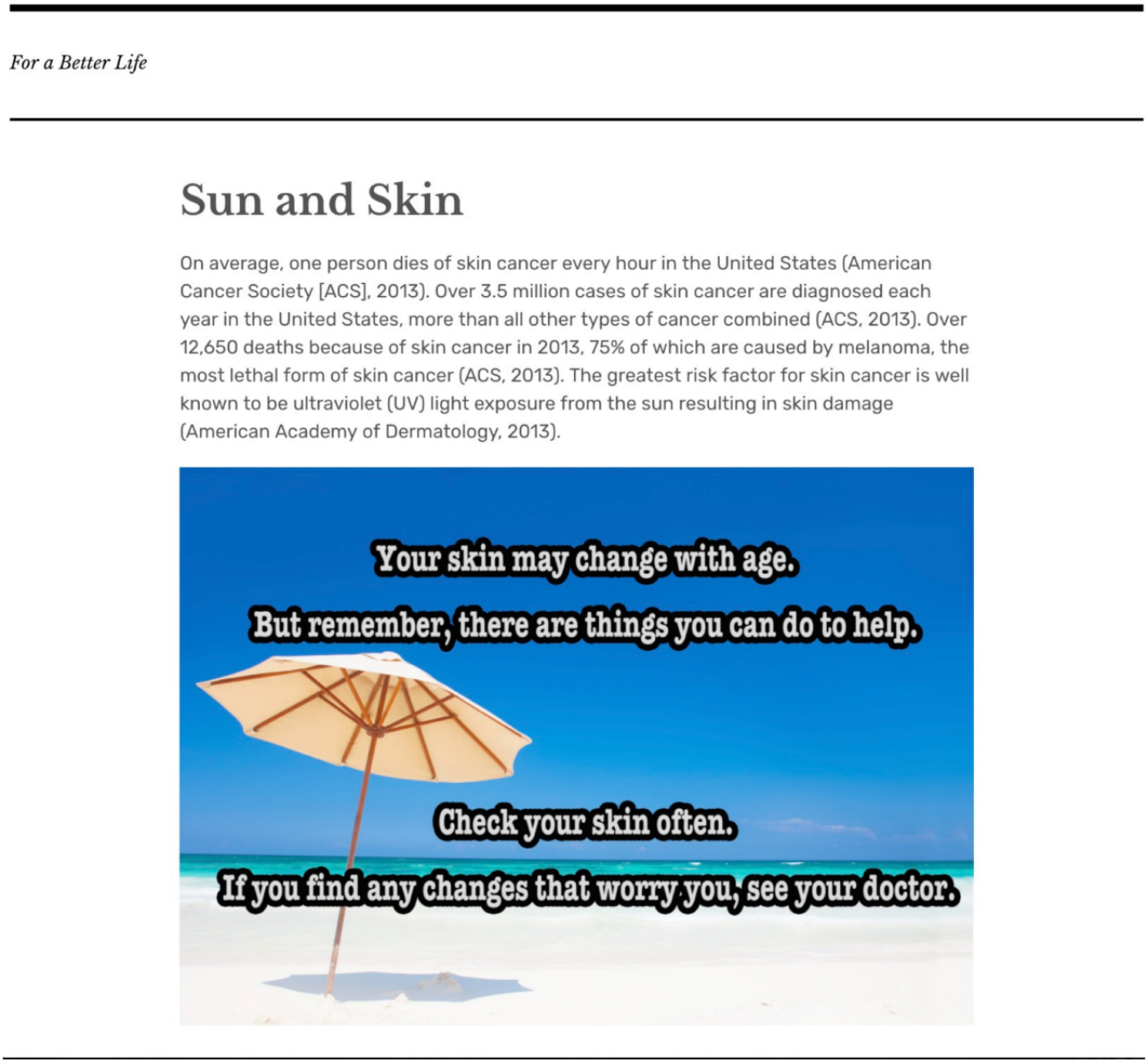
Participants who signed up for the study were asked to come to a communication laboratory. As the independent variables were between-subjects factors, participants were randomly assigned to 1 of the 8 experimental conditions. Participants in the big screen size conditions were provided with an Apple iPad Air (9.7-inch screen), whereas participants in the small screen size conditions were provided with an Apple iPhone 6 or 6s (4.7-inch screen). After being provided with a mobile device, each participant was instructed to explore a health website and read all the information displayed on the website. When participants finished exploring the health websites, they were instructed to complete a questionnaire in Qualtrics using identical laboratory computers.

Experimental Treatment Conditions

A total of 4 websites were created for the experiment to test the 2 (*human presence*: people vs no people)×2 (*screen size*: big screen vs small screen)×2 (*interactivity*: high vs low) between-subjects design. Although all websites had the same title (*Sun and Skin*), webpage layout (Figure 1), and health information about skin cancer, sunburn, and aging, they differed in terms of interactive features and images with or without human presence. The high interactivity condition included interactive function such as *clicking* and *zooming* that allowed participants to interactively access the website content, whereas the low interactivity condition simply loaded content with minimum participant control. The human presence conditions contained images of humans in relation to sun protection objects, whereas the human absence condition strictly contained the sun

protection objects (eg, sunscreen and hat). Screenshots of website pages are shown in Figures 2 and 3 as examples. Screen size was manipulated through mobile devices with different screen sizes.

Figure 1. Screenshot of the website homepage.



Measures

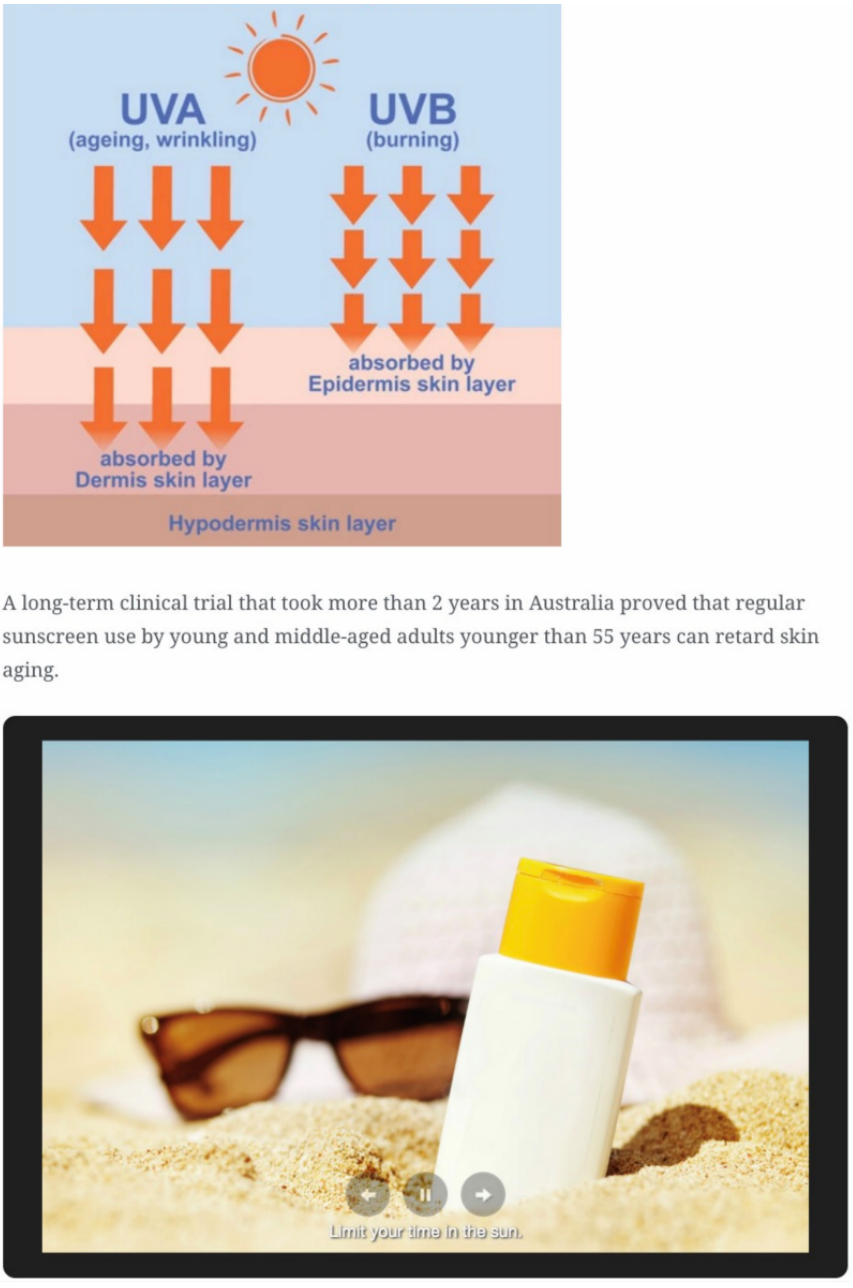
Manipulation Checks

Two self-report questions were used as manipulation checks to examine the effectiveness of the manipulations of the experimental conditions. The manipulation of *perceived interactivity* was assessed using 3 items adapted from the study by Kalyanaraman and Sundar [30]. The manipulation of *human presence* was assessed by a 3-point question (1=*disagree*, 2=*neutral*, and 3=*agree*) asking the extent to which the participants agreed that they saw a human figure on the website.

Primary Outcome

Behavioral intentions were measured with five 5-point Likert-type items reflecting participants' behavioral intention to use sun protection mentioned in the health message [31], such as "In the future, how often do you intend to use sunscreen with Sun-Protection Factor (SPF) 15 or higher on your face when you were in the sun" (alpha=.80, mean 3.32, SD 0.83).

Figure 2. Screenshot of the webpage of high interactivity and human absence condition.



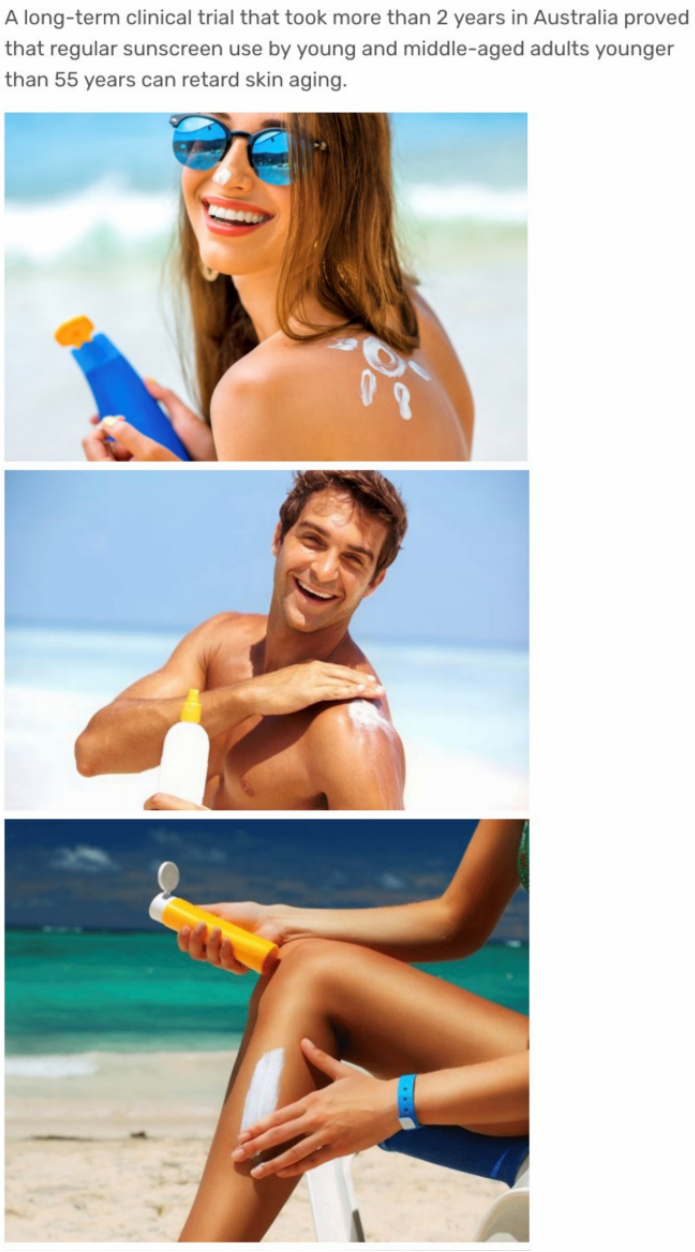
Secondary Outcomes

Attitudes toward skin cancer message were measured using five 7-point statements from a reliable scale in previous studies [26,32] such as asking the respondents to "indicate whether you feel that the messages on the website you just viewed was believable or not " with response options anchored with *not believable* and *believable*. The remaining items included *not informative* or *informative*, *not insightful* or *insightful*, *not interesting* or *interesting*, and *not clear* or *clear*. Items were averaged to create an attitude index (alpha=.75, mean 5.94, SD 0.77).

Cognitive trust was measured using 4 items from previous work on trust and credibility of online information [33,34], asking, "to what extent do you agree or disagree with the following statements: The health information I just read was accurate/accurate/reliable/credible/believable" on a 10-point Likert scale ranging from 1 (*not at all*) to 10 (*extremely*; alpha=.89, mean 7.81, SD 1.65). Similarly, *affective trust* was measured using items from Koh and Sundar's [34] dimensions of trust that are designed to capture emotion-driven trust of information. Moreover, four 10-point Likert-type items ranging from 1 (*not at all*) to 10 (*extremely*) were used to measure whether the respondents felt that the health information they just read was *likable*, *interested in my well-being*, *empathetic*, and *warm* (alpha=.81, mean 6.79, SD 1.70).

Heuristic processing was measured using four 7-point Likert-type items adapted from the study by Griffin et al [35], such as, *When I encounter information about this topic, I focus on only a few key points* (alpha=.71, mean 2.59, SD 0.90). *Systematic processing* was measured using five 7-point Likert-type items adopted from the same questionnaire, such as, "After I encounter information about this topic, I am likely to stop and think about it" (alpha=.78, mean 5.01, SD 1.16).

Figure 3. Screenshot of the webpage of low interactivity and human presence condition.



Covariates

Participants who had paid attention on media to skin cancer information, sun protection information, or both types of information are more likely to be familiar with the related information. Thus, *media attention* was used as a covariate in this study. Participants' *media attention* was adapted from the study by Brossard and Nisbet [36]. Participants were asked to indicate their level of agreement on 4 items with a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), including items such as "I have paid attention to information related to skin cancer and/or sun protection in the past" (alpha=.91).

Statistical Analysis

A power analysis was conducted to ensure that the current sample size was sufficient for testing the hypotheses, with a power of 81% to detect a medium effect size for *F* test. A series of analysis of covariance (ANCOVA) tests were used to test the main effects and interaction effects of human presence, interactivity, and screen size on the primary and secondary outcomes. Media attention to skin cancer and/or sun protection information, race or ethnicity (white vs nonwhite), and sex were controlled as covariates in all ANCOVA analyses.

Results

Sample

The initial sample consisted of 147 undergraduate students, but data cleaning yielded an analytical sample size of 134 participants (cases with missing data were excluded). The mean age of the participants was 19.94 years (SD 2.22). More than 60.0% of the sample was female (83/134, 61.9%). Participants identified as white (76/134, 56.7%), black (15/134, 11.2%), Hispanic (19/134, 14.2%), Asian (15/134, 11.2%), and other (9/134, 6.7%).

Manipulation Checks

Independent sample *t* test was used to check the manipulation of interactivity. According to the results, participants in high interactivity condition (mean 3.25, SD 0.98) scored higher on perceived interactivity than participants in low modality condition (mean 2.87, SD 1.08, t_{132}=2.31; *P*=.04). Similarly, participants in human presence condition (mean 2.71, SD 0.68) scored higher on seeing people in the pictures than participants in human absence condition (mean 1.08, SD 0.41, t_{132}=17.03; *P*<.001). Results of all ANCOVA tests are presented in Table 1.

Table 1. Results of analysis of covariance tests (6 rows represent six 3-way analyses of covariance tests).

	Covariates[a]			Experimental manipulations[a]			Interactions[a]			
Outcome	Sex	Race or ethnicity	Media attention	Human presence	Screen size	Interactivity	Human presence × screen size	Human presence × interactivity	Screen size × interactivity	3-way interaction
Behavioral intentions	.02[b]	.09	.67	<.001[b]	<.001[b]	<.001[b]	.15	.008[b]	<.001[b]	.02[b]
Attitudes	.06	.31	.26	.02[b]	.47	.12	.13	.23	.02[b]	.48
Affective trust	.98	.76	.10	.006[b]	<.001[b]	.12	.90	.52	.61	.95
Cognitive trust	.74	.09	.71	<.001[b]	.007[b]	.64	.35	.16	.48	.25
Systematic processing	.06	.81	.51	.34	.04[b]	.70	.34	.59	.85	.13
Heuristic processing	.16	.55	.84	.07[b]	.03[b]	.23	.73	.49	.38	.20

[a]Values are *P* values of the tests.

[b]Values less than .05 indicate statistical significance.

Behavioral Intentions

Main Effects

Human presence had a main effect on behavioral intentions ($F_{1,123}$=14.90, *P*<.001, η_p=0.11); participants in the human presence condition (mean 3.50, SD 0.62) had greater intention to use sun protection than did participants in the human absence condition (mean 3.12, SD 0.98). The main effect of screen size was also significant on behavioral intentions ($F_{1,123}$=25.34, *P*<.001, η_p=0.17). Participants in the large screen size condition (mean 3.56, SD 0.68) had greater intention to use sun protection than did participants in the small screen size condition (mean 3.02, SD 0.91). Interactivity also demonstrated a main effect on behavioral intentions ($F_{1,123}$=14.37, *P*<.001, η_p=0.11). Participants in the high interactivity condition (mean 3.46, SD 0.68) had greater intention to use sun protection than did participants in the low interactivity condition (mean 3.16, SD 0.96). Female participants were more likely to use sun protection than male participants (*P*=.02).

Interaction Effects

There was a significant 2-way interaction effect of screen size and interactivity on behavioral intentions ($F_{1,123}$=23.75, *P*<.001, η_p=0.16). Large screen size increased the effects of interactivity (high: mean 3.48, SE 0.11; low: mean=3.61 , SE 0.11) on behavioral intentions (Figure 4).

Furthermore, there was also a significant 2-way interaction effect of human presence and interactivity on behavioral intentions ($F_{1,123}$=7.17, *P*=.008, η_p=0.06), indicating human presence helped reinforcing the effects of interactivity (high: mean=3.54 , SE 0.11; low: mean 3.41 , SE 0.12) on behavioral intentions. This indicates that although human absence and low interactivity individually led to diminished intentions to act on the message, coupling these with high interactivity and human presence mitigated these effects and positively influenced and intentions to act (Figure 5).

Finally, a significant 3-way interaction effect of screen size, human presence, and interactivity on behavioral intentions was found ($F_{1,123}$=5.43, *P*=.02, η_p=0.04; Figure 6). Human presence increased behavioral intentions in the small screen size conditions for both high and low interactivity such that highest behavioral intention scores were reported among those who viewed websites with human presence on small screen size devices with high interactivity (mean 3.47, SE 0.16) or low interactivity (mean 3.44, SE 0.17). In the large screen size condition, behavioral intention scores were the highest in the human presence condition with high (mean 3.61, SE 0.15) or low interactivity (mean 3.76, SE 0.16). In conclusion, human presence had a consistently positive impact on intentions across different levels of interactivity and screen size. Specifically, there was a marked difference in intention to use sun protection, depending on the low and high interactivity in the absence of a human presence on a small screen.

Figure 4. Two-way interaction of screen size and interactivity on intentions.

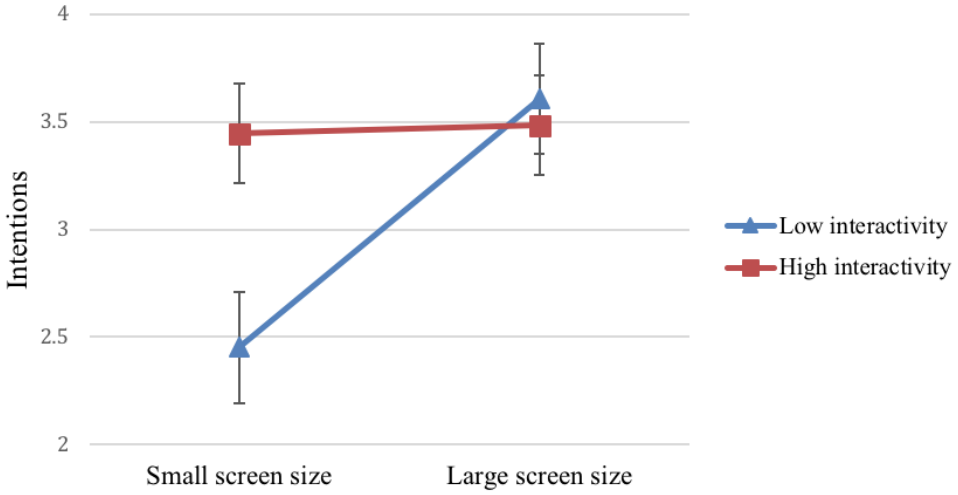


Figure 5. Two-way interaction of human presence and interactivity on intentions.

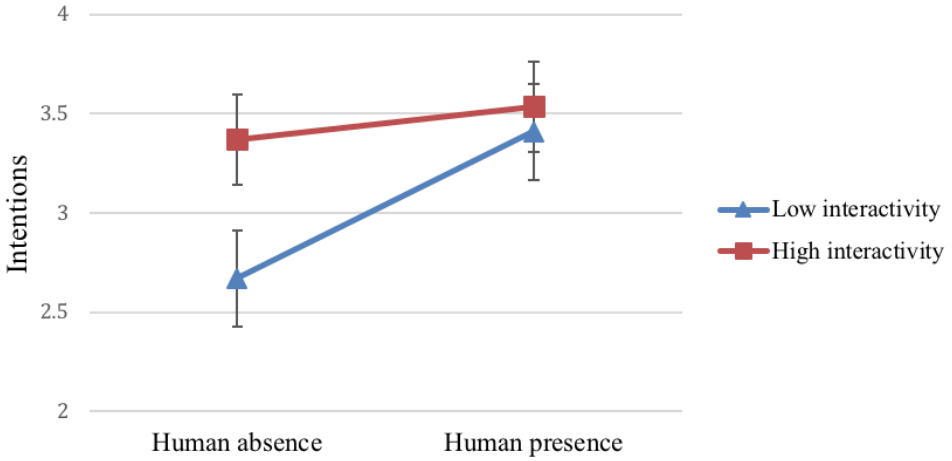
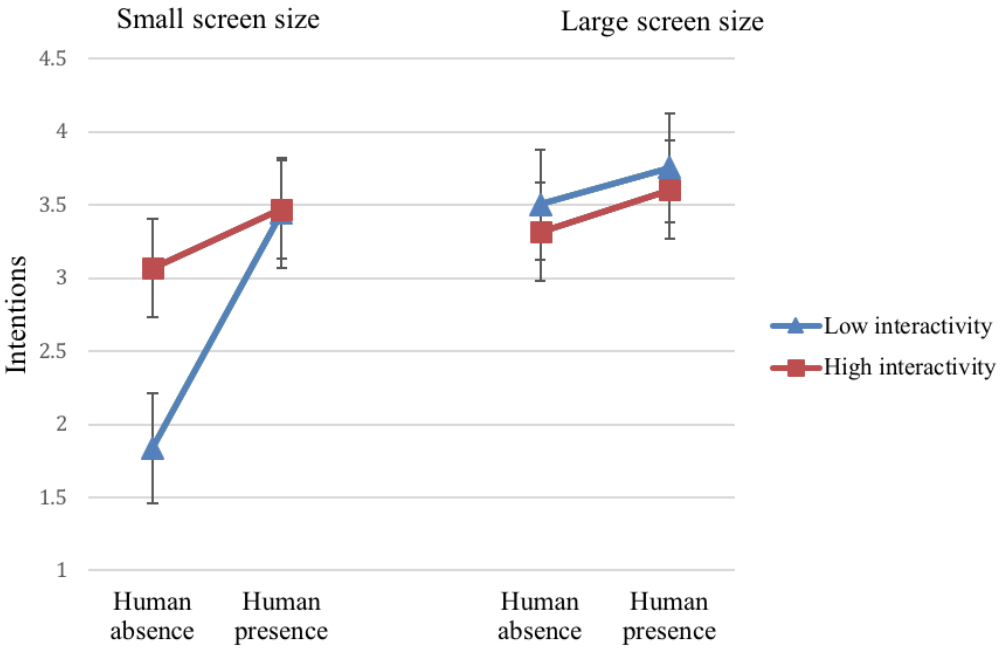


Figure 6. Three-way interaction of screen size, interactivity, and human presence on intentions.



Attitudes

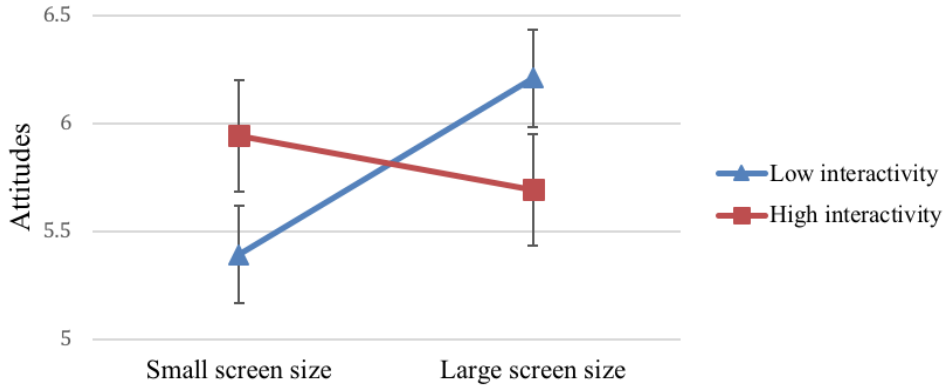
Main Effect

Only human presence had a main effect on attitudes toward health information ($F_{1,123}$=6.10, P=.02, η_p=0.05); participants in the human presence condition (mean 6.10, SD 0.75) had more favorable attitudes toward health information than participants in the human absence condition (mean 5.76, SD 0.76).

Interaction Effects

A significant 2-way interaction effect of screen size and interactivity on attitudes toward health information was observed ($F_{1,123}$=5.72, P=.02, η_p=0.04). Large screen size had stronger effects on attitudes in the low interactivity condition (mean 6.23, SE 0.12), whereas small screen size had stronger effects on attitudes in high interactivity condition (mean 5.93, SE 0.13; Figure 7). These results indicated that although small screens and low interactivity individually led to diminished attitudes toward the health message and intentions to act on the message, coupling these with high interactivity and large screen sizes mitigated these effects and positively influenced attitudes and intentions.

Figure 7. Two-way interaction of screen size and interactivity on attitudes.



Affective and Cognitive Trust

There was a significant effect of human presence on affective trust ($F_{1,123}$=7.93, P=.006, η_p=0.06) and cognitive trust ($F_{1,123}$=13.22, P<.001, η_p=0.10). In addition, screen size was observed to have a main effect on multidimensional trust. Large screen size led to more affective trust ($F_{1,123}$=13.57, P<.001, η_p=0.10); participants in the large screen size condition (mean 7.22, SE 1.74) reported higher affective trust than small screen size condition (mean 6.23, SE 1.50). Small screen size led to more cognitive trust ($F_{1,123}$=7.64, P=.007, η_p=0.06); participants in the small screen size condition (mean 8.28, SE 1.53) reported higher affective trust than large screen size condition (mean 7.52, SE 1.60). No interaction effects were found for affective or cognitive trust.

Information Processing

Screen size had a significant effect on heuristic-systematic processing. Large screen size led to more heuristic processing ($F_{1,123}$=4.74, P=.03, η_p=0.04), and small screen size led to more systematic processing ($F_{1,123}$=4.22, P=.04, η_p=0.03). Participants who viewed information on large screen devices reported higher heuristic processing (mean 2.74, SE 0.99) than those in small screen size condition (mean 2.39, SE 0.76). Participants in the small screen size (mean 5.28, SE 0.88) condition had higher systematic processing than participants in the large screen size condition (mean 4.82, SE 1.29). This indicates that screen size did differ on influencing heuristic-systematic information processing. No interaction effects were found for heuristic or systematic processing.

Discussion

Principal Findings

This study was an experimental study that aimed to examine the impact of a mobile-based educational health program on promoting sun protection behavioral intentions among college students. The individual and combinatory effects of technological factors, such as screen size and interactivity, and social factors, such as human presence human on the primary outcome including intention to use sun protection, and secondary outcomes including attitudes toward the message, trust of the message, and information processing were examined. The preliminary results regarding increased intention to use sun protection supported the promising influences of the educational program, and the results also demonstrated implications for the design of a future sun protection intervention delivered through mobile technologies among college students.

Generally, it was found that the presence of human characters in the educational message was very effective in garnering favorable attitudes, trust, and intentions to act on the sun protection advice in the message. More specifically, the presence of human characters influenced one's affective trust, one's attitude toward the educational message, and one's intention to use sun protection as promoted in the message. This finding regarding the effects of human presence on attitudes, trust, and intentions is consistent with previous research about the role of presence in advertising, online shopping, and social commerce [37-39], which suggested the persuasive role of human presence in health behavior science. It was also found that delivering educational health messages on a large iPad screen led to greater intentions, greater affective trust, and heuristic processing,

whereas messages on a smaller iPhone screen led to cognitive trust and systematic processing. These results are consistent with previous literature regarding the effects of presentation mode on trust [20]. Although higher interactivity did not lead to more favorable attitudes toward the message, it did lead to greater intentions to act on the educational message, confirming the impact of digital interactivity on health-related behavioral outcomes [26].

According to the results of the 2-way interaction effects, attitudes toward the message and intention to use sun protection were positively affected by low interactivity and large screen size. High interactivity and human presence together exerted the highest scores for behavioral intention to use sun protection. The 3-way interaction suggested that behavioral intention was consistently and positively impacted by human presence across the different levels of interactivity and screen size. Interactivity level did not make a huge difference when 3 factors were present together. However, there was a marked difference in intention to use sun protection between the interaction of low interactivity and absence of a human character on a small screen and the interactions of other combinations.

In sum, this study suggests that mHealth programs may increase health outcomes by not only manipulating the screen size and implementing interactivity modalities but also by placing an emphasis on social factors that represent human presence. This human presence serves to both engage the user in a communicative perspective and act as a physical model for actionable behavioral outcomes, such as applying sunscreen. Furthermore, the results demonstrate the significance of screen size in the effectiveness of mobile-based educational program. It is also worth noting that the larger screen size would increase the size of the human characters represented in the educational materials, which may augment the impact of human presence. As the results suggest, a small screen requires the educational materials to be combined with other factors to be more effective. For instance, although small screens are effective for learning when implementing both high interactivity and human presence, these effects are mitigated with low interactivity and in the absence of human characters.

These findings of this study suggest that effective digital education promoting health-related outcomes should include human characters, a large screen size, and some level of modality interactivity. Although mobile device use is ubiquitous among the study sample of college students, it is unclear how different mobile device types may influence college students' cognitive and behavioral intentions and attitudes toward health-related educational content [8,40-42]. A clear implication of this study is that health practitioners/educators designing health educational messages using mobile technologies must account for the effectiveness and form of educational messages across different devices and the suitability of these different messages for students with different access to different mobile technologies.

Limitations and Future Directions

This study also had a number of limitations. First, although the study was adequately powered to test for message effects on the outcomes of interest, the small sample size leads to concerns about the generalizability of results from this single study. In addition, the observation of statistically significant differences in outcomes of interest in an experimental study does not imply that changing intentions or message processing outcome would necessarily produce clinically significant changes in sun protection behavior. A future formal health intervention with a large sample is needed to examine the efficacy of mHealth programs designed based on the findings of this study and their ability to produce meaningful behavioral changes among college students. Moreover, the human presence condition was rather simplistic, essentially referring to the physical presence or absence of a human character. Future work should explore this variable with greater degree of gradation, such as placing individuals in different situational and contextual environments, comparing the effects across different actions performed by human characters, and comparing the effects of multiple human characters included in a single educational setting. Furthermore, the human characters represented in the stimuli materials were depicted in still images. The effectiveness of human characters may be more refined by exploring the impact of physical movement and voice. In addition, the effects for manipulating the presentation of human characters across sex, race, and ethnicity should also be explored in future work, particularly for different groups of subjects.

Furthermore, this study did not control for participants' familiarity with the devices used in the experiment. For instance, users of iPads may be more or less acceptant of messages delivered on iPads. Alternatively, users may have psychological attachment to certain devices, which may have a stronger influence than the size of the screen. Critically, certain groups of individuals may not have access to devices used in the study, which could impact their responsiveness to the content delivered on the devices. Future work would benefit from exploring these device-specific complexities, as this would prevent a bias toward users who have access to and familiarity with such a high-cost device.

Finally, the health information was delivered on a mobile version website on the mobile devices. Future studies can investigate whether participants prefer receiving health information on mobile apps or a website on mobile devices.

Conclusions

This study aimed to investigate how mobile digital technologies and a social factor of a mobile-based educational program may influence health-related outcomes. The observed results suggest feature-specific recommendations for educational program design. Specifically, results indicate that intention to use sun protection remains relatively constant across levels of interactivity and human presence on a large screen, whereas the negative impact of low interactivity is more pronounced on a smaller screen. In sum, this study suggests social factors should be integrated with digital technologies to maximize the effects when designing and delivering health-related educational messages in mobile-based programs that are aiming at generating actionable intent for health behavior.

Conflicts of Interest

None declared.

References

1. Gershenwald JE, Guy Jr GP. Stemming the rising incidence of melanoma: calling prevention to action. J Natl Cancer Inst 2016 Jan;108(1):pii: djv381 [FREE Full text] [doi: 10.1093/jnci/djv381] [Medline: 26563358]
2. Gambla WC, Fernandez AM, Gassman NR, Tan MC, Daniel CL. College tanning behaviors, attitudes, beliefs, and intentions: a systematic review of the literature. Prev Med 2017 Dec;105:77-87. [doi: 10.1016/j.ypmed.2017.08.029] [Medline: 28867504]
3. Cottrell RR, McClamroch L, Bernard AL. Melanoma knowledge and sun protection attitudes and behaviors among college students by gender and skin type. Am J Health Educ 2013 Feb 25;36(5):274-278. [doi: 10.1080/19325037.2005.10608196]
4. Saraiya M, Glanz K, Briss PA, Nichols P, White C, Das D, et al. Interventions to prevent skin cancer by reducing exposure to ultraviolet radiation: a systematic review. Am J Prev Med 2004 Dec;27(5):422-466. [doi: 10.1016/j.amepre.2004.08.009] [Medline: 15556744]
5. Harris Poll. Pearson. 2015. Pearson Student Mobile Device Survey 2015. National Report: College Students URL: https://www.pearsoned.com/wp-content/uploads/2015-Pearson-Student-Mobile-Device-Survey-College.pdf [WebCite Cache ID 78LjPVvlC]
6. Shyshkanova G, Zaytseva T, Frydman O. Mobile technologies make education a part of everyday life. Inf Learn Sci 2017 Nov 13;118(11/12):570-582. [doi: 10.1108/ILS-03-2017-0019]
7. Yang SH. Exploring college students' attitudes and self-efficacy of mobile learning. Turk Online J Educ Technol 2012;11(4):148-154 [FREE Full text]
8. Echenique EG, Molías LM, Bullen M. Students in higher education: social and academic uses of digital technology. Int J Educ Technol High Educ 2015 Jan 2;12(1):25-37. [doi: 10.7238/rusc.v12i1.2078]
9. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
10. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
11. Bandura A. Social cognitive theory of mass communication. In: Bryant J, Zillmann D, editors. Media Effects: Advances in Theory and Research. New York: Routledge; 2009:121-153.
12. Herman CP, Roth DA, Polivy J. Effects of the presence of others on food intake: a normative interpretation. Psychol Bull 2003 Nov;129(6):873-886. [doi: 10.1037/0033-2909.129.6.873] [Medline: 14599286]
13. McFerran B, Dahl DW, Fitzsimons GJ, Morales AC. I'll have what she's having: effects of social influence and body type on the food choices of others. J Consum Res 2010 Apr 1;36(6):915-929. [doi: 10.1086/644611]
14. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
15. Smedley BD, Syme SL, Committee on Capitalizing on Social Science and Behavioral Research to Improve the Public's Health. Promoting health: intervention strategies from social and behavioral research. Am J Health Promot 2001;15(3):149-166. [doi: 10.4278/0890-1171-15.3.149] [Medline: 11265579]
16. Jan SR, Ullah F, Ali H, Khan F. Enhanced and effective learning through mobile learning an insight into students perception of mobile learning at university level. Int J Sci Res Sci Eng Technol 2016;2(2):674-681 [FREE Full text]
17. Lombard M, Ditton T. At the heart of it all: the concept of presence. J Comput-Mediat Comm 1997;3(2). [doi: 10.1111/j.1083-6101.1997.tb00072.x]
18. Grabe ME, Lombard M, Reich RD, Bracken CC, Ditton TB. The role of screen size in viewer experiences of media content. Vis Commun Q 1999 Mar;6(2):4-9. [doi: 10.1080/15551399909363403]
19. Raptis D, Tselios N, Kjeldskov J, Skov MB. Does Size Matter?: Investigating the Impact of Mobile Phone Screen Size on Users' Perceived Usability, Effectiveness and Efficiency. In: Proceedings of the 15th International Conference on Human-Computer Interaction With Mobile Devices and Services. 2013 Presented at: MobileHCI'13; August 27-30, 2013; Munich, Germany p. 127-136. [doi: 10.1145/2493190.2493204]
20. Kim KJ, Sundar SS. Mobile persuasion: can screen size and presentation mode make a difference to trust? Hum Commun Res 2015 Apr 21;42(1):45-70. [doi: 10.1111/hcre.12064]
21. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
22. Johnson D, Grayson K. Cognitive and affective trust in service relationships. J Bus Res 2005 Apr;58(4):500-507. [doi: 10.1016/S0148-2963(03)00140-1]
23. Sundar SS, Xu Q, Bellur S, Oh J, Jia H. Modality is the Message: Interactivity Effects on Perception and Engagement. In: CHI'10 Extended Abstracts on Human Factors in Computing Systems. 2010 Presented at: CHI EA'10; April 10-15, 2010; Atlanta, Georgia, USA p. 4105-4110. [doi: 10.1145/1753846.1754110]

24. Sundar SS, Xu Q, Bellur S, Oh J, Jia H. Beyond Pointing and Clicking: How Do Newer Interaction Modalities Affect User Engagement? In: CHI'11 Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI EA'11; May 7-12, 2011; Vancouver, BC, Canada p. 1477-1482. [doi: 10.1145/1979742.1979794]
25. Lu Y, Kim Y, Dou XY, Kumar S. Promote physical activity among college students: using media richness and interactivity in web design. Comput Human Behav 2014 Dec;41:40-50. [doi: 10.1016/j.chb.2014.08.012]
26. Oh J, Sundar SS. How does interactivity persuade? An experimental test of interactivity on cognitive absorption, elaboration, and attitudes. J Commun 2015 Feb 27;65(2):213-236. [doi: 10.1111/jcom.12147]
27. Niu Z. Examining the effects of interactivity on skin cancer websites: a feasibility study on web-based skin cancer intervention. Ann Behav Med 2018;52:S9 [FREE Full text]
28. Lustria ML. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. J Assoc Inf Sci Technol 2007 Apr;58(6):766-776. [doi: 10.1002/asi.20557]
29. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychol Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
30. Kalyanaraman S, Sundar SS. The psychological appeal of personalized content in web portals: does customization affect attitudes and behavior? J Commun 2006 Mar;56(1):110-132. [doi: 10.1111/j.1460-2466.2006.00006.x]
31. Cafri G, Thompson JK, Roehrig M, van den Berg P, Jacobsen PB, Stark S. An investigation of appearance motives for tanning: the development and evaluation of the physical appearance reasons for tanning scale (PARTS) and its relation to sunbathing and indoor tanning intentions. Body Image 2006 Sep;3(3):199-209. [doi: 10.1016/j.bodyim.2006.05.002] [Medline: 18089223]
32. Sundar SS. Multimedia effects on processing and perception of online news: a study of picture, audio, and video downloads. Journalism Mass Commun Q 2016 Jun 25;77(3):480-499. [doi: 10.1177/107769900007700302]
33. Hu Y, Sundar SS. Effects of online health sources on credibility and behavioral intentions. Commun Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
34. Koh YJ, Sundar SS. Heuristic versus systematic processing of specialist versus generalist sources in online media. Hum Commun Res 2010 Apr 1;36(2):103-124. [doi: 10.1111/j.1468-2958.2010.01370.x]
35. Griffin RJ, Neuwirth K, Giese J, Dunwoody S. Linking the heuristic-systematic model and depth of processing. Commun Res 2016 Jun 30;29(6):705-732. [doi: 10.1177/009365002237833]
36. Brossard D, Nisbet MC. Deference to scientific authority among a low information public: understanding US opinion on agricultural biotechnology. Int J Public Opin Res 2006 Mar 13;19(1):24-52. [doi: 10.1093/ijpor/edl003]
37. Fortin DR, Dholakia RR. Interactivity and vividness effects on social presence and involvement with a web-based advertisement. J Bus Res 2005 Mar;58(3):387-396. [doi: 10.1016/S0148-2963(03)00106-1]
38. Hassanein K, Head M. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. Int J Hum-Comput St 2007 Aug;65(8):689-708. [doi: 10.1016/j.ijhcs.2006.11.018]
39. Lu B, Fan W, Zhou M. Social presence, trust, and social commerce purchase intention: an empirical research. Comput Human Behav 2016 Mar;56:225-237. [doi: 10.1016/j.chb.2015.11.057]
40. Blackwell CK, Lauricella AR, Wartella E. Factors influencing digital technology use in early childhood education. Comput Educ 2014 Aug;77:82-90. [doi: 10.1016/j.compedu.2014.04.013]
41. Henderson M, Selwyn N, Aston R. What works and why? Student perceptions of 'useful' digital technology in university teaching and learning. Stud High Educ 2015 Feb 25;42(8):1567-1579. [doi: 10.1080/03075079.2015.1007946]
42. Collins A, Halverson R. Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America. New York: Teachers College Press; 2018.

Abbreviations

ANCOVA: analysis of covariance
mHealth: mobile health
UV: ultraviolet

Edited by G Eysenbach; submitted 15.02.19; peer-reviewed by M Wakkee, N Dexter, C Mather; comments to author 11.04.19; revised version received 15.05.19; accepted 16.05.19; published 26.08.19.

Please cite as:
Niu Z, Jeong DC, Coups EJ, Stapleton JL
An Experimental Investigation of Human Presence and Mobile Technologies on College Students' Sun Protection Intentions: Between-Subjects Study
JMIR Mhealth Uhealth 2019;7(8):e13720
URL: http://mhealth.jmir.org/2019/8/e13720/
doi:10.2196/13720
PMID:31452523

Original Paper

Long-Term Outcomes of a Therapist-Supported, Smartphone-Based Intervention for Elevated Symptoms of Depression and Anxiety: Quasiexperimental, Pre-Postintervention Study

Marcos Economides[1], PhD; Kristian Ranta[1], MSc; Albert Nazander[1], MSc; Outi Hilgert[1], MD; Philippe R Goldin[2], PhD; Anu Raevuori[1,3,4], MD, PhD; Valerie Forman-Hoffman[1], MPH, PhD

[1]Meru Health Inc, Palo Alto, CA, United States

[2]Betty Irene Moore School of Nursing, University of California Davis, Sacramento, CA, United States

[3]Clinicum, Department of Public Health, University of Helsinki, Helsinki, Finland

[4]Department of Adolescent Psychiatry, Helsinki University Central Hospital, Helsinki, Finland

Corresponding Author:
Marcos Economides, PhD
Meru Health Inc
470 Ramona Street
Palo Alto, CA, 94301
United States
Phone: 1 (650) 740 9404
Email: marcos@mhealthoutcomes.com

Abstract

Background: Depression is one of the most common mental health disorders and severely impacts one's physical, psychological, and social functioning. To address access barriers to care, we developed Ascend—a smartphone-delivered, therapist-supported, 8-week intervention based on several evidence-based psychological treatments for depression and anxiety. A previous feasibility study with 102 adults with elevated depression reported that Ascend is associated with a postintervention reduction in depression symptoms.

Objective: We aimed to examine whether Ascend is associated with a reduction in symptoms of anxiety, and importantly, whether reductions in symptoms of depression and anxiety are maintained up to 12-months postintervention.

Methods: We assessed whether the previously reported, end-of-treatment improvements seen in the 102 adults with elevated symptoms of depression extended up to 12 months posttreatment for depression symptoms (measured by the Patient Health Questionnaire-9 [PHQ-9]) and up to 6 months posttreatment for anxiety symptoms (added to the intervention later and measured using the Generalized Anxiety Disorder-7 [GAD-7] scale). We used linear mixed effects models with Tukey contrasts to compare time points and reported intention-to-treat statistics with a sensitivity analysis.

Results: The intervention was associated with reductions in symptoms of depression that were maintained 12 months after the program (6.67-point reduction in PHQ-9 score, 95% CI 5.59-7.75; $P<.001$; Hedges g=1.14, 95% CI 0.78-1.49). A total of 60% of the participants with PHQ-9 scores above the cutoff for major depression at baseline (PHQ≥10) reported clinically significant improvement at the 12-month follow-up (at least 50% reduction in PHQ-9 score and postprogram score <10). Participants also reported reductions in symptoms of anxiety that were maintained for at least 6 months after the program (4.26-point reduction in GAD-7 score, 95% CI 3.14-5.38; $P<.001$; Hedges g=0.91, 95% CI 0.54-1.28).

Conclusions: There is limited evidence on whether outcomes associated with smartphone-based interventions for common mental health problems are maintained posttreatment. Participants who enrolled in Ascend experienced clinically significant reductions in symptoms of depression and anxiety that were maintained for up to 1 year and 6 months after the intervention, respectively. Future randomized trials are warranted to test Ascend as a scalable solution to the treatment of depression and anxiety.

(JMIR Mhealth Uhealth 2019;7(8):e14284) doi:10.2196/14284

KEYWORDS

digital health; depression; anxiety; mindfulness; CBT; online intervention; smartphone intervention

Introduction

Depression is a common mental health disorder and one of the leading causes of disease burden and disability worldwide [1-3]. Individuals with depression have a reduced capacity to work and function in daily life, causing major economic and societal costs [4-6]. By 2030, depression is predicted to pose the largest burden of disease in high-income countries, surpassing heart disease, dementia, and alcohol-related disorders [1].

Although there are several effective pharmacological [7] and psychological [8,9] treatments for depression, less than half of all individuals who require treatment actually receive it [10,11]. Barriers to treatment include financial and time constraints, long wait periods, a shortage of trained professionals, and fear of stigmatization [12,13]. When treated with antidepressants, an estimated 30%-50% of patients do not experience significant symptom reduction [14], up to 80% of patients report at least one mild-to-severe side effect [15], and few patients maintain a state of long-term remission [16]. Estimates suggest that up to 70% of individuals with depressive disorders have a comorbid anxiety disorder, which renders treatment even more challenging [17]. Further, up to 75% of patients referred to in-person psychotherapy either do not enter treatment or discontinue treatment prematurely [18,19]. Thus, there is an urgent need for novel, evidence-based treatments for depression and anxiety to overcome these barriers.

Recently, digital interventions delivered via smartphone apps have been developed as a means to address this need [20]. Smartphone ownership has seen rapid worldwide growth [21], and survey data suggest that smartphone-based interventions may be preferred over other online formats by health care consumers [22,23]. Smartphone-based interventions offer several advantages over traditional treatment modalities including large-scale accessibility and scalability; low costs; patient anonymity and privacy; standardized content that is less dependent on therapist skills; flexible usage at a self-determined time and pace, which is thought to enhance self-efficacy [24]; monitoring of activity, symptoms, and progression in real time; provision of personalized feedback, motivational support, and targeted care; and potential to improve adherence to treatment [25]. Preliminary evidence suggests that smartphone-based interventions are a promising means to treat depression and anxiety, with recent meta-analyses reporting small-to-moderate reductions in clinical symptoms across 27 studies [26,27].

Despite these promising results, several important questions regarding the design, efficacy, and implementation of smartphone-based interventions for common mental health disorders remain unanswered. First, few authors report on the outcomes of commercially developed smartphone-based interventions for depression and anxiety, making it difficult to evaluate their utility [25,28]. Second, despite being generally effective for the short-term treatment of symptoms of depression and anxiety, there is a dearth of evidence regarding how long the beneficial effects of smartphone-based interventions and other online interventions are maintained posttreatment [29]. This is important because an estimated 50% or more of patients with major depression or generalized anxiety disorder will relapse within 6 to 12 months after the end of an initially effective treatment [30-32]. Thus, investigation of the long-term outcomes associated with smartphone-based interventions for depression and anxiety is an important step toward understanding their true real-world effectiveness. Lastly, despite an increased interest in transdiagnostic interventions that target depression and anxiety concurrently [33], most smartphone-based interventions for mental health are disorder-specific, and it remains unknown whether smartphone-based interventions can address symptoms of depression and anxiety when they are comorbid.

We recently evaluated the feasibility of the Meru Health *Ascend* intervention, a novel, 8-week smartphone-based intervention for elevated symptoms of depression and anxiety, assisted by a remote therapist [34]. The intervention was found to be feasible and was associated with a postintervention reduction in depression. In this study, we extend these findings by investigating whether the previously reported postintervention reductions in depression are maintained at a 1-year follow-up, whether *Ascend* is associated with reductions in comorbid symptoms of anxiety, and whether any postintervention reductions in symptoms of anxiety are maintained at a 6-month follow-up. Thus, this study aims to lend important insights into the real-world, long-term impact of smartphone-based interventions for the treatment of common mental health problems. We hypothesized that postintervention reductions in symptoms of depression would be maintained at the 1-year follow-up and that Ascend would be associated with reductions in the symptoms of anxiety, which would persist at the 6-month follow-up. Lastly, as a secondary objective, we aimed to examine whether participant demographics or intervention engagement were predictive of symptom change or attrition at follow-up.

Methods

Research Design

We used a quasiexperimental research design that included a single-arm, pre- and postintervention assessment of outcomes. Symptoms of depression were measured before the intervention ("baseline"); at the end of the 8-week intervention; and at 1, 3, 6, and 12 months postintervention. Symptoms of anxiety were measured at baseline; week 4 of the intervention; the end of the 8-week intervention; and 1, 3, and 6 months postintervention.

Participants

This study included adult patients treated at the Meru Health Online Clinic, a national remote health care provider that currently operates in the United States and Finland. The clinic has had a rolling enrolment since March 2017. At the time of this study, 197 enrollees had passed the 6-month postintervention outcome window (recruited between March

2017 and June 2018), and 102 passed the 12-month postintervention outcome window (recruited between March and December 2017). Our primary analysis includes the latter group, although we also report the Patient Health Questionnaire (PHQ-9) results from the former group with 6-month postintervention outcomes only. The *Ascend* intervention was primarily intended to treat symptoms of depression; however, many individuals with depression also have comorbid symptoms of anxiety [17]. Thus, to evaluate whether *Ascend* is also associated with reductions in symptoms of anxiety, Generalized Anxiety Disorder-7 (GAD-7) scale measures were added to the intervention assessment in December 2017. At the time of this study, 102 of the original 197 enrollees with 6-month postintervention PHQ-9 outcomes also had GAD-7 (anxiety) data (recruited between January and June 2018), while 12-month postintervention anxiety data were not yet available for any participants. Note that the 102 participants with 6-month postintervention GAD-7 data are different from the 102 participants with 12-month postintervention PHQ-9 data.

Participants were recruited via online Facebook advertisements that sought participants for a smartphone-based intervention for depression that included self-guided smartphone-delivered content, private access to a therapist via messaging, and an anonymous group-chat feature with other participants. Prior to the intervention, participants were given free access to the app and trained on how to use the group-chat feature and communicate with their assigned therapist. Participant demographics (age, gender, and antidepressant medication status) were acquired at the start of the intervention via an intake questionnaire administered online. Since medication status was introduced in July 2017, these data are absent for the first 33 participants. Outcome measures were administered via the app for all within-intervention time points (including immediately postintervention) and via email for all follow-up time points.

For inclusion, participants had to provide informed consent via the Meru Health app, own a smartphone, have at least mild symptoms of depression (a score≥5 on the PHQ-9 at baseline), and acknowledge/demonstrate the ability to commit to a minimum of 20 minutes of practice per day for 6 days per week across the 8-week intervention (as judged by both the participant and their assigned therapist). Exclusion criteria included a previous suicide attempt, severe active suicidal ideation with a specific plan, severe self-harm, active substance abuse, and a history of psychosis. Inclusion/exclusion criteria were assessed prior to enrolment via phone interviews between study participants and intervention therapists, as per the standard treatment procedure at the Meru Health Online Clinic. Participants were not compensated for their time but could participate in the intervention for free. All participants provided informed consent for their anonymized data to be used for research purposes prior to engaging with the intervention. All procedures were reviewed by the Pearl Institutional Review Board, which granted exemption for analyses of previously collected and deidentified data. All procedures performed were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

An a priori sample size calculation was performed for comparing patient-reported outcome measures (PHQ-9 and GAD-7 scores) at baseline and follow-up time points. Using an alpha level of .05, a power of 0.8, and a medium effect size of 0.5, 33 subjects were needed. Thus, this study was sufficiently powered to detect a medium effect, even after accounting for substantial dropout over the course of the 1-year follow-up period.

Intervention

The Meru Health *Ascend* intervention has been described in detail previously [34]. Briefly, the intervention consists of 8 modules delivered sequentially over an 8-week period, which include content derived from evidence-based practices such as mindfulness-based stress reduction [35], mindfulness-based cognitive therapy [36], cognitive-behavioral therapy [37], and behavioral activation therapy [38]. The content includes text; video; audio-guided mindfulness meditation exercises; infographics that illustrate cognitive-behavioral therapy principles; and journal prompts. Daily content and practices range from 10 to 30 minutes, except for the first day of each week, in which a series of introductory videos extend the content to a maximum of 45 minutes. The Meru Health app can be used on both Android and iOS and is designed to be platform-agnostic and thus equivalent across different operating systems.

A licensed therapist (employed by Meru Health) provides support to participants via messaging (and less frequently, phone calls), throughout the intervention. As part of this support, therapists review practice logs using a provider "dashboard" and electronic medical records (that detail participant engagement and patient-reported outcomes to date) to monitor individual participant progress. Therapists aim to spend approximately 20 minutes (on average) supporting each participant per week of the intervention (including initiating contact at least 2-3 times per week), but are at liberty to adjust the levels of support in accordance with each participant's individual progress. In addition, participants are free (and encouraged) to contact their allocated therapist when they require additional support. Such two-way interaction is designed to create a system of support that is structured while being tailored to each participant's personal preference and needs. Further, therapists are instructed to conduct a phone-based assessment for any participants that show signs of mental deterioration during or immediately after the intervention. In case of an emergency, such as severe suicidality, the intervention includes a written security plan, which all participants are required to review with their therapist before engaging with the intervention.

Participants are enrolled in groups of 10-15 individuals that work through the intervention at the same time and can provide anonymous support to one another via a discussion board within the app. Specifically, participants can post anonymous reflections on practices and lessons to the discussion board, to which their therapist can respond freely, and to which other group members can respond with prewritten empathy statements and emoticons. Free cross-talk between participants is not allowed.

In Finland, Meru Health is approved by the Finnish National Supervisory Authority for Welfare and Health (Valvira approval number V/25535/2017) and is compliant with the European Union General Data Protection Regulation. In the United States,

Meru Health is compliant with HIPAA (Health Insurance Portability and Accountability Act of 1996) legislation. All protected health information is kept in a HIPAA-compliant electronic medical record, which is housed in cloud-based storage systems hosted by a company named Datica [39]. All data are encrypted in transit, end-to-end, and at rest.

Measures

Patient Health Questionnaire

The PHQ-9 is a 9-item depression scale, derived from the full PHQ and is one of the most widely used instruments to screen for the presence and severity of depression in primary care [40]. Participants rate each item on a Likert scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 27. In general, a score of 10 or above suggests the presence of major depression, and scores of 5, 10, 15, and 20 are taken as cut-off points for mild, moderate, moderately severe, and severe depression, respectively. The PHQ-9 has excellent internal consistency (Cronbach α of 0.89 in primary care settings) and excellent test-retest reliability [41]. In their original validation study, Kroenke and colleagues [40] defined a clinically significant improvement in depression as a 50% reduction in the PHQ-9 score combined with a postintervention score of <10 (for participants with baseline scores≥10), and this definition has been further validated in a comparison study [42]. A similar yet more liberal definition for clinically significant change was also proposed by Löwe and colleagues [43] as a PHQ-9 score reduction of ≥5.

Generalized Anxiety Questionnaire

The GAD-7 is a 7-item scale used extensively in outpatient and primary care settings to screen for the presence and severity of an anxiety disorder [44]. Participants rate each item on a Likert scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 21. In general, a score of 10 or above is suggestive of the presence of anxiety to the extent that further evaluation is warranted, and scores of 5, 10, and 15 are taken as cut-off points for mild, moderate, and severe anxiety, respectively. The GAD-7 has excellent reliability and internal consistency (Cronbach α of 0.89) and has been validated in both the general population and primary care settings [44,45]. A clinically significant improvement in anxiety symptoms has previously been defined as a GAD-7 score reduction of ≥3 [46].

Statistical Analysis

Measures of Engagement

Descriptive statistics were calculated for participant demographics and week-by-week engagement metrics. For each participant, we calculated a measure of intervention engagement, defined as total days in which >3 minutes of app-based meditation was completed. We used a threshold of 3 minutes, as this corresponds to the shortest meditation session available in the intervention. We used logistic regression to explore whether baseline characteristics were predictive of the completion of outcome measures at 6 and 12 months postintervention. Explanatory variables included age, gender, country, intervention engagement (as defined above), and baseline PHQ-9 or GAD-7 scores.

Patient-Reported Outcomes

Outcome measures were analyzed using an intention-to-treat analysis in which all participants with outcome measures at baseline were included, regardless of intervention engagement or attrition. We used linear mixed effects models (LMMs) implemented through LME4 [47] and Tukey contrasts to compare between time points (using the "multcomp" package [48]) in the statistical computing software R [computer software] (version 3.5.2. Vienna, Austria: R Foundation for Statistical Computing). LMMs are capable of handling missing data and are considered superior to other ITT approaches such as the last observation carried forward [49]. We modeled a single within-subject factor "time" (as a fixed effect) and a separate baseline for each participant (random-intercept model). Time was modeled as a categorical predictor, and we therefore did not enforce a linear relationship between time and outcome measures. We also ran the model while controlling for participant age, gender, country, total intervention engagement (total active days), and total therapist contact, both with and without two-way interactions with time, which produced equivalent results. We report the contrast estimate, 95% CI of the estimate, and *P* value. *P* values<.05 were considered significant. The estimated marginal mean and standard error for each time point was calculated using the "emmeans" package in R.

Sensitivity Analysis

Since LMMs rely on data being "missing at random," an assumption that is difficult to verify in clinical research, we also implemented an LMM-based pattern-mixture model (PMM) for the analysis of PHQ-9 (depression) scores. PMMs are based on a joint modeling of outcomes and missingness and can account for data that is "missing not at random" [50]. We modified the estimated marginal mean (EMM) from the previously described LMM at the 6- and 12-month time points based on specific clinical assumptions about participants with missing data. We identified three patterns of data: (1) participants with complete outcome data at baseline and 12-month follow-up (n=52); (2) participants who were lost to follow-up during or immediately postintervention (n=21); and (3) participants who were lost to follow-up at the 1-, 3-, 6-, or 12-month follow-up time points (n=29). We used the conservative approach of assuming that participants belonging to pattern 2 did not benefit from the intervention, and we set the EMM at 6 and 12 months of follow-up equal to the EMM at baseline for these participants. For group 3, we estimated the EMM at the 6- and 12-month follow-ups as being equal to the EMMs across the first three time points of the intervention (baseline to week 4). The overall treatment effect at 6 or 12 months of follow-up was then defined as a linear combination of EMMs for the three different patterns, weighted by the proportion of participants belonging to each pattern. The standard error for this estimate was calculated using the delta method (via the "car" package in R [51]), and a test of the null hypothesis of no treatment effect was performed using a Wald statistic. A full description of the LMM-based PMM procedure is provided in a previous paper [52].

Effect Size Calculation

For comparisons with previous literature, we calculated effect size (Hedges *g*) immediately postintervention and at 6 and 12 months after the intervention as the difference in outcome scores from baseline divided by the pooled weighted SD (based on observed outcome scores only). For the PMM, effect size (Cohen *d*) was calculated as the estimated reduction in outcome score (from the LMM-based PMM) divided by the SD of observed outcome scores at baseline. We also calculated the percentage of participants who met the definition for clinically significant improvement in PHQ-9 symptoms postintervention and at the 12-month follow-up, defined as either a 5-point reduction in PHQ-9 score from baseline or a 50% reduction in PHQ-9 score combined with a postintervention PHQ-9 score<10 (for participants with baseline scores ≥10).

Predictors of Outcome Change

Lastly, we performed exploratory multiple regressions to test whether participant demographics and engagement metrics predicted outcome change from baseline to 6 and 12 months postintervention with regard to PHQ-9 scores and from baseline to 6 months postintervention with regard to GAD-7 scores. In each case, we excluded participants who did not have an available change score and modeled the following predictor variables: age, gender, baseline PHQ-9 or GAD-7 score, total active days, and total days with therapist contact. Since antidepressant status was only available for a subset of participants, we repeated each analysis including and excluding antidepressant status. In addition, we excluded meditation minutes from all models, as this was highly correlated with active days. Predictors associated with a P value<.05 were considered significant.

Results

Demographics and Participation

Participant demographics for the primary group of participants with 12-month PHQ-9 data are presented in Table 1. On an average, participants were 33 years of age, had baseline symptoms above the cutoff for major depression (mean PHQ-9 score=12.8), and were predominantly female. In addition, 80 participants were based in Finland, while 22 were based in the United States. Participants based in the United States had significantly less PHQ-9 symptoms at baseline than those in Finland (United States: mean 8.33, SD 4.40; Finland: mean 13.9, SD 4.82; t_{101}=4.82; P<.001). Approximately one-third of the participants were taking antidepressant medication at the start of the intervention, and 20 participants (19.6%) dropped out of the intervention, where dropout was defined as less than 4 weeks of active participation during the 8-week intervention combined with incomplete PHQ-9 (depression) scores immediately postintervention.

Table 1. Participant demographics and relevant baseline data for the primary study cohort (with Patient Health Questionnaire-9 outcomes available up to 12-months postintervention).

Demographics and baseline data	All participants	Completed 8-week outcomes	Completed 6-month outcomes	Completed 12-month outcomes	Did not complete 8-week outcomes	Did not complete 6-month outcomes	Did not complete 12-month outcomes
Total participants, n (%)	102 (100)	83 (81.4)	44 (43.1)	52 (51)	19 (18.6)	58 (56.9)	50 (49)
Age in years, mean (SD)	32.9 (10.3)	33.2 (10.4)	33.5 (11.0)	31.3 (11.0)	31.9 (10.0)	32.5 (9.9)	34.6 (9.4)
Gender, n (%)							
Male	23 (22.5)	19 (22.9)	11 (25)	10 (19.2)	4 (21.0)	12 (20.7)	13 (26)
Female	79 (77.5)	64 (77.1)	33 (75)	42 (80.8)	15 (79.0)	46 (79.3)	37 (74)
Antidepressants, n (%)							
Yes	25 (24.5)	18 (21.7)	7 (15.9)	7 (13.5)	7 (36.8)	18 (31.0)	18 (36)
No	44 (43.1)	37(44.6)	24 (54.6)	22 (42.3)	7 (36.8)	20 (34.5)	22 (44)
Unknown	33 (32.4)	28 (33.7)	13 (29.6)	23 (44.2)	5 (26.3)	20 (34.5)	10 (20)
Country, n (%)							
Finland	80 (78.4)	67 (80.7)	36 (81.8)	48 (92.0)	13 (68.0)	44 (75.9)	32 (64.0)
Unites States	22 (21.6)	16 (19.3)	8 (18.2)	4 (8.0)	6 (32.0)	14 (24.1)	18 (36.0)
Baseline PHQ-9[a] score, mean (SD)[b]	12.8 (5.2)	13.0 (5.4)	13.2 (5.8)	14.3 (5.0)	11.6 (4.6)	12.5 (4.8)	11.2 (5.0)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]Significantly associated with the presence of PHQ-9 scores at the 12-month follow-up.

Just over half (51%) of all participants reported a PHQ-9 outcome at 12 months postintervention (Table 1). Logistic regression revealed that participants with higher depression scores at baseline (B=–0.11, P=.02) and those who engaged with the intervention on more days (B=–0.05, P=.009) were more likely to complete the PHQ-9 at 12 months postintervention. When including country as a covariate, participants in the United States were less likely to complete PHQ-9 data at 12 months postintervention than participants in

Finland (*B*=1.89, *P*=.01), although this may have been driven by differences in baseline PHQ-9 severity.

We also report participant demographics for a separate group of patients with anxiety outcomes available at the 6-month follow-up in Table 2. On an average, baseline GAD-7 scores suggest that participants had symptoms above the cutoff for moderate anxiety (mean GAD-7 score=10.7), despite being recruited on the basis of symptoms of depression. Age, gender, country, and baseline PHQ-9 symptoms were similar to those of the primary study cohort presented in Table 1. Intervention engagement positively predicted completion of the GAD-7 scale at 6 months postintervention (*B*=–0.07, *P*<.001), but baseline GAD-7 scores did not (*P*=.99).

Table 2. Participant demographics and relevant baseline data for patients with Generalized Anxiety Disorder-7 outcomes available at 6-months postintervention.

Demographics and baseline data	All participants	Completed 8-week outcomes	Completed 6-month outcomes	Did not complete 8-week outcomes	Did not complete 6-month outcomes
Total participants, n (%)	102 (100)	74 (72.6)	45 (44.1)	28 (27.5)	57 (55.9)
Age in years, mean (SD)	31.7 (11.4)	31.5 (11.7)	31.0 (12.0)	32.1 (10.8)	32.2 (10.9)
Gender, n (%)					
Male	16 (15.7)	9 (12.2)	4 (8.9)	7 (25.0)	12 (21.1)
Female	86 (84.3)	65 (87.8)	41 (91.9)	21 (75.0)	45 (78.9)
Antidepressants, n (%)					
Yes	47 (46.1)	37 (50.0)	22 (48.9)	10 (35.7)	25 (43.9)
No	53 (52.0)	36 (48.7)	23 (51.1)	17 (60.1)	30 (52.6)
Unknown	2 (1.9)	1 (1.4)	0 (0)	1 (3.6)	2 (3.5)
Country, n (%)					
Finland	90 (88.2)	71 (95.9)	45 (100)	19 (67.9)	12 (21.1)
United States	12 (11.8)	3 (4.1)	0 (0)	9 (32.1)	45 (78.9)
Baseline GAD-7[a] score, mean (SD)	10.7 (4.9)	11.0 (4.8)	10.8 (4.8)	9.85 (4.9)	10.5 (4.9)
Baseline PHQ-9[b] score, mean (SD)	13.4 (4.9)	13.6 (4.7)	13.1 (4.4)	12.9 (5.5)	13.7 (5.4)

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]PHQ-9: Patient Health Questionnaire-9.

Engagement

Week-by-week intervention engagement rates are summarized in Table 3. On an average, participants engaged with the intervention on 31.3 days (SD 13.5, min=3, max=56), which corresponds to 55.9% of intervention days (where engagement is defined as >3 minutes of app-based mindfulness meditation on a given day). Participants completed an average of 9.79 hours of mindfulness-based exercises (SD 5.01, min=0.79, max=24.7) and had contact with their therapist on 13.1 days (SD 8.34, min=0, max=35) or 23.4% of intervention days. In addition, 68 participants (66.7%) completed at least one app-based meditation practice on each of the 8 weeks of the intervention. As reported in a previous publication [34], the mean number of days of intervention engagement ($F_{7,707}$=42.4, *P*<.001) and the mean number of days of therapist contact ($F_{7,707}$=5.37, *P*<.001) decreased from week 1 to week 8 (Table 3). Of the 102 participants, 81.4% completed the PHQ-9 immediately postintervention, while 43.1% and 51% completed the PHQ-9 at 6 and 12 months postintervention, respectively.

Table 3. Week-by-week intervention engagement across all participants. “Active days” corresponds to >3 minutes of app-based mindfulness practice on a given day.

Metric	Week							
	1	2	3	4	5	6	7	8
Active days, mean (SD)	5.02 (1.81)	4.65 (1.89)	4.54 (1.95)	4.13 (1.95)	3.76 (2.19)	3.59 (2.34)	2.95 (2.21)	2.67 (2.20)
Meditation minutes, mean (SD)	93.9 (39.0)	105.9 (52.6)	80.6 (40.0)	76.1 (47.6)	82.6 (60.5)	67.7 (55.2)	39.0 (39.8)	41.5 (39.5)
Days with therapist contact, mean (SD)	2.02 (1.27)	1.90 (1.53)	1.65 (1.40)	1.75 (1.40)	1.47 (1.42)	1.46 (1.45)	1.43 (1.43)	1.42 (1.37)

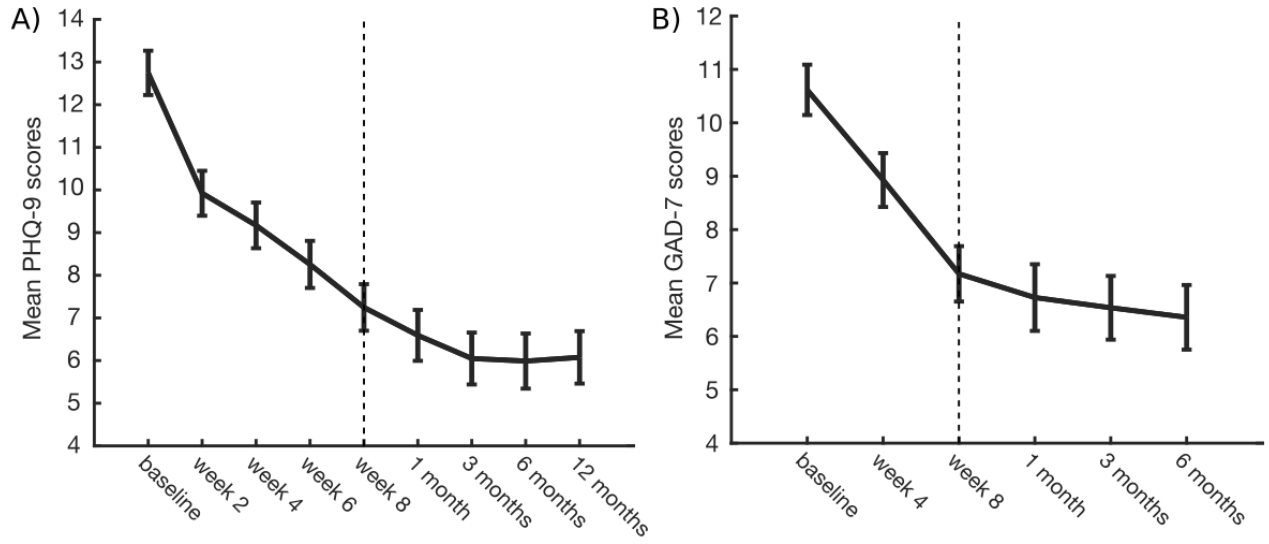
Patient-Reported Outcomes

Depression Symptoms

Participants reported clinically significant improvements in depression symptoms (PHQ-9 scores) from baseline to postintervention (5.50-point reduction, 95% CI 4.58-6.42; *P*<.001, Hedges *g*=1.02, 95% CI 0.71-1.32; Figure 1, panel A). This improvement was maintained at 6 months (6.76-point reduction, 95% CI 5.61-7.90; *P*<.001; Hedges *g*=1.30, 95% CI 0.91-1.68) and 12 months (6.67-point reduction; 95% CI 5.59-7.75; *P*<.001; Hedges *g*=1.14, 95% CI 0.78-1.49) postintervention. The improvement at 6 months remained robust when including an additional 95 participants (total n=197) with outcomes available at 6 months postintervention only (6.53-point reduction, 95% CI 5.73-7.33; *P*<.001; Hedges *g*=1.28, 95% CI 1.00-1.55).

When considering participants with PHQ-9 scores≥10 at baseline (n=83), 48% reported a clinically significant improvement immediately postintervention, which increased to 60% at the 12-month follow-up (defined as postintervention score<10 combined with ≥50% symptom reduction). When including all participants and using the more liberal definition of a 5-point reduction in PHQ-9 symptoms, 56% reported clinically significant improvements immediately postintervention, which increased to 77% at the 12-month follow-up.

Figure 1. Estimated marginal means for PHQ-9 (A) and GAD-7 (B) scores across all available timepoints. The dotted line indicates the last week of the 8-week intervention. Error bars represent standard error of the mean. PHQ-9: Patient Health Questionnaire (9-item version); GAD-7: Generalized Anxiety Disorder (7-item version).



Sensitivity Analysis

We also estimated PHQ-9 scores at 6 and 12 months postintervention under the conservative assumption that participants with missing data did not experience any long-term benefit from the intervention or benefitted only marginally (see Methods). This approach revealed more modest yet significant reductions in PHQ-9 scores relative to baseline at both 6 months (4.35-point reduction, 95% CI 3.65-5.06; *P*<.001; Cohen *d*=0.83, 95% CI 0.43-1.24) and 12 months (4.31-point reduction, 95% CI 3.63-4.99; *P*<.001; Cohen *d*=0.82, 95% CI 0.42-1.23) postintervention.

Anxiety Symptoms

Participants reported clinically significant improvements in anxiety (GAD-7 scores) from baseline to postintervention (3.45-point reduction, 95% CI 2.51-4.38; *P*<.001; Hedges *g*=0.69, 95% CI 0.38-1.00; Figure 1, panel B). This improvement was maintained at 6 months postintervention (4.26-point reduction, 95% CI 3.14-5.38; *P*<.001; Hedge *g*=0.91, 95% CI 0.54-1.28).

Predictors of Outcome Change

Individuals with higher PHQ-9 symptoms (*b*=0.54, *P*<0.001) or higher GAD-7 symptoms (*b*=0.63, *P*<.001) at baseline were likely to experience larger reductions in depression (at the 12-month follow-up) and anxiety (at the 6-month follow-up), respectively. However, none of the reported engagement metrics or participant demographics were predictive of score change from baseline to 6- and 12-month follow-ups, for either PHQ-9 or GAD-7 (all *P*>.05).

Discussion

Principal Findings

The Meru Health *Ascend* intervention is a newly developed, smartphone-based, therapist-supported intervention for depression and anxiety, designed to overcome common barriers to treatment. A recent feasibility study reported that the intervention is feasible and associated with reduced depression symptoms immediately after the intervention [34]. This follow-up study extends these findings by demonstrating that the intervention is associated with clinically significant reductions in symptoms of both depression and anxiety and that these reductions are maintained up to 1 year and 6 months after intervention completion, respectively.

Understanding whether improvements in symptoms associated with *Ascend* are maintained long-term is important, as evidence suggests that a large proportion of patients are likely to relapse

following treatment for depression or anxiety [30,32]. Although relapse rates associated with other smartphone-based interventions for depression and anxiety are largely unknown, previous research suggests that the risk of relapse is substantially lower following psychotherapy than following pharmacotherapy, where as many as 50%-75% of patients are likely to experience significant return of symptoms within 12 months of withdrawal from the latter [53,54]. Our results suggest that, similar to in-person psychotherapy, *Ascend* may be associated with enduring effects that extend for up to 12 months beyond the end of treatment. However, further work is needed to understand the long-term effectiveness of *Ascend* and other smartphone-based interventions for depression and anxiety under controlled conditions and relative to conventional treatment methods.

Our analysis revealed a larger effect size for the reduction in depression symptoms 12 months postintervention (g=1.14) than was reported by a recent meta-analysis [26] comparing smartphone-based interventions for depression to inactive control groups (g=0.56). However, this study was uncontrolled, and we therefore caution the reader not to overinterpret this difference. Indeed, the large effect reported here is consistent with other within-group (uncontrolled) effect sizes for app- and online-based depression interventions [55-57]. Further, a recent meta-analysis reported a similar uncontrolled effect size of g=1.29 for the reduction in depression symptoms associated with transdiagnostic, internet-delivered cognitive-behavioral therapy for depression and anxiety at follow-up [33]. Together, this suggests that the aforementioned discrepancy may stem from the lack of a comparison group in our study. We observed that the reduction in depression symptoms remained significant when making a conservative assumption that participants with missing data did not experience any long-term benefit from the intervention. However, the average reduction in depression under this assumption was marginally below the threshold for clinical significance (5-point change in PHQ-9 symptoms).

Similarly, our study revealed a slightly higher effect size for the reduction in anxiety symptoms immediately after the intervention (g=0.69) and 6 months postintervention (g=0.91) than was recently reported in a meta-analysis [27] comparing smartphone-based interventions for anxiety symptoms to inactive or waitlist control groups (g=0.45) [27]. This discrepancy may also be explained by the fact that this study was uncontrolled, and the effect size reported here may thus be an overestimate. In addition, the interventions included in the aforementioned meta-analysis were approximately 2 weeks shorter (on average) than the current 8-week intervention. Further, this study recruited participants on the basis of elevated symptoms of depression (as opposed to anxiety), making it difficult to make direct comparisons. Interestingly, the slightly higher effect size reported here is consistent with the results from online (as opposed to app-based) psychological interventions for anxiety disorders [58,59] and is, in fact, slightly lower than the effect size reported by a recent meta-analysis [33] of transdiagnostic, internet-delivered cognitive-behavioral therapy for depression and anxiety (uncontrolled pretest to follow-up, g=1.29).

The finding that *Ascend* was associated with reductions in symptoms of anxiety as well as depression is unsurprising, given that the intervention comprises evidence-based practices that have previously been shown to be effective at treating both conditions [60]. In addition, depression and anxiety are highly comorbid, suggesting the presence of shared underlying mechanisms that may respond to similar treatment methods [61]. Indeed, baseline scores suggest that a large proportion of participants were experiencing anxiety symptoms before the intervention, despite being recruited on the basis of having symptoms of depression. However, the magnitude of symptom reduction associated with *Ascend* was slightly larger for depression than for anxiety, likely reflecting the fact that the intervention was designed to address depression as the primary condition. Although previous online psychological interventions have often separated the treatment of depression and anxiety symptoms, there has been a recent trend toward the development of transdiagnostic interventions that target mechanisms common to multiple psychiatric disorders. These results add to a growing body of evidence highlighting the feasibility of such interventions [33] and suggest that smartphone apps may have the potential to address symptoms of depression and anxiety concurrently.

Two important but largely unanswered questions are who is likely to benefit the most from app-based psychological interventions and what factors are predictive of the degree to which individuals benefit over time. Our previous feasibility study suggested that a greater volume of app-based practice predicted the occurrence of fewer depressive symptoms 4 weeks after completing the *Ascend* intervention [34]. In this study, the degree of engagement with *Ascend* did not impact depression or anxiety scores at 6 or 12 months postintervention, although this analysis was likely underpowered due to a large proportion of missing data at these time points. It is also likely that a portion of the variance in scores at these time points was driven by factors not captured by our study. Further research is needed to understand the factors and components of the intervention that are predictive of long-term benefits.

Limitations

As with the former feasibility study [34], this study used a nonrandomized, uncontrolled design, which precludes causal inferences regarding the intervention and patient-reported outcome changes. In addition, the effect sizes reported in this study are likely an overestimate of the true treatment effect relative to an active control or treatment as usual.

Further, approximately one-third of the participants were taking antidepressants during (and presumably after) the intervention, making it difficult to preclude the possibility that the reduction in depression symptoms was caused or maintained by antidepressants [62] or by the tendency for a proportion of depressed patients to naturally recover over a 12-month period [63] and not by this intervention. Future randomized controlled trials that compare *Ascend* to treatment as usual over an extended period are required to fully address these limitations. The Meru Health Online Clinic could also consider collecting information on the antidepressant status at follow-up (and not

just at baseline) to better tease apart the long-term effects associated with the current intervention versus antidepressants.

Further, although engagement with the intervention was high, approximately half of all participants did not complete questionnaires at 6 and 12 months postintervention. Thus, our estimates of the percentage of participants that experienced clinically significant improvements in depression symptoms was based on the participants with complete data only. Further, participants with fewer symptoms of depression postintervention may have been more likely to engage with questionnaires, biasing estimates of the long-term treatment effect. This was addressed by using a robust pattern-mixture model approach and applying the most conservative clinical assumptions about the treatment effect for participants who dropped out or who were lost to follow-up. Nevertheless, such assumptions are inherently unverifiable, and recent guidelines highlight the importance of minimizing the likelihood of incomplete outcome data [64].

Although this study suggests that the intervention is associated with reductions in depression and anxiety symptoms, it does not address the underlying mechanisms or mediators of outcome change. Since the intervention encompasses techniques and practices from multiple approaches (cognitive behavioral therapy, mindfulness meditation, and behavioral activation therapy) as well as remote therapist and peer support, it is difficult to differentiate which components of the intervention are the most effective. Such information could assist with the design and optimization of future iterations of *Ascend* and other app-based interventions for mental health [65].

Finally, the study included self-selected participants who may have shown a higher degree of motivation to engage with the intervention. Further, the majority of participants were female, and although the study included individuals from both Finland and the United States, the majority of participants were based in Finland. Moreover, the Meru Health Clinic does not currently track patient race and ethnicity, making it difficult to assess the generalizability of the findings. Together, this suggests that this study sample may not be representative of the wider population of individuals with elevated symptoms of depression or anxiety, and future studies should address this issue by using more robust recruitment strategies and more representative study samples.

Conclusions

Depression is a serious and growing problem that causes individual suffering and huge economic and societal costs worldwide. Many individuals with depression are unable to access appropriate treatment, with high costs and a lack of trained professionals being major barriers. Scalable, low-cost, app-based interventions such as *Ascend* are designed to overcome these barriers and may help to significantly reduce the burden of anxiety and depression. Further research is needed to investigate the efficacy of *Ascend* in comparison to control groups and other established treatments for depression and anxiety.

Acknowledgments

We wish to express our gratitude to all participants for their efforts and to the staff and therapists at Meru Health who were involved in developing, implementing, and supporting the study intervention.

Authors' Contributions

ME, KR, AR, and VF-H conceived and designed the study. ME performed statistical analysis, interpreted the data, and drafted the manuscript. KR, AN, OH, and AR oversaw the design and creation of the study intervention. AN and OH oversaw data collection and archiving. VF-H contributed to interpretation of the data. All authors revised the manuscript and approved the final content.

Conflicts of Interest

All authors are employed, receive a salary, and/or hold equity at Meru Health Inc. KR serves as the chief executive officer of Meru Health Inc and owns a large share of the company’s stocks. AN serves as the chief technology officer of Meru Health Inc and owns a large share of the company’s stocks.

References

1. 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]
2. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]
3. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007 Sep 8;370(9590):851-858. [doi: 10.1016/S0140-6736(07)61415-9] [Medline: 17826170]
4. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]

5. Papakostas GI, Ionescu DF. Updates and trends in the treatment of major depressive disorder. J Clin Psychiatry 2014 Dec;75(12):1419-1421. [doi: 10.4088/JCP.14ac09610] [Medline: 25551237]
6. Wade AG, Häring J. A review of the costs associated with depression and treatment noncompliance: the potential benefits of online support. Int Clin Psychopharmacol 2010 Sep;25(5):288-296. [Medline: 20715299]
7. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018 Dec 07;391(10128):1357-1366. [doi: 10.1016/S0140-6736(17)32802-7] [Medline: 29477251]
8. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]
9. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
10. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
11. SAMHSA: Substance Abuse and Mental Health Services Administration. 2017 Sep 01. Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/report/key-substance-use-and-mental-health-indicators-united-states-results-2016-national-survey [accessed 2019-02-01]
12. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
13. Weil TP. Insufficient dollars and qualified personnel to meet United States mental health needs. J Nerv Ment Dis 2015 Apr;203(4):233-240. [doi: 10.1097/NMD.0000000000000271] [Medline: 25816044]
14. El-Hage W, Leman S, Camus V, Belzung C. Mechanisms of antidepressant resistance. Front Pharmacol 2013 Nov 22;4:146 [FREE Full text] [doi: 10.3389/fphar.2013.00146] [Medline: 24319431]
15. Wang S, Han C, Bahk W, Lee S, Patkar AA, Masand PS, et al. Addressing the Side Effects of Contemporary Antidepressant Drugs: A Comprehensive Review. Chonnam Med J 2018 May;54(2):101-112 [FREE Full text] [doi: 10.4068/cmj.2018.54.2.101] [Medline: 29854675]
16. Pigott HE. STAR*D: A Tale and Trail of Bias. Ethical Human Psychology and Psychiatry 2011;13(1):6-28.
17. Coplan JD, Aaronson CJ, Panthangi V, Kim Y. Treating comorbid anxiety and depression: Psychosocial and pharmacological approaches. World J Psychiatry 2015 Dec 22;5(4):366-378. [doi: 10.5498/wjp.v5.i4.366] [Medline: 26740928]
18. Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
19. Grant K, McMeekin E, Jamieson R, Fairfull A, Miller C, White J. Individual therapy attrition rates in a low-intensity service: a comparison of cognitive behavioural and person-centred therapies and the impact of deprivation. Behav Cogn Psychother 2012 Mar;40(2):245-249. [doi: 10.1017/S1352465811000476] [Medline: 21851754]
20. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
21. Poushter J. Pew Research Center: Global Attitudes & Trends. 2016 Feb. Smartphone ownership and internet usage continues to climb in emerging economies URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/
22. Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
23. Torous J, Chan SR, Yee-Marie TS, Behrens J, Mathew I, Conrad EJ, et al. Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
24. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
25. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
26. 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]

27. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
28. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
29. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
30. Rickels K, Etemad B, Khalid-Khan S, Lohoff FW, Rynn MA, Gallop RJ. Time to relapse after 6 and 12 months' treatment of generalized anxiety disorder with venlafaxine extended release. Arch Gen Psychiatry 2010 Dec;67(12):1274-1281. [doi: 10.1001/archgenpsychiatry.2010.170] [Medline: 21135327]
31. Johansson O, Lundh L, Bjärehed J. 12-Month Outcome and Predictors of Recurrence in Psychiatric Treatment of Depression: A Retrospective Study. Psychiatr Q 2015 Sep;86(3):407-417. [doi: 10.1007/s11126-015-9341-y] [Medline: 25597030]
32. Sim K, Lau WK, Sim J, Sum MY, Baldessarini RJ. Prevention of Relapse and Recurrence in Adults with Major Depressive Disorder: Systematic Review and Meta-Analyses of Controlled Trials. Int J Neuropsychopharmacol 2015 Jul 07;19(2). [doi: 10.1093/ijnp/pyv076] [Medline: 26152228]
33. Pasarelu CR, Andersson G, Bergman NL, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
34. Goldin PR, Lindholm R, Ranta K, Hilgert O, Helteenvuori T, Raevuori A. Feasibility of a Therapist-Supported, Mobile Phone-Delivered Online Intervention for Depression: Longitudinal Observational Study. JMIR Form Res 2019 Jan 22;3(1):e11509 [FREE Full text] [doi: 10.2196/11509] [Medline: 30682726]
35. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness meditation for everyday life. New York City: Hachette Books; 1994.
36. Morgan D. Mindfulness-based cognitive therapy for depression: a new approach to preventing relapse. Psychother Res 2003 Mar;13(1):123-125. [doi: 10.1080/713869628] [Medline: 22475168]
37. Beck AT. Cognitive Therapy and the Emotional Disorders. London, UK: Penguin; 1979.
38. Jacobson NS, Martell CR, Dimidjian S. Behavioral Activation Treatment for Depression: Returning to Contextual Roots. Clinical Psychology: Science and Practice 2001;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]
39. Datica. URL: https://datica.com/
40. 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]
41. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
42. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord 2010 Dec;127(1-3):122-129. [doi: 10.1016/j.jad.2010.04.030] [Medline: 20569992]
43. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
44. 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]
45. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
46. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials. Psychosom Med 2016;78(6):716-727 [FREE Full text] [doi: 10.1097/PSY.0000000000000322] [Medline: 27187854]
47. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 2015;67(1):1-48.
48. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J 2008 Jun;50(3):346-363. [doi: 10.1002/bimj.200810425] [Medline: 18481363]
49. Baraldi AN, Enders CK. An introduction to modern missing data analyses. J Sch Psychol 2010 Feb;48(1):5-37. [doi: 10.1016/j.jsp.2009.10.001] [Medline: 20006986]
50. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharm Stat 2013;12(6):337-347. [doi: 10.1002/pst.1549] [Medline: 23292975]
51. Fox J, Weisberg S. An R Companion to Applied Regression. Thousand Oaks, CA: SAGE Publications; 2018.
52. Ratitch B, OKelly M. Pattern-Mixture Models as Linear Combinations of Least Squares Means from MMRM with Delta Method Variance Estimation. 2012 Presented at: PharmaSUG 2012; May 13-16, 2012; San Francisco, CA.

53. Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'Reardon JP, et al. Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. Arch Gen Psychiatry 2005 Apr;62(4):417-422. [doi: 10.1001/archpsyc.62.4.417] [Medline: 15809409]
54. Evans MD, Hollon SD, DeRubeis RJ, Piasecki JM, Grove WM, Garvey MJ, et al. Differential relapse following cognitive therapy and pharmacotherapy for depression. Arch Gen Psychiatry 1992 Oct;49(10):802-808. [doi: 10.1001/archpsyc.1992.01820100046009] [Medline: 1417433]
55. Schlosser DA, Campellone TR, Truong B, Anguera JA, Vergani S, Vinogradov S, et al. The feasibility, acceptability, and outcomes of PRIME-D: A novel mobile intervention treatment for depression. Depress Anxiety 2017 Jun;34(6):546-554 [FREE Full text] [doi: 10.1002/da.22624] [Medline: 28419621]
56. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. Internet Interventions 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
57. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13(1):49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
58. Hofmann SG, Smits JAJ. Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. J Clin Psychiatry 2008 Apr;69(4):621-632 [FREE Full text] [Medline: 18363421]
59. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
60. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety - a systematic review. Health Psychol Behav Med 2014 Jan 01;2(1):841-881 [FREE Full text] [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
61. Zbozinek TD, Rose RD, Wolitzky-Taylor KB, Sherbourne C, Sullivan G, Stein MB, et al. Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. Depress Anxiety 2012 Dec;29(12):1065-1071 [FREE Full text] [doi: 10.1002/da.22026] [Medline: 23184657]
62. Simon GE. Long-term prognosis of depression in primary care. Bull World Health Organ 2000;78(4):439-445 [FREE Full text] [Medline: 10885162]
63. Spijker J, de Graaf R, Bijl RV, Beekman ATF, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 2002 Sep;181:208-213. [doi: 10.1192/bjp.181.3.208] [Medline: 12204924]
64. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med 2012 Oct 4;367(14):1355-1360 [FREE Full text] [doi: 10.1056/NEJMsr1203730] [Medline: 23034025]
65. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Hollon SD. Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. Psychother Res 2019 Dec;29(1):15-29. [doi: 10.1080/10503307.2017.1395922] [Medline: 29115185]

Abbreviations

EMM: estimated marginal mean
GAD-7: Generalized Anxiety Disorder (7-item version)
HIPAA: Health Insurance Portability and Accountability Act
LMM: linear mixed effects model
PHQ-9: Patient Health Questionnaire (9-item version)
PMM: pattern mixture model

Edited by A Aguilera; submitted 05.04.19; peer-reviewed by HMH Lo, J Nicholas, A Baglione, A Neal-Barnett; comments to author 28.05.19; revised version received 26.06.19; accepted 24.07.19; published 26.08.19.

Please cite as:
Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, Forman-Hoffman V
Long-Term Outcomes of a Therapist-Supported, Smartphone-Based Intervention for Elevated Symptoms of Depression and Anxiety: Quasiexperimental, Pre-Postintervention Study
JMIR Mhealth Uhealth 2019;7(8):e14284
URL: http://mhealth.jmir.org/2019/8/e14284/
doi:10.2196/14284
PMID:31452521

Original Paper

Predictors of Retention in an Adult Text Messaging Smoking Cessation Intervention Program: Cohort Study

Kara P Wiseman[1], MPH, PhD; Kisha I Coa[2], MPH, PhD; Yvonne M Prutzman[1], MPH, PhD

[1]Tobacco Control Research Branch, Behavioral Research Program, National Cancer Institute, Bethesda, MD, United States

[2]ICF, Fairfax, VA, United States

Corresponding Author:
Kara P Wiseman, MPH, PhD
Tobacco Control Research Branch
Behavioral Research Program
National Cancer Institute
9606 Medical Center Drive, MSC 9761
Room 3E208
Bethesda, MD, 20892-9761
United States
Phone: 1 2402765873
Fax: 1 240 276 7894
Email: kara.wiseman@nih.gov

Abstract

Background: Mobile health tools such as text messaging programs can support smoking cessation. However, high rates of disengagement from these tools decrease their effectiveness.

Objective: The purpose of this study was to identify user characteristics associated with retention in an adult text messaging smoking cessation intervention.

Methods: Adults initiating a quit attempt using the publicly available program SmokefreeTXT between March 6 and June 21, 2016 (n=6215), were included. Data were collected to assess nicotine dependence, frequency of being around other smokers, time of the day for cigarette cravings, extrinsic and intrinsic motivation to quit smoking, confidence in quitting, and long-term intention to be smoke free. Multivariable survival analysis modeling for time to opt out was conducted to identify characteristics associated with opting out over the course of the intervention, adjusting for age, sex, and smoking frequency, reset of the quit date by the user, and the number of days enrolled before initiating the quit attempt. Among those who opted out, multivariable multinomial logistic regression analysis was used to identify predictors of opting out early (within 3 days and between 4 and 7 days into the quit attempt) compared to opting out late (more than 7 days into the quit attempt), adjusting for the same confounders.

Results: Survival analyses indicated that younger age, female sex, higher levels of nicotine dependence, lower intention to be smoke free, and enrolling in SmokefreeTXT ≤1 week before initiating the quit attempt were associated with an increased risk of opting out. For example, users who smoked within 5 minutes of waking up were 1.17 times more likely to opt out than those who smoked more than 5 minutes after waking up (95% CI 1.01-1.35). Among users who opted out from SmokefreeTXT, logistic regression modeling indicated that compared to users who were never or rarely around other smokers, those who were sometimes around other smokers had 1.96 times more likely to opt out within the first 3 days of the quit attempt (95% CI 1.18-3.25). In addition, compared to users with high levels of long-term quit intention, users with lower levels of intention had 1.80 times the odds of opting out between 4 and 7 days into the quit attempt (95% CI 1.02-3.18). Users who reset their quit date after initiating a quit attempt were less likely to opt out at either time point compared with those who did not reset their quit date.

Conclusions: Several user characteristics are associated with retention in an adult text messaging smoking cessation program. These results provide guidance on potential characteristics that should be addressed in future text messaging smoking cessation programs. Providing additional support to users with these characteristics may increase retention in text messaging programs and ultimately lead to smoking cessation.

(JMIR Mhealth Uhealth 2019;7(8):e13712) doi:10.2196/13712

KEYWORDS

smoking cessation; text-messaging; mHealth; engagement

Introduction

Although smoking rates continue to decline, smoking is responsible for more than 480,000 deaths annually in the United States [1]. Mobile health (mHealth) smoking cessation programs are a cost-effective way to reach many smokers, particularly smokers who are underserved by traditional treatment modalities (eg, in-person or telephone counseling) [2,3]. Text messaging smoking cessation programs have been found to as much as double a smoker's likelihood of quitting [4,5]; however, high user dropout rates reduce the potential benefit of these programs [6,7]. A recent study found that smokers who left a text-based cessation program were less likely to abstain from smoking at 6 months posttreatment compared with those who remained in the program for longer [8], suggesting that increasing retention in a cessation program may increase the likelihood of quitting.

There are many potential reasons for someone to leave a cessation program prior to program completion, including failure to initiate a quit attempt or relapse, low initial motivation to quit, a high level of nicotine dependence and withdrawal symptoms, inability to resist cravings, or lack of sufficient social support during a quit attempt [9,10]. Text-based cessation programs have the potential to address some of these issues, but it is still unclear which users are more likely to leave a cessation program, and to date, the most relevant characteristics related to opting out of text messaging cessation programs have not been identified [4]. A better understanding of the characteristics associated with opting out may illuminate potential avenues for program enhancement and help identify the types of program users who may benefit most from new or additional program content.

The purpose of this study was to determine whether potentially relevant and addressable user characteristics measured during program enrollment were associated with opting out from SmokefreeTXT, one of the most widely used text-based smoking cessation programs available in the United States.

Methods

SmokefreeTXT Cessation Program

SmokefreeTXT is a free, nationally available, fully automated text-messaging smoking cessation program for adults run by the National Cancer Institute [7]. Smokers interested in quitting smoking can sign up for the program using a Web enrollment form or short message service (text) opt in. The program includes up to 2 weeks of preparation messages and 6 weeks of postquit date messages, and messages vary in content and frequency relative to the quit date set by the user. Text messages provide general motivation support, tips on preparing to quit, advice on managing cravings, suggestions for smoke-free activities, relevant smoking facts, and recognition of cessation milestones [7]. The program is bidirectional; users can text keywords (ie, "crave," "slip," or "mood") at any time to receive on-demand support. Self-reported smoking status, mood, and craving levels are measured using within-program assessment questions delivered during the program. Users can reset their quit date at any time (eg, after a slip) to restart the program. Users can opt out of the program at any time by texting "STOP." Since the program began in 2011, over 150,000 people have enrolled in SmokefreeTXT [11].

Data Collection and Study Population

Design and Measures

Data on age, sex, and smoking frequency are routinely collected from all SmokefreeTXT users at the time of program enrollment. For this study, eight additional baseline items, adapted from validated scales when available, were added to examine smoking context and motivational characteristics associated with opting out. The smoking context characteristics measured the level of nicotine dependence (measured as time to first cigarette: "How soon after you wake up do you smoke your first cigarette?" [12]), frequency of reminders to smoke ("My life is full of reminders to smoke" [13]), time of the day for cigarette cravings ("When do you crave cigarettes the most?" [13]), and frequency of being around other smokers ("How often are you around people who are smoking?"). The motivational characteristics measured extrinsic motivation for quitting smoking ("I would try to quit smoking because others want me to quit smoking" [14]), intrinsic motivation for quitting smoking ("I would try to quit smoking because quitting smoking is an important thing for me to do" [14]), confidence in quitting ("I feel able to meet the challenge of quitting smoking"), and long-term cessation intention ("I intend to be smokefree one year from now" [15]). To minimize respondent burden, only single-item measures were used, and each SmokefreeTXT user received only two of the eight additional questions at the time of enrollment. Specifically, with each page refresh, two items were randomly selected from the bank of eight items for inclusion on the Web enrollment form. Under the National Institutes of Health policy, assessment of quality improvement processes does not require institutional review board approval and is covered by the Terms of Agreements that users agree to when signing up for the SmokefreeTXT program [16]. All data were nonidentifiable, and unique user identifiers replaced user telephone numbers before data were received from the text-message program vendor.

Study Population

The study population consisted of the first quit attempt of all users who signed-up for SmokefreeTXT using the Web enrollment form between March 6, 2016, and June 21, 2016 (N=6831 unique users). Of these, 616 were excluded (571 opted out of SmokefreeTXT before reaching their initial quit date, 41 set quit dates outside of the 2-week window [range: 46 days until the initial quit date to 2 days after a quit date], and 4 had unreliable ages [eg, age >99 years]), leaving 6215 users in the analytic sample.

Analysis

Coding

Opting Out of the Program

Two outcome variables were created to measure the timing for opting out of the program. The date and time any user texts "STOP" to opt out are saved, as are the date and time of enrollment. The number of days that a user was enrolled before opting out was determined by calculating the difference between a user's initial quit date and the date of opting out. Users can reset their quit date by texting "NEW" to SmokefreeTXT. Receipt of this message is date and time stamped and when this occurs, a new row of data is created to capture program data relative to the new quit date. If users reset their quit date within the program, the program data capturing their most up-to-date quit date were used to determine timing of opting out. For users who did not opt out, time to opt out was censored at 42 days (length of the full 6-week intervention starting on the quit date). Previous findings show that the majority of users opting out of SmokefreeTXT do so in the first 3 days and up to the first week after initiating their quit attempt [7]; therefore, to understand the user characteristics associated with opting out early compared to opting out later in the program, a three-level variable was created among users who opted out of the program (n=3259) in order to investigate opting out within 3 days after the initial quit date and opting out between 4 and 7 days after the initial quit attempt compared with opting out more than 7 days after the initial quit date.

Characteristics of Interest

The eight baseline items added to the Web enrollment form were the primary characteristics of interest for this study and were coded using the distribution of responses in the study population. These characteristics were level of nicotine dependence (two levels: smokes within 5 minutes of waking up and smokes more than 5 minutes after waking), frequency of reminders to smoke (two levels: very true and a little true or a little/very untrue), time of the day for cigarette cravings (two levels: craves cigarettes in the morning, afternoon, or evening and craves cigarettes the same amount all times of day), frequency around other smokers (three levels: never or rarely around other smokers, sometimes around other smokers, and very often around other smokers), extrinsic motivation for quitting smoking (three levels: very true, a little true, and a little/very untrue), intrinsic motivation for quitting smoking (two levels: very true and a little true or a little/very untrue), confidence in quitting (three levels: very true, a little true, and a little/very untrue), and long-term cessation intention (two levels: strongly agree and agree or disagree/strongly disagree).

Potential Confounders

Information collected from users at the time of program enrollment were considered as potential confounders: age (18-29, 30-39, 40-49, and ≥50 years), sex (male or female), and smoking frequency (smokes every day or smokes less often than every day). Two additional characteristics were also included. First, a variable was created to indicate if a user had ever reset their quit date (ie, yes [used the keyword "NEW" coded as yes] or no). Lastly, the number of days that a user was enrolled before starting their cessation attempt was included to capture receipt of cessation preparation messages (0, 1-7, and 8-14 days).

Statistical Analyses

SAS 9.4 (SAS Institute, Cary NC) was used for all analyses. Descriptive analyses were performed using Chi-square tests. Multivariable survival analysis modeling for days to opt out was used to determine user characteristics associated with an increased risk of opting out. For all models, an initial confounder-only model (age, sex, smoking frequency, reset of the quit date by the user, and days enrolled before starting the quit attempt) was created. Thereafter, eight separate models were created with each independent variable of interest. Due to the data collection design, each user only answered two of the eight items of interest, with each item being randomly pulled from the bank of the eight items; thus, information on all eight items is not available for each user, and multiple independent variables of interest could not be included in the same model. Frequency around other smokers had a violation of the proportional hazards assumption; further investigation revealed that the assumption was violated at 14 days after the quit date; thus, two models were created for this user characteristic—one for opting out in the first 14 days and one for opting out after 14 days. Models provided adjusted hazard ratios and 95% CIs. Among users who opted out, multivariable multinomial logistic regression modeling was performed to examine the association between each user characteristic and opting out within 3 days (or between 4 and 7 days) compared to opting out after 7 days. Again, an initial confounder-only model was created and each user characteristic of interest was added to the confounder-only model. Logistic regression models provided odds ratios and 95% CIs.

Results

Overall, the majority of SmokefreeTXT users were women (69.4%), smoked every day (92.4%), had frequent reminders to smoke (68.7%), and craved cigarettes at all times of the day (67.1%) (Table 1). Most users reported high extrinsic and intrinsic motivation for quitting smoking (43.1% and 91.2% reported very true, respectively). Just over one-third of the users signed up for the program on their quit date (37.5%). In addition, 81% of users did not reset their quit date during the program. Slightly more than half (52.4%) opted out during the 6-week course of the program. User characteristics by the timing of the opt-out variable can be found in Multimedia Appendix 1.

Table 1. Descriptive characteristics of SmokefreeTXT users from March 3, 2016, to June 21, 2016.

Variable	Total, n (%)	Opted out during the program		*P* value[a]
		Yes, n (%)	No, n (%)	
Number of users	6215 (100)	3259 (52.4)	2956 (47.6)	
Age (years)				<.001[b]
18-29	1823 (29.3)	1052 (32.3)	771 (26.1)	
30-39	1823 (29.3)	957 (29.4)	866 (29.3)	
40-49	1288 (20.7)	659 (20.2)	629 (21.3)	
≥50	1281 (20.6)	591 (18.1)	690 (23.3)	
Sex				.009[b]
Male	1904 (30.6)	951 (29.2)	953 (32.2)	
Female	4311 (69.4)	2308 (70.8)	2003 (67.8)	
Smoking frequency[c]				.16
less often than every day	468 (7.6)	231 (7.2)	237 (8.1)	
Every day	5672 (92.4)	2992 (92.8)	2680 (91.9)	
Time to first cigarette (min)[d, e]				.13
>5	1025 (62.7)	511 (60.9)	514 (64.6)	
≤5	610 (37.3)	328 (39.1)	282 (35.4)	
Frequent reminders to smoke[d]				.90
Not true[f]	472 (31.3)	252 (31.5)	220 (31.2)	
Very true	1035 (68.7)	549 (68.5)	486 (68.8)	
Frequency around other smokers[d]				.13
Never or rarely	295 (18.8)	145 (18.1)	150 (19.5)	
Sometimes	506 (32.2)	247 (30.8)	259 (33.7)	
Very often	771 (49.1)	411 (51.2)	360 (46.8)	
Craves cigarettes at a specific time of day[d]				.46
No	1023 (67.1)	536 (67.9)	487 (66.2)	
Yes	502 (32.9)	253 (32.1)	249 (33.8)	
Extrinsic motivation to quit[d]				.56
Very true	685 (43.1)	347 (42.5)	338 (43.8)	
A little true	502 (31.6)	259 (31.7)	243 (31.5)	
A little or very untrue	402 (25.3)	211 (25.8)	191 (24.7)	
Intrinsic motivation to quit[d]				.19
Not true[f]	134 (8.8)	79 (9.7)	55 (7.8)	
Very true	1392 (91.2)	739 (90.3)	653 (92.2)	
Confidence in quitting smoking[d]				.35
A little or very untrue	295 (19.0)	163 (19.6)	132 (18.4)	
A little true	652 (42.0)	354 (42.5)	298 (41.5)	
Very true	605 (39.0)	316 (37.9)	289 (40.2)	
Long-term quit intention[d, g]				.21
Other responses[h]	182 (12.6)	105 (13.6)	77 (11.4)	

Variable	Total, n (%)	Opted out during the program		*P* value[a]
		Yes, n (%)	No, n (%)	
Strongly agree	1266 (87.4)	667 (86.4)	599 (88.6)	
Reset quit date during quit attempt				<.001[b]
No	5032 (81.0)	2757 (84.6)	2275 (77.0)	
Yes	1183 (19.0)	502 (15.4)	681 (23.0)	
Days enrolled before starting quit attempt				<.001[b]
0	2332 (37.5)	1316 (40.4)	1016 (34.4)	
1-7	2545 (41.0)	1361 (41.8)	1184 (40.1)	
8-14	1338 (21.5)	582 (17.9)	756 (25.6)	

[a]*P* value from the Chi-square test.

[b]These values are statistically significant at an alpha level of .05.

[c]Sum does not add to the total due to missing values.

[d]Sum does not add to the total, as users were only given two of eight items at sign up (see Methods section for details).

[e]Time to first cigarette after waking up in the morning.

[f]A little true, a little untrue, or very untrue.

[g]Users were asked about their intention to be smoke free 1 year from signing up.

[h]Agree, disagree, or strongly disagree.

Survival analyses revealed that, except for smoking frequency, each variable included in the confounder-only model was associated with opting out (Figure 1, Multimedia Appendix 2). Specifically, women, users younger than 50 years of age, and users who signed up for SmokefreeTXT on their quit date or within 1 week of their quit date were more likely to opt out than others (Figure 1A). Users who reset their quit date after initiating their attempt were less likely to opt out than those who did not reset their quit date. Among the characteristics of interest, people who smoked within 5 minutes of waking up, were sometimes around other smokers, and had lower long-term quit intention were more likely to opt out (Figure 1B). Specifically, users who smoked within 5 minutes of waking were 1.17 times more likely to opt out than those who smoked more than 5 minutes after waking (95% CI 1.01-1.35). Users with less than high long-term intention to be smoke free were 1.29 times more likely to opt out than those with the highest levels of long-term quit intention (95% CI 1.04-1.59). Compared to users who were never or rarely around other smokers, those who were sometimes around other smokers were 1.34 times more likely to opt out during the first 14 days of the program (95% CI 1.03-1.73).

Among all users who opted out, several characteristics were associated with opting out early (ie, within the first 3 days of the quit attempt or between 4 and 7 days into the quit attempt vs later; Figure 2, Multimedia Appendix 3). Specifically, compared to users aged ≥50 years, those aged 18-29 years had 1.33 and 1.45 times the odds of opting out within 3 days or between 4 and 7 days into the quit attempt, respectively (95% CI 1.04-1.69 and 95% CI 1.08-1.94, respectively; Figure 2A). Users who reset their quit date after initiating the quit attempt were less likely to opt out of SmokefreeTXT early compared to users who did not reset their quit date (within 3 days: odds ratio 0.08, 95% CI 0.06-0.12; between 4 and 7 days: odds ratio 0.17, 95% CI 0.12-0.25). Compared to users who were never or rarely around other smokers, those who were sometimes around other smokers had 1.96 times the odds of opting out within the first 3 days (95% CI 1.18-3.25; Figure 2B). Lastly, compared to users with the highest levels of long-term quit intention, users with lower levels of long-term quit intention had 1.80 times the odds of opting out between 4 and 7 days into the quit attempt (95% CI 1.02-3.18).

Figure 1. Adjusted survival analysis describing predictors of opting out of SmokefreeTXT, presenting the results of 10 adjusted models. Full model information available is in Multimedia Appendix 2. Panel A presents results from one confounder-only model (age, sex, smoking frequency, reset of quit date by the user, and days enrolled before start of the quit attempt). Panel B presents nine separate survival models with all confounders plus each user characteristic of interest. Violation of the proportional hazards assumption was found for the user characteristic "frequency around other smokers"; models were stratified at 14 days at the point where violation occurred and are presented separately.

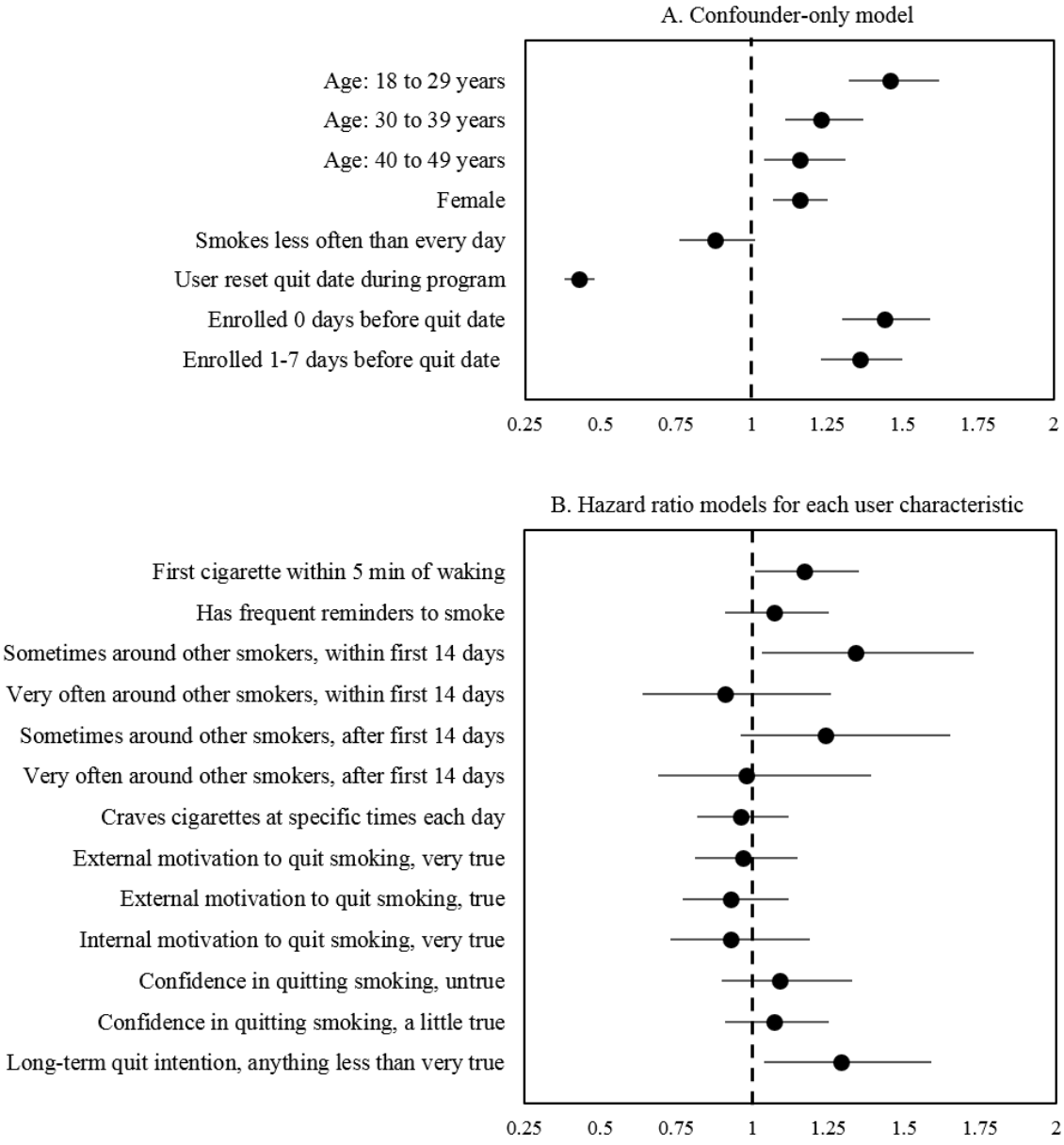
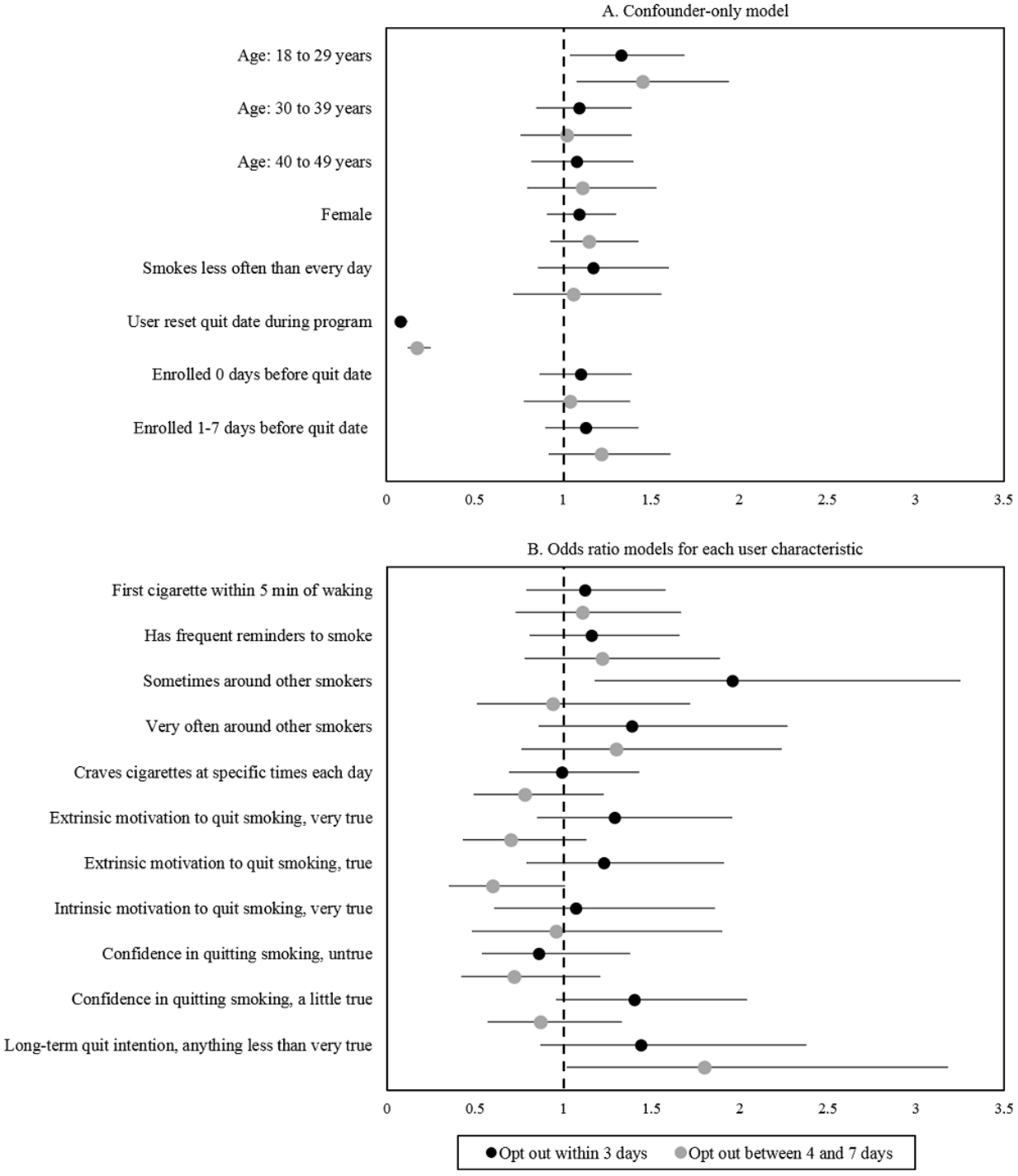


Figure 2. Multivariable multinomial logistic regression model for users who opted out of SmokefreeTXT, comparing users who opted out within 3 days and between 4 and 7 days to those opting out after 7 days. This figure presents results of nine adjusted models. Full model information is available in Multimedia Appendix 3. Panel A presents results from one confounder-only model (age, sex, smoking frequency, reset of quit date by the user, and days enrolled before start of the quit attempt). Panel B presents eight separate logistic regression models with all confounders plus each user characteristic of interest.



Discussion

This study assessed smoking context and motivational characteristics associated with opting out of the SmokefreeTXT program in two ways: opting out over the entire cessation intervention and opting out at specific periods early in a cessation attempt. Overall, almost 50% of SmokefreeTXT users were retained in the program over the course of the 6-week intervention. Several user characteristics seemed to influence retention in SmokefreeTXT; these were found to vary in terms of strength of the association and timing relative to the cessation attempt. Specifically, younger age, female sex, and higher levels of nicotine dependence were associated with increased rates of opting out over the entire intervention. Being around other smokers was uniquely associated with opting out early in the quit attempt. Lower long-term intention to stay smoke free was

associated with increased rates of opting out over the entire intervention and had a particularly pronounced association with retention in days 4-7 of the quit attempt.

To our knowledge, this is the first study to identify characteristics associated with opting out by using data from a real-world implementation of a text messaging smoking cessation program. Previous analyses of SmokefreeTXT user data found similar rates of opting out among users enrolled between 2012 and 2014 [7]. Therefore, it appears that program changes may be needed to increase retention in this widely used program. The predictors of opting out found in this study are largely consistent with those shown to be predictive of relapse following smoking cessation [17-21], which suggests that users may be leaving the program because they have returned to smoking. This potential pattern of opting out in parallel with relapse is supported by an analysis by Heminger et al [8], who found that participants who opted out of the text messaging smoking cessation program Text2Quit were less likely to be abstinent at 6 months compared with those who did not opt out of the program. It is uncertain how generalizable their results are to the SmokefreeTXT user population, as their analysis occurred in the context of a clinical trial and most participants remained enrolled in the program. The potential differences in how users interact with text-messaging programs within and outside a research setting are potentially large, given that in the study by Heminger et al [8], only 30.2% of participants texted "STOP" [8], yielding a 22% difference between their study and the opt out rates found in the SmokefreeTXT program. To date, there has been no real-world evaluation of the association between program retention and smoking abstinence. The results of this study identified several user characteristics of interest that could be addressed with modifications to the currently available text messaging cessation program.

It is possible that addressing the characteristics associated with opting out might improve retention in the SmokefreeTXT program, which could be done by going beyond the currently available "one-size-fits-all" approach, by potentially including elements of tailoring. Tailoring to specific influential characteristics may increase the relevance of the program to each user and provide more salient tips and resources, thereby increasing retention [22,23]. Previous studies have shown the effectiveness of tailored text message–based cessation interventions [24-26], but to our knowledge, none have been implemented outside a research setting. Further, only one study has compared tailored to nontailored text messaging cessation programs [27]. Thus, the full potential impact of tailoring within text messaging cessation programs and the features requiring tailoring the most are unknown. The results of the survival analysis provide information about the characteristics of users who had significantly higher rates of opting out over the entire 42-day SmokefreeTXT program, highlighting several characteristics that could be the focus of tailoring throughout text messaging–based cessation programs.

Users younger than 50 years of age were more likely to opt out of SmokefreeTXT. Further, those aged 18-29 years had the highest risk of opting out over the entire intervention and early in the quit attempt. Previous research among young adult smokers have identified several unique features that may make quitting more difficult in this group. For example, there is evidence that young adults' identity as smokers continues to evolve [28] and young adults who smoke frequently might still not identify as smokers [29-33]. This could be a result of younger smokers having lower rates of daily smoking compared to older adult smokers [1]. It is therefore possible that even while making a quit attempt, traditional cessation programs might not resonate with young adult smokers. Previous evaluations of motivations to quit smoking among young adult smokers have identified physical fitness and worry about long-term health effects of smoking as motivators to quit [34,35]. Thus, incorporating additional intervention content in these areas may assist in keep young adult smokers engaged in a cessation program. In addition, it is possible that identifying ways to help reinforce a young adult's identity as a smoker could maintain motivation to quit in order to avoid the long-term health harms already associated with smoking [36]. Lastly, young adults have the highest rates of technology adoption [37]. Although a text messaging program might align with how young adults are using devices, it may be harder for an intervention to resonate with high-technology adopters if it does not include the most up-to-date functionality. Continued exploration of the barriers to quitting and remaining engaged in a cessation attempt among this group is imperative to provide up-to-date, relevant content within cessation programs.

Logistic regression analyses identified two characteristics found to be particularly important early in the quit attempt: being around other smokers sometimes and having low long-term quit intention. However, the timing of these associations differed, even within the first week of the quit attempt. Including messages to address these characteristics within the first days of a quit attempt may increase retention. Frequency of being around other smokers can be an indicator of how frequently a smoker initiating a cessation attempt might experience cues and urges to smoke, which can derail a quit attempt [38-40]. Being around other smokers may also present a perceived smoking opportunity to smokers undertaking a quit attempt, which can also increase cravings to smoke [41]. In this study, compared to users who were never around other smokers, only those who were sometimes around other smokers had higher odds of opting out. There was no increased risk of opting out for users who were frequently around other smokers, compared to users who were never around other smokers. This finding could indicate that people who are only sometimes around other smokers (almost one-third of SmokefreeTXT users) face unexpected cravings that they are not prepared for and indicates the potential need for additional skill building early in the quit attempt to help plan for and resist unexpected cravings. One way to do this could be to provide additional reminders about the on-demand keyword functionality embedded within text message programs to people who are only intermittently around other smokers in order to assist with cravings and increase the use of real-time support during their quit attempt.

People with lower levels of long-term quit intention had an increased risk of opting out over the entire intervention and had higher odds of opting out between days 4 and 7 of the quit attempt. This finding is consistent with the Theory of Planned Behavior (TPB), which suggests that without intention,

subsequent behavioral changes will not occur [15]. A previous meta-analysis determined that the TPB applies to smoking behavior, specifically that smoking intentions predict smoking behavior and that the key constructs of the TPB (attitudes, subjective norms, and perceived behavioral control) were all associated with smoking intentions [42]. Health behavior theories like the TPB are promoted for use in interventions to change addictive behaviors [43-45]. Using the TPB by incorporating additional strategies to change social norms, attitudes, and perceived behavioral control may help maintain engagement of smokers undergoing a quit attempt.

Users who smoked within 5 minutes of waking were more likely to opt out prior to program completion compared to those who reported a longer interval to their first cigarette of the day. Time to first cigarette is an indicator of the level of nicotine dependence [12], and smokers with higher levels of nicotine dependence are more likely to relapse [46,47]. The level of nicotine dependence represents a potentially important dimension for treatment tailoring; smokers with higher versus lower levels of nicotine dependence might benefit from different forms, doses, or schedules of pharmaceutical support to achieve long-term smoking cessation. For example, evaluations of pharmacotherapy for cessation show that smokers with higher levels of nicotine dependence benefit more from combination pharmacotherapy (ie, combinations of smoking cessation medications) than smokers with lower levels of nicotine dependence. Currently, nicotine replacement therapy (NRT) or other pharmaceutical cessation aids are not provided to SmokefreeTXT users. Nonetheless, there is potential opportunity to increase the use of NRT, particularly among SmokefreeTXT users with the highest levels of nicotine dependence, by tailoring the intervention to provide additional suggestions and information on the use of NRT in this group. Additionally, the use of pharmaceutical therapies for cessation is higher among former smokers who visited a physician in the year before they quit smoking [48]; thus, identifying ways to integrate clinical care, where access to pharmaceuticals is increased, and digital cessation care might also lead to more successful quitting attempts and motivation to stay engaged with a text message–based cessation program. One example of integration of clinical and digital cessation is the use of the electronic health record to identify patients who currently smoke, which then prompts providers to offer tobacco dependence counseling and medication and include a referral to the local quit line [49,50]. Within this framework, it could also be feasible for clinical practices to include referral to other digital tools like text messaging, depending on patient preference.

People who enrolled within 1 week of starting their quit attempt were more likely to opt out than those who enrolled between 1 and 2 weeks before initiating their quit attempt. One interpretation of this result is that greater exposure to preparation messages increased the odds of successfully quitting smoking. This finding is in line with clinical practice guidelines, which recommend that smokers create a quit plan and take time to prepare for their quit attempt [51]. Previous text messaging–based cessation interventions have also included time for preparation by requiring participants to set quit dates in the future [26,52,53]. However, when implemented in a real-world setting, there is no oversight about how much preparation a user receives and in our study population, almost 38% of users opted to initiate their quit attempt immediately and thus received no preparation messages. Further, requiring a specific amount of preparation may create a barrier to accessing support for a smoker who is ready to quit immediately. Since SmokefreeTXT users are allowed to set their own quit date within a 2-week period, it is unclear what the impact would be if all users received the same amount of preparation messages. Further, it is unknown how users are selecting their initial quit date in this setting. Future research should examine the motivations and preferences of users setting their own quit dates and continue to explore this critical period of the cessation attempt.

People who used the feature of resetting their quit date were less likely to opt out of SmokefreeTXT. Use of this feature was the only independent variable to show a protective effect with opting out and was associated with early and overall program retention. It is possible that use of this functionality represents smokers remaining committed to their cessation goals, especially as use of this feature is optional. Staying committed to a quit attempt is extremely important early on when the risk of smoking relapse is the highest; therefore, utilization of this feature could represent people who might have left after failing to initiate their quit attempt (ie, smoked on quit day), but instead of giving up on quitting smoking altogether and opting out, they immediately recommitted to quitting by using the reset feature. Thus, there is potential value in reframing a smoking lapse not as a failure but as part of the process of cessation that can still help smokers get to their end goal of being smoke free. Further investigation is needed to better understand this phenomenon and the best way to promote re-engagement in a cessation attempt after a slip or another setback.

This study has several limitations that need to be considered. First, given the real-world implementation of this study, participant burden needed to be minimized; therefore, only two additional baseline items were added per user to the SmokefreeTXT sign up webpage. As a result, complete information for all the independent variables of interest were not available for each user. However, as these data are all missing completely at random, future research could consider employing imputation techniques to allow for simultaneous examination of all independent variables of interest. Second, to reduce respondent burden, only single-item measures were used for data collection. However, validated items and scales were used when available. Additionally, this study focused on smoking context and motivational characteristics associated with retention and therefore did not measure all potential user characteristics of interest, such as measures of depression or stress, or other substance use behaviors like alcohol use. Fortunately, it appears that SmokefreeTXT users tolerated the additional items, as sign-up rates to the program were not reduced during the study. Therefore, a future implementation could consider adding more than two new items to the sign up webpage to expand data collection. Lastly, this study was unable to account for the potential influence of technology-related characteristics; for example, previous use of text-messaging programs and cell phone availability during working hours or

assessment of users' input on why they decided to opt out. Future qualitative studies might be particularly well suited to explore the larger contextual factors that impact a smokers' use and disuse of technology-based interventions.

This study identified several important smoking context and motivational characteristics associated with opting out of a smoking cessation text-messaging program. Although some characteristics were associated with opting out over the entire 42-day cessation program, others were only influential very early into the quit attempt. The user characteristics and programmatic features found to be associated with opting out could be addressed through program changes, such as tailoring of new messages or integration of mHealth tools with clinical care. For example, program changes to incorporate tailoring for characteristics like age, frequency of being around other smokers, and motivations to quit smoking might increase the salience of the program to users. Integration with clinical care could help support smokers with higher levels of nicotine dependence by facilitating use of NRTs and other pharmaceutical cessation aids. These modifications, in combination with promoting the use of preparation messages and helping users remain committed to quitting smoking, could increase retention and engagement, which could ultimately increase success in quitting smoking.

Acknowledgments

The Smokefree.gov Initiative is supported via National Cancer Institute contracts to ICF (contract #HHSN261201400002B and #HHSN26100007). The research presented in this paper is that of the authors and does not necessarily reflect the official position of the National Cancer Institute or the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

User characteristics by the timing of the opt-out variable.

[PDF File (Adobe PDF File), 86KB - mhealth_v7i8e13712_app1.pdf]

Multimedia Appendix 2

Adjusted survival analysis describing predictors of opting out of SmokefreeTXT.

[PDF File (Adobe PDF File), 86KB - mhealth_v7i8e13712_app2.pdf]

Multimedia Appendix 3

Characteristics associated with opting out early.

[PDF File (Adobe PDF File), 88KB - mhealth_v7i8e13712_app3.pdf]

References

1. The health consequences of smoking—50 years of progress: A Report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services; 2014.
2. Hall A, Cole-Lewis H, Bernhardt J. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
3. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
4. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
5. Ybarra M, Jiang Y, Free C, Abroms L, Whittaker R. Participant-level meta-analysis of mobile phone-based interventions for smoking cessation across different countries. Prev Med 2016 Dec;89:90-97 [FREE Full text] [doi: 10.1016/j.ypmed.2016.05.002] [Medline: 27154349]
6. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
7. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: A national text message smoking cessation intervention. Tob Control 2017 Dec;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]

8. Heminger C, Boal A, Zumer M, Abroms L. Text2Quit: An analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Dec;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
9. Bahadir A, Iliaz S, Yurt S, Ortakoylu M, Bakan N, Yazar E. Factors affecting dropout in the smoking cessation outpatient clinic. Chron Respir Dis 2016 May;13(2):155-161 [FREE Full text] [doi: 10.1177/1479972316629953] [Medline: 26846679]
10. Stolz D, Scherr A, Seiffert B, Kuster M, Meyer A, Fagerström KO, et al. Predictors of success for smoking cessation at the workplace: A longitudinal study. Respiration 2014;87(1):18-25 [FREE Full text] [doi: 10.1159/000346646] [Medline: 23594795]
11. ICF. SmokefreeTXT Cumulative Analysis, 2012-2017. ICF Report 2018.
12. Heatherton T, Kozlowski L, Frecker R, Fagerström KO. The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
13. Edelen M. The PROMIS smoking assessment toolkit--background and introduction to supplement. Nicotine Tob Res 2014 Sep;16 Suppl 3:S170-S174 [FREE Full text] [doi: 10.1093/ntr/ntu086] [Medline: 25118225]
14. Williams G, McGregor H, Sharp D, Levesque C, Kouides R, Ryan R, et al. Testing a Self-determination Theory intervention for motivating tobacco cessation: Supporting autonomy and competence in a clinical trial. Health Psychol 2006 Jan;25(1):91-101. [doi: 10.1037/0278-6133.25.1.91] [Medline: 16448302]
15. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-T]
16. Smokefree.gov. Smokefreetxt - Terms URL: https://smokefree.gov/smokefreetxt-terms [accessed 2019-02-14] [WebCite Cache ID 76BTubcsf]
17. Cepeda-Benito A, Doogan N, Redner R, Roberts M, Kurti A, Villanti A, et al. Trend differences in men and women in rural and urban U.S. settings. Prev Med 2018 Dec;117:69-75 [FREE Full text] [doi: 10.1016/j.ypmed.2018.04.008] [Medline: 29627511]
18. García-Rodríguez O, Secades-Villa R, Flórez-Salamanca L, Okuda M, Liu S, Blanco C. Probability and predictors of relapse to smoking: Results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Drug Alcohol Depend 2013 Oct 01;132(3):479-485 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.03.008] [Medline: 23570817]
19. Smith P, Bessette A, Weinberger A, Sheffer C, McKee S. Sex/gender differences in smoking cessation: A review. Prev Med 2016 Dec;92:135-140 [FREE Full text] [doi: 10.1016/j.ypmed.2016.07.013] [Medline: 27471021]
20. Huang W, Hsu H, Chang B, Chang F. Factors correlated with success rate of outpatient smoking cessation services in Taiwan. Int J Environ Res Public Health 2018 Dec 10;15(6) [FREE Full text] [doi: 10.3390/ijerph15061218] [Medline: 29890766]
21. Reynoso J, Susabda A, Cepeda-Benito A. Gender differences in smoking cessation. J Psychopathol Behav Assess 2005 Sep;27(3):227-234. [doi: 10.1007/s10862-005-0638-2]
22. Strecher V, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: Randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
23. Graham A, Jacobs M, Cohn A, Cha S, Abroms L, Papandonatos G, et al. Optimising text messaging to improve adherence to web-based smoking cessation treatment: A randomised control trial protocol. BMJ Open 2016 Mar 30;6(3):e010687 [FREE Full text] [doi: 10.1136/bmjopen-2015-010687] [Medline: 27029775]
24. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]
25. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: The initial trial of TXT-2-Quit: Randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 30;1(2):e17 [FREE Full text] [doi: 10.2196/mhealth.2522] [Medline: 25098502]
26. Abroms L, Boal A, Simmens S, Mendel J, Windsor R. A randomized trial of Text2Quit: A text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
27. Skov-Ettrup L, Ringgaard L, Dalum P, Flensborg-Madsen T, Thygesen L, Tolstrup J. Comparing tailored and untailored text messages for smoking cessation: A randomized controlled trial among adolescent and young adult smokers. Health Educ Res 2014 Apr;29(2):195-205. [doi: 10.1093/her/cyt112] [Medline: 24399268]
28. Tombor I, Shahab L, Herbec A, Neale J, Michie S, West R. Smoker identity and its potential role in young adults' smoking behavior: A meta-ethnography. Health Psychol 2015 Oct;34(10):992-1003 [FREE Full text] [doi: 10.1037/hea0000191] [Medline: 25622078]
29. Levinson A, Campo S, Gascoigne J, Jolly O, Zakharyan A, Tran Z. Smoking, but not smokers: Identity among college students who smoke cigarettes. Nicotine Tob Res 2007 Aug;9(8):845-852. [doi: 10.1080/14622200701484987] [Medline: 17654297]
30. Berg C, Lust K, Sanem J, Kirch M, Rudie M, Ehlinger E, et al. Smoker self-identification versus recent smoking among college students. Am J Prev Med 2009 Apr;36(4):333-336. [doi: 10.1016/j.amepre.2008.11.010] [Medline: 19201148]

31. Ridner S, Walker K, Hart J, Myers J. Smoking identities and behavior: Evidence of discrepancies, issues for measurement and intervention. West J Nurs Res 2010 Jun;32(4):434-446. [doi: 10.1177/0193945909354904] [Medline: 20685903]
32. Choi Y, Choi S, Rifon N. "I smoke but I am not a smoker": Phantom smokers and the discrepancy between self-identity and behavior. J Am Coll Health 2010;59(2):117-125. [doi: 10.1080/07448481.2010.483704] [Medline: 20864438]
33. Hoek J, Maubach N, Stevenson R, Gendall P, Edwards R. Social smokers' management of conflicted identities. Tob Control 2013 Jul;22(4):261-265. [doi: 10.1136/tobaccocontrol-2011-050176] [Medline: 22351792]
34. Wellman R, O'Loughlin E, Dugas E, Montreuil A, Dutczak H, O'Loughlin J. Reasons for quitting smoking in young adult cigarette smokers. Addict Behav 2018 Feb;77:28-33. [doi: 10.1016/j.addbeh.2017.09.010] [Medline: 28946012]
35. Villanti A, Bover Manderski M, Gundersen D, Steinberg M, Delnevo C. Reasons to quit and barriers to quitting smoking in US young adults. Fam Pract 2016 Apr;33(2):133-139 [FREE Full text] [doi: 10.1093/fampra/cmv103] [Medline: 26733658]
36. Guillory J, Lisha N, Lee Y, Ling P. Phantom smoking among young adult bar patrons. Tob Control 2017 Dec;26(2):153-157 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052821] [Medline: 27048205]
37. Pew Research Center: Internet & Technology. 2018. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-05-20] [WebCite Cache ID 78b0Y98Iq]
38. Nørregaard J, Tønnesen P, Petersen L. Predictors and reasons for relapse in smoking cessation with nicotine and placebo patches. Prev Med 1993 Mar;22(2):261-271. [doi: 10.1006/pmed.1993.1021] [Medline: 8483863]
39. Shiffman S, Paty J, Gnys M, Kassel J, Hickcox M. First lapses to smoking: Within-subjects analysis of real-time reports. J Consult Clin Psychol 1996 Apr;64(2):366-379. [Medline: 8871421]
40. Motschman C, Germeroth L, Tiffany S. Momentary changes in craving predict smoking lapse behavior: A laboratory study. Psychopharmacology (Berl) 2018 Dec;235(7):2001-2012. [doi: 10.1007/s00213-018-4898-4] [Medline: 29704217]
41. Sayette M, Dimoff J. In search of anticipatory cigarette cravings: The impact of perceived smoking opportunity and motivation to seek treatment. Psychol Addict Behav 2016;30:277-286. [Medline: 27099960]
42. Topa G, Moriano J. Theory of planned behavior and smoking: Meta-analysis and SEM model. Subst Abuse Rehabil 2010;1:23-33 [FREE Full text] [doi: 10.2147/SAR.S15168] [Medline: 24474850]
43. Webb T, Sniehotta F, Michie S. Using theories of behaviour change to inform interventions for addictive behaviours. Addiction 2010 Nov;105(11):1879-1892. [doi: 10.1111/j.1360-0443.2010.03028.x] [Medline: 20670346]
44. Lee C, Middlestadt S, Seo D, Lin H, Macy J, Park S. Incorporating environmental variables as precursor background variables of the Theory of Planned Behavior to predict quitting-related intentions: A comparative study between adult and young adult smokers. Arch Public Health 2018;76:66 [FREE Full text] [doi: 10.1186/s13690-018-0311-3] [Medline: 30410755]
45. Hagger M, Leaver E, Esser K, Leung C, Te Pas N, Keatley D, et al. Cue-induced smoking urges deplete cigarette smokers' self-control resources. Ann Behav Med 2013 Dec;46(3):394-400. [doi: 10.1007/s12160-013-9520-8] [Medline: 23720186]
46. Ockene J, Emmons K, Mermelstein R, Perkins K, Bonollo D, Voorhees C, et al. Relapse and maintenance issues for smoking cessation. Health Psychol 2000 Jan;19(1S):17-31. [Medline: 10709945]
47. Transdisciplinary Tobacco Use Research Center (TTURC) Tobacco Dependence, Baker T, Piper M, McCarthy D, Bolt D, Smith S, et al. Time to first cigarette in the morning as an index of ability to quit smoking: Implications for nicotine dependence. Nicotine Tob Res 2007 Nov;9 Suppl 4:S555-S570 [FREE Full text] [doi: 10.1080/14622200701673480] [Medline: 18067032]
48. Soulakova J, Crockett L. Unassisted quitting and smoking cessation methods used in the United States: Analyses of 2010-2011 Tobacco Use Supplement to the Current Population Survey Data. Nicotine Tob Res 2017 Dec 13;20(1):30-39 [FREE Full text] [doi: 10.1093/ntr/ntw273] [Medline: 27798084]
49. Lindholm C, Adsit R, Bain P, Reber P, Brein T, Redmond L, et al. A demonstration project for using the electronic health record to identify and treat tobacco users. WMJ 2010 Dec;109(6):335-340 [FREE Full text] [Medline: 21287886]
50. Adsit R, Fox B, Tsiolis T, Ogland C, Simerson M, Vind L, et al. Using the electronic health record to connect primary care patients to evidence-based telephonic tobacco quitline services: A closed-loop demonstration project. Transl Behav Med 2014 Sep;4(3):324-332 [FREE Full text] [doi: 10.1007/s13142-014-0259-y] [Medline: 25264471]
51. 2008 PHS Guideline Update Panel. Treating tobacco use and dependence: 2008 update U.S. Public Health Service Clinical Practice Guideline executive summary. Respir Care 2008 Sep;53(9):1217-1222 [FREE Full text] [Medline: 18807274]
52. 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 02;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
53. Ybarra M, Holtrop J, Prescott T, Rahbar M, Strong D. Pilot RCT results of Stop My Smoking USA: A text messaging-based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399 [FREE Full text] [doi: 10.1093/ntr/nts339] [Medline: 23348969]

Abbreviations

mHealth: mobile health
NRT: nicotine replacement therapy

TPB: Theory of Planned Behavior

Edited by G Eysenbach; submitted 14.02.19; peer-reviewed by S Parackal, P Delir Haghighi, D Christofferson; comments to author 28.03.19; revised version received 23.05.19; accepted 27.05.19; published 01.08.19.

Please cite as:
Wiseman KP, Coa KI, Prutzman YM
Predictors of Retention in an Adult Text Messaging Smoking Cessation Intervention Program: Cohort Study
JMIR Mhealth Uhealth 2019;7(8):e13712
URL: https://mhealth.jmir.org/2019/8/e13712/
doi:10.2196/13712
PMID:31373278

Original Paper

German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS)

Johannes Knitza[1,2], MD; Koray Tascilar[1], MD; Eva-Maria Messner[3], MA; Marco Meyer[2,4], MD; Diana Vossen[2,5], MD; Almut Pulla[2,5], MD; Philipp Bosch[6], MD; Julia Kittler[1,2], MD; Arnd Kleyer[1,2], MD; Philipp Sewerin[2,7], MD; Johanna Mucke[2,7], MD; Isabell Haase[2,7], MD; David Simon[1,2], MD; Martin Krusche[2,8], MD

[1]Department of Internal Medicine 3 – Rheumatology and Immunology, Friedrich-Alexander University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany

[2]Working Group Young Rheumatology, German Society for Rheumatology, Berlin, Germany

[3]Department for Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

[4]Rheumatologie, Klinische Immunologie, Nephrologie, Asklepios Klinik Altona, Hamburg, Germany

[5]Rheinisches Rheumazentrum Meerbusch, St Elisabeth Hospital, Meerbusch, Germany

[6]Department of Rheumatology and Immunology, Medical University Graz, Graz, Austria

[7]Department of Rheumatology and Hiller Research Unit Rheumatology, Heinrich Heine University, Düsseldorf, Germany

[8]Department of Rheumatology and Clinical Immunology, Charité Universitätsmedizin, Berlin, Germany

Corresponding Author:
Johannes Knitza, MD
Department of Internal Medicine 3 – Rheumatology and Immunology
Friedrich-Alexander University Erlangen-Nürnberg
University Hospital Erlangen
Ulmenweg 18
Erlangen,
Germany
Phone: 49 0913 1853 4742
Fax: 49 0913 1853 6448
Email: johannes.knitza@uk-erlangen.de

Abstract

Background: Chronic rheumatic diseases need long-term treatment and professional supervision. Mobile apps promise to improve the lives of patients and physicians. In routine practice, however, rheumatology apps are largely unknown and little is known about their quality and safety.

Objective: The aim of this study was to provide an overview of mobile rheumatology apps currently available in German app stores, evaluate app quality using the Mobile Application Rating Scale (MARS), and compile brief, ready-to-use descriptions for patients and rheumatologists.

Methods: The German App Store and Google Play store were systematically searched to identify German rheumatology mobile apps for patient and physician use. MARS was used to independently assess app quality by 8 physicians, 4 using Android and 4 using iOS smartphones. Apps were randomly assigned so that 4 apps were rated by all raters and the remaining apps were rated by two Android and two iOS users. Furthermore, brief app descriptions including app developers, app categories, and features were compiled to inform potential users and developers.

Results: In total, 128 and 63 apps were identified in the German Google Play and App Store, respectively. After removing duplicates and only including apps that were available in both stores, 28 apps remained. Sixteen apps met the inclusion criteria, which were (1) German language, (2) availability in both app stores, (3) targeting patients or physicians as users, and (4) clearly including rheumatology or rheumatic diseases as subject matter. Exclusion criteria were (1) congress apps and (2) company apps with advertisements. Nine apps addressed patients and 7 apps addressed physicians. No clinical studies to support the effectiveness and safety of apps could be found. Pharmaceutical companies were the main developers of two apps. Rheuma Auszeit was the only app mainly developed by a patient organization. This app had the highest overall MARS score (4.19/5). Three out of 9 patient apps featured validated questionnaires. The median overall MARS score was 3.85/5, ranging from 2.81/5 to 4.19/5. One patient-targeted and one physician-targeted app had MARS scores >4/5. No significant rater gender or platform (iOS/Android)

differences could be observed. The overall correlation between app store ratings and MARS scores was low and inconsistent between platforms.

Conclusions: To our knowledge, this is the first study that systematically identified and evaluated mobile apps in rheumatology for patients and physicians available in German app stores. We found a lack of supporting clinical studies, use of validated questionnaires, and involvement of academic developers. Overall app quality was heterogeneous. To create high-quality apps, closer cooperation led by patients and physicians is vital.

(JMIR Mhealth Uhealth 2019;7(8):e14991) doi:10.2196/14991

KEYWORDS

mobile apps; eHealth; rheumatology; mHealth; Mobile Application Rating Scale

Introduction

There is great potential in using eHealth tools, especially in chronic rheumatic diseases [1]. In the anticipated reality of Rheumatology 4.0 computer-aided diagnostic systems allowing precise and quick diagnosis [2], mobile apps and other eHealth tools could improve the positions of all stakeholders, including patients, physicians, health insurance companies, and the pharmaceutical industry.

The use of diagnostic decision support systems could shorten the time to correct diagnosis, even for rare diseases [3]. Once a correct diagnosis is established, patients and physicians need to maintain disease control, which necessitates continuous monitoring of treatment adherence, accurate symptom tracking, and surveillance of adverse treatment effects. eHealth is promising to increase the quantity, quality, and availability of medical data, thus allowing a more precise and personalized treatment. A recent study showed that remote monitoring of disease activity using physical activity trackers precisely detects flareups in patients with rheumatoid arthritis [4]. This is a good example in which an accurate clinical assessment is accomplished using an eHealth tool without necessitating a direct patient-physician encounter. Such tools could drastically increase the efficiency of health care delivery.

The development of apps is becoming easier and less expensive thanks to the lack of restrictions on interventions in app stores. These low market barriers attract various businesses that seek to seize the opportunity of entering the profitable health care market [5]. This leads to considerable heterogeneity regarding security and quality in general.

Quality indicators for health care–related apps beyond the app store star ratings, comments, and number of downloads are largely unavailable. Trust marks and certification labels (like CE marking indicating conformity with health, safety, and environmental protection standards for products sold within the European Economic Area) for apps are rarely found [6], making quality assessment of an app a challenge for the end user. A number of tools have been proposed to this end [7,8]. Among the relatively established tools to rate app quality is the validated Mobile Application Rating Scale (MARS) [9]. Since its publication in 2015, it has been used to rate various medical mobile apps [10,11]. The MARS score is based on a 5-point Likert scale in four sections with multiple items: engagement (5 items), functionality (4 items), aesthetics (3 items), and information quality (7 items). In addition, there is a subjective section consisting of 4 items.

A New Zealand study recently reported the results of a MARS assessment evaluating patient apps for rheumatoid arthritis and found a lack of high-quality apps [10]. Such systematic quality assessments are scarce and represent an unmet need. A recent survey conducted by the Working Group Young Rheumatology of the German Society for Rheumatology (Arbeitsgemeinschaft Junge Rheumatologie, or rheumadocs) showed that medical app use among German rheumatologists increased by 12% during two years, yet rheumatologists were aware of only two recommendable apps specific to rheumatology (RheumaHelper, RheumaLive) [5].

To our knowledge, no systematic quality assessment of rheumatology apps available in German app stores has yet been carried out. Therefore, the aim of this study was to identify and evaluate rheumatology-specific German mobile apps targeting patients or physicians.

Methods

Selection of Mobile Apps

An extensive German App Store and Google Play search was performed from May 1-31, 2018. This search included the following search terms: “Rheuma” OR “Rheumatologie” OR “Arthritis” OR “Psoriasisarthritis” OR “PsA” OR “PsoA” OR “Rheumatoide Arthritis” OR “RA” OR “Morbus Bechterew” OR “Spondylarthritis” OR “Spondylitis ankylosans” OR “Spondylopathie” OR “axSpa” OR “Spondylarthropathie” OR “Spa” OR “Spondylitis ankylosans” OR “ankylosierende Spondylarthritis” OR “Kollagenose” OR “SLE” OR “Systemischer Lupus Erythematodes” OR “Sklerodermie” OR “Lupus” OR “Sjögren” OR “APS” OR “Antiphospholipidsyndrom” OR “systemische Sklerose” OR “SSc” OR “Polymyositis” OR “Dermatomyositis” OR “RZA” OR “Riesenzellarteriitis” OR “Riesenzellarteritis” OR “EGPA” OR “GPA” OR “eosinophile Granulomatose mit Polyangiitis” OR “Granulomatose mit Polyangiitis” OR “PAN” OR “panarteriitis nodosa” OR “polyarteriitis nodosa” OR “mikroskopische Polyangiitis” OR “Morbus Behcet” OR ”Takayasu Arteriitis” OR “Kawasaki Syndrom” OR “Arteriitis temporalis” OR “PMR” OR “Polymyalgie” OR “Polymyalgia rheumatica” OR “reaktive Arthritis” OR “enteropathische Arthritis” OR “Vaskulitis” OR “FMF” OR “familiäres Mittelmeerfieber” OR “Autoinflammation” OR “AOSD.”

All 8 raters (Multimedia Appendix 1) were physicians currently completing their rheumatology fellowships. Four physicians were using Android phones and four physicians were using iPhones to individually search for the terms in the associated app stores. Raters stated no conflict of interest regarding the industry surrounding the apps being rated. Searches were performed from May 1-31, 2018.

App inclusion criteria were (1) German language, (2) availability in both app stores, (3) targeting patients or physicians as users, and (4) clearly including rheumatology or rheumatic diseases as subject matter. Exclusion criteria were (1) congress apps and (2) company apps with advertisements.

App Evaluation

As recommended by the MARS developers, all raters viewed the training video by Stoyanov et al [9], and each app was tested for at least 10 minutes. The raters agreed on the relevance of all MARS items to this project. Before rating their assigned apps, all raters evaluated two apps selected for training purposes (previously excluded from the analysis) and discussed their results to ensure a similar understanding of the MARS items and process.

Four of the apps were rated by all raters and the remaining apps were randomly assigned to raters by creating a stratified randomization list using a virtual urn method without replacement, such that each app would be rated by two Android and two iPhone users. Apps were downloaded and rated from July 1-31, 2018. Furthermore, information was collected regarding target group, target disease, content, developer, availability of privacy policy statement, medical product status, and current app store rating. Availability of scientific studies was checked via Google, Google Scholar, PubMed, the developer website, and the app stores.

Statistical Analysis

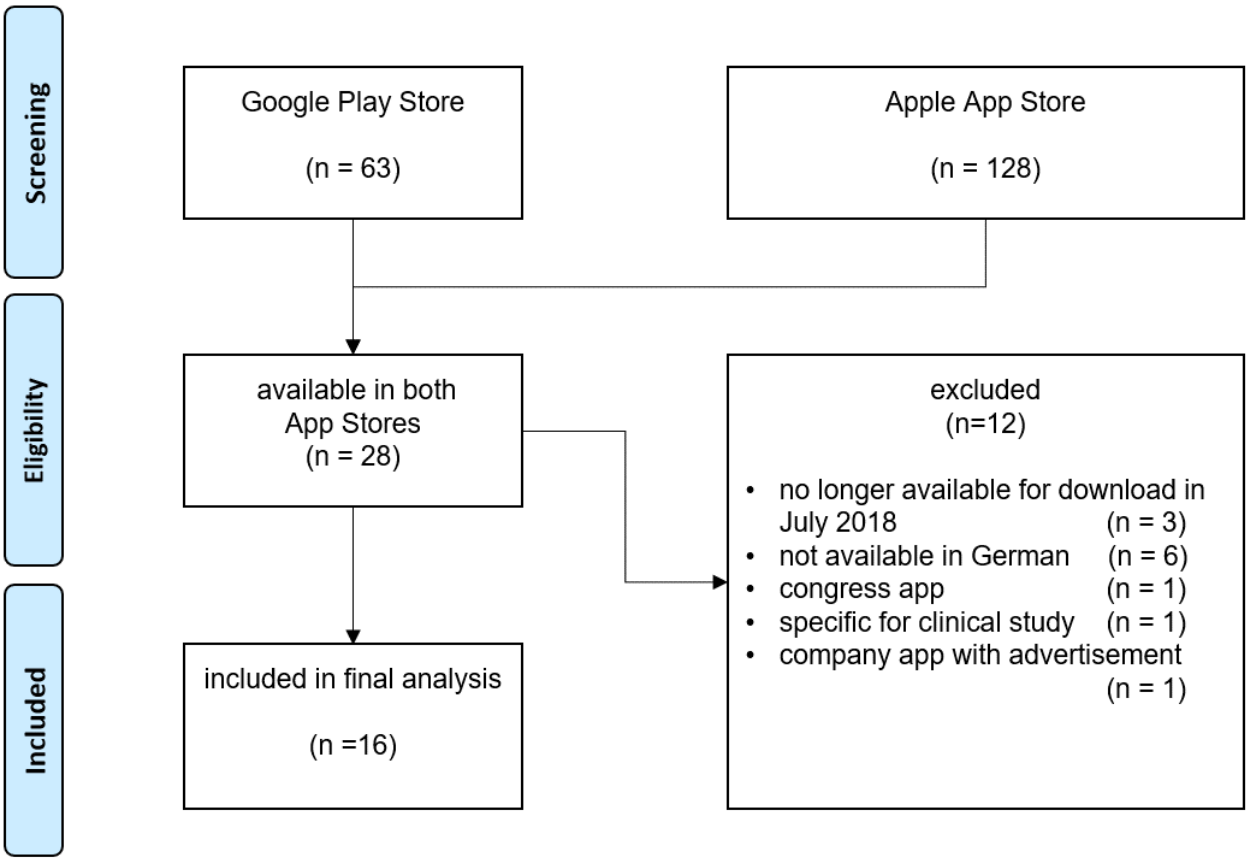
MARS section scores were calculated by taking the arithmetic mean of each item score in the section, while the overall score was the arithmetic mean of the section scores (excluding subjective quality). Overall scores and section scores were summarized as median and range for each app, and apps were ranked based on the median overall MARS score. We analyzed item score deviations by section and rater using a random intercept–only mixed-effects linear regression model including the individual item scores as the dependent variable, a random effects term for the rater, and nested random effects terms for the MARS section and app. Using random intercepts from this model, we estimated how the item scores in each section in each app deviated from the overall mean item score to rank and plot the importance of the sections within each app. Similarly, we plotted the random effect intercepts and respective 95% confidence intervals for raters to rank the raters by their deviation from the overall mean item score as a measure of rater bias. We analyzed the effect of rater gender and operating system on ratings by adding respective fixed effect terms to the model and reported their coefficients and 95% confidence intervals. Random intercept and fixed effect term confidence intervals spanning both sides of 0 were considered insignificant. We constructed scatter plots of MARS scores for each app and platform against their respective store ratings and calculated Pearson correlation coefficients both across platforms and separately. Finally, we analyzed interrater agreement at item, section, and overall score levels for raters from a rater sample, namely ICC2k (two-way random, average measures, absolute agreement) [12]. All data analysis was performed using the open source R software v3.5.3 (The R Foundation).

Results

Selection of Mobile Apps

In total 128 and 63 apps were identified in the German App Store and Google Play, respectively. After removing duplicates and only including apps that were available in both stores, 28 apps remained. Three previously included apps were no longer available for download in July 2018 and were excluded; 9 apps were removed—6 were not available in German, one was a congress app, one a specific app for a clinical study, and one an ergo therapy advertisement app—so there remained 16 final apps for analysis (Figure 1). During the analysis, Psoriapp was no longer available in Google Play and could only be rated by iOS raters. Android rater 2 downloaded Rheumatologie Visuell but the log-in repeatedly failed. The same rater was unable to successfully download Rheuma Edu although it was available in Google Play. iOS rater 2 had the same problem with the Rheumatologie Visuell app.

Figure 1. App selection process flowchart.



Characteristics of Mobile Apps

Tables 1 and 2 display the characteristics of the analyzed apps. Nine apps were designed for patient use, and 7 for physician use. The following rheumatologic diseases were targeted: rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (SpA), juvenile idiopathic arthritis (JIA), systemic lupus erythematosus (SLE), vasculitis, and giant cell arteritis. Thirteen apps were rheumatology specific, and 3 apps were nonspecific. All physician-targeted apps focused on education. Most of these apps were text and graphic based, focusing on guidelines. Other physician apps incorporated videos (Rheuma Edu), audio files (Meditorium), and case images (Rheumatologie Visuell). Three apps consisted of a score calculator. Eight out of 9 patient apps had a diary function of some sort. Rheuma Auszeit, the only patient app without a diary function, provided video and audio instructions for mental and physical exercises. Only 3 out of 8 diary patient apps consisted of validated disease activity questionnaires. Most apps provided a reminder function. Two out of 9 patient apps provided a service to exchange experiences via private or group messages.

Only one app, Rheuma Auszeit, was developed mainly by a patient organization; 2 apps were mainly developed by pharmaceutical companies. Five other apps were financially supported by pharmaceutical companies. All patient apps were free of charge, but 2 of the physician-targeted apps required in-app purchases to function completely. According to the associated website, the MyTherapy app has been used for an adherence study with type 2 diabetes patients; however, no details for this study were stated and the study could not be identified using Google Scholar or PubMed. Privacy policy statements were available for all apps except the ASAS App. Three patient apps were classified as medical products, all constructed by the same developer (STAR Healthcare Management GmbH).

Table 1. Target group, target disease, and developer of included rheumatology apps.

App	Target group	Target disease	Developer	Category
Rheuma Auszeit	Patient	RA[a]	Deutsche Rheuma-Liga Bundesverband eV	Education
Meditorium	Physician	Nonspecific	SchiLu Media UG	Education
RheumaGuide	Physician	RA[a], PsA[b], SpA[c]	MedMedia Verlag und Mediaservice GmbH	Education
ASAS App	Physician	SpA[c]	Assessment of SpondyloArthritis International Society	Education, calculator
RheumaLive	Patient	RA[a]	STAR Healthcare Management GmbH	Diary, reminder
Pain Companion	Patient	Nonspecific	Sanovation AG	Diary, exchange
MyTherapy	Patient	Nonspecific	Smartpatient GmbH	Diary, reminder
Psoriapp	Patient	PsA[b]	Novartis Pharma GmbH	Diary, reminder
Rheumatologie Visuell	Physician	Rheumatic	Georg Thieme Verlag KG	Education
PsALive	Patient	PsA[b]	STAR Healthcare Management GmbH	Diary, reminder
AxSpaLive	Patient	SpA[c]	STAR Healthcare Management GmbH	Diary, reminder
Lupuslog	Patient	SLE[d]	GlaxoSmithKline PLC	Diary, reminder
Rheuma Edu	Physician	Rheumatic	Pomcanys Marketing AG	Education
ANCA–Assoziierte Vaskulitiden	Physician	ANCA[e]–associated vasculitis	Börm Bruckmeier Verlag GmbH	Education
RheumaBuddy	Patient	RA[a], JIA[f]	DAMAN P/S	Diary, exchange
Rheumatologie App	Physician	RA[a], vasculitis	Börm Bruckmeier Verlag GmbH	Education

[a]RA: rheumatoid arthritis.

[b]PsA: psoriatic arthritis.

[c]SpA: ankylosing spondylitis.

[d]SLE: systemic lupus erythematosus.

[e]ANCA: antineutrophil cytoplasmic antibody.

[f]JIA: juvenile idiopathic arthritis.

Table 2. Characteristics of included rheumatology apps.

App	Technical aspects/content	Studies available	Privacy policy available	Medical product	Price
Rheuma Auszeit	Videos and audio files	No	Yes	No	Free
Meditorium	Audio files	No	Yes	No	2.99 € to 47.99 € (US $3 to $54)
RheumaGuide	Diagnostic and therapeutic guidelines, score calculator	No	Yes	No	Free
ASAS App	Diagnostic and therapeutic guidelines, score calculator	No	No	No	Free
RheumaLive	Diary[a], medication reminder, export function	No	Yes	Yes	Free
Pain Companion	Diary, group discussion, private messages, export function	No	Yes	No	Free
MyTherapy	Tracking, medication/task reminder	No[b]	Yes	No	Free
Psoriapp	Diary, medication reminder, export function	No	Yes	No	Free
Rheumatologie Visuell	Rheumatology images	No	Yes	No	Free
AxSpaLive	Diary[a], medication reminder, export function	No	Yes	Yes	Free
PsALive	Diary[a], medication reminder, export function	No	Yes	Yes	Free
Lupuslog	Diary, reminder, pictures, export function, local weather	No	Yes	No	Free
Rheuma Edu	Videos	No	Yes	No	6.49 € (US $7)
ANCA–Assoziierte Vaskulitiden	Diagnostic and therapeutic guidelines	No	Yes	No	Free
RheumaBuddy	Diary, group discussion, private messages, reminder, export function	No	Yes	No	Free
Rheumatologie App	Diagnostic and therapeutic guidelines, score calculator	No	Yes	No	Free

[a]This allowed tracking of validated rheumatology-specific questionnaires.

[b]Developer website states clinical study, yet no details could be identified using Google Scholar or Pubmed.

App Ratings

The overall MARS scores ranged from 2.81 to 4.19. The apps were ranked by median overall score. The individual MARS score ratings by each rater and their range are presented in Figure 2. The individual MARS section scores by each rater and their ranges are displayed in Multimedia Appendix 2.

Random intercepts and 95% confidence intervals from the mixed-effects linear regression analysis are presented in Figure 3, summarizing the mean deviation of item scores and their 95% confidence intervals by section and their ranking within each app. This figure shows that subjective quality was the section in which item scores were most often significantly lower compared to the overall mean (Pain Companion, Psoriapp, RheumaBuddy, and Rheumatologie App). Information was the section in which item scores were most often significantly higher compared to the overall mean (Rheuma Auszeit, ASAS App, and Rheumatologie Visuell). For the aesthetics, functionality, and engagement sections there was no app with significantly lower item score deviations.

Rater agreement on overall MARS score at app level was poor and imprecise (ICC2k 0.53, 95% CI 0.08 to 0.81) whereas the interrater agreement for section scores (ICC2k 0.82, 95% CI 0.76 to 0.88) and individual item scores (ICC2k 0.84, 95% CI 0.81 to 0.86) were good. Random intercepts for observers from the mixed-effects model are presented in Figure 4.

The point estimates and confidence intervals show that 3 of the 4 iOS raters were significantly biased with respect to mean item scores either in the positive or negative direction, whereas the random intercepts for the Android raters were similar. However, adding the operating system as a fixed effect in the regression model did not seem to be associated with an overall significant difference in item scores (β=–0.10, 95% CI –0.44 to 0.24; *P*=.57 for iOS, compared to Android). Finally, the mixed-effects model with rater gender as a fixed effect also shows that the adjusted difference between item scores between male and female raters was small and imprecise (β=0.08, 95% CI –0.27 to 0.44; *P*=.64 for male gender) and does not suggest a gender effect on item scores.

MARS and app store ratings, including the range and number of ratings, are shown in Multimedia Appendix 3. App store ratings were retrieved on April 21, 2019. At the time, Psoriapp was not available in both app stores. For all apps, Google Play had more ratings than App Store. MyTherapy had by far the most ratings (24,408).

Figure 2. Mobile Application Rating Scale overall ratings of included apps.

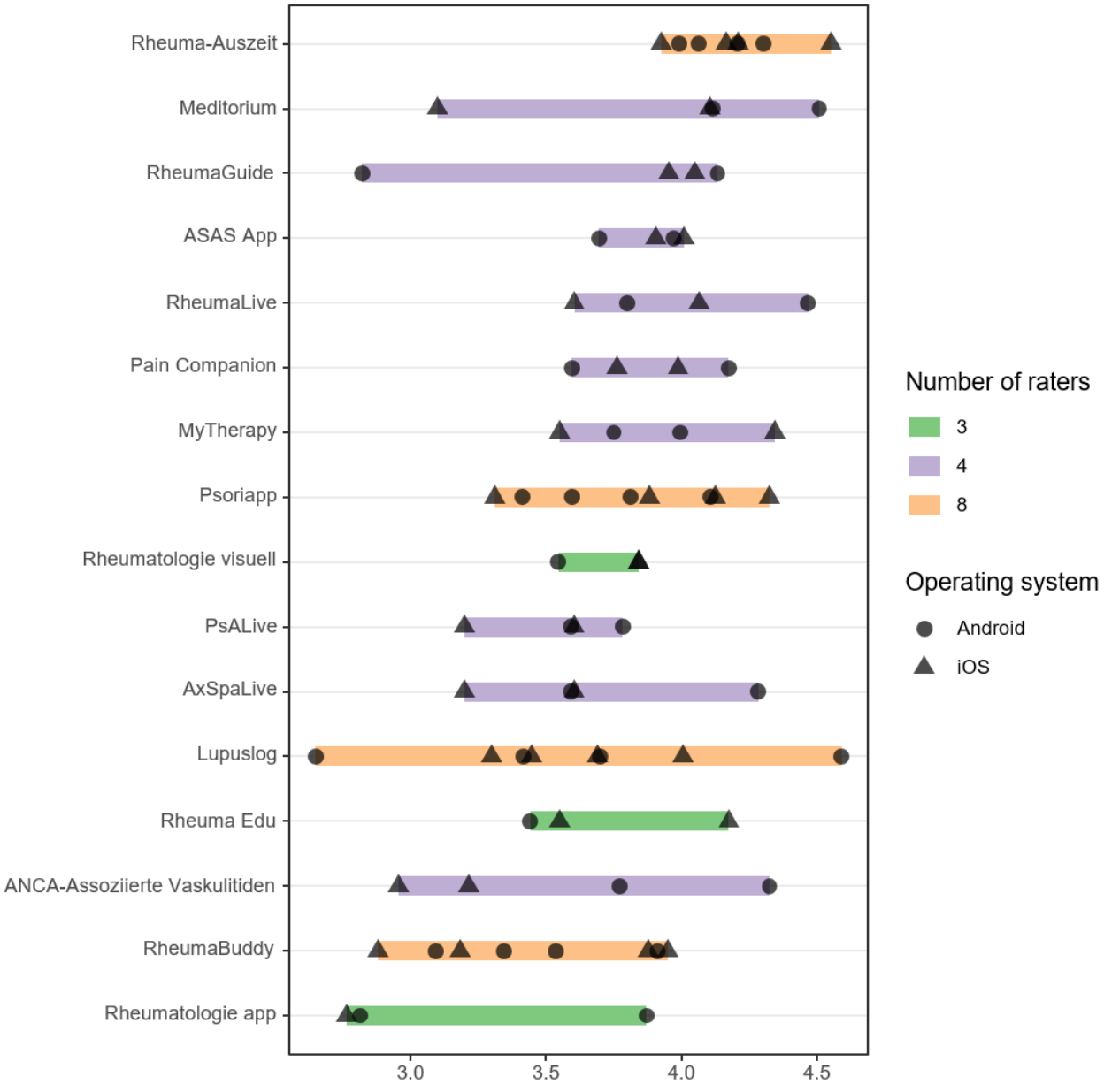


Figure 3. Mobile Application Rating Scale section item scores by section and app.

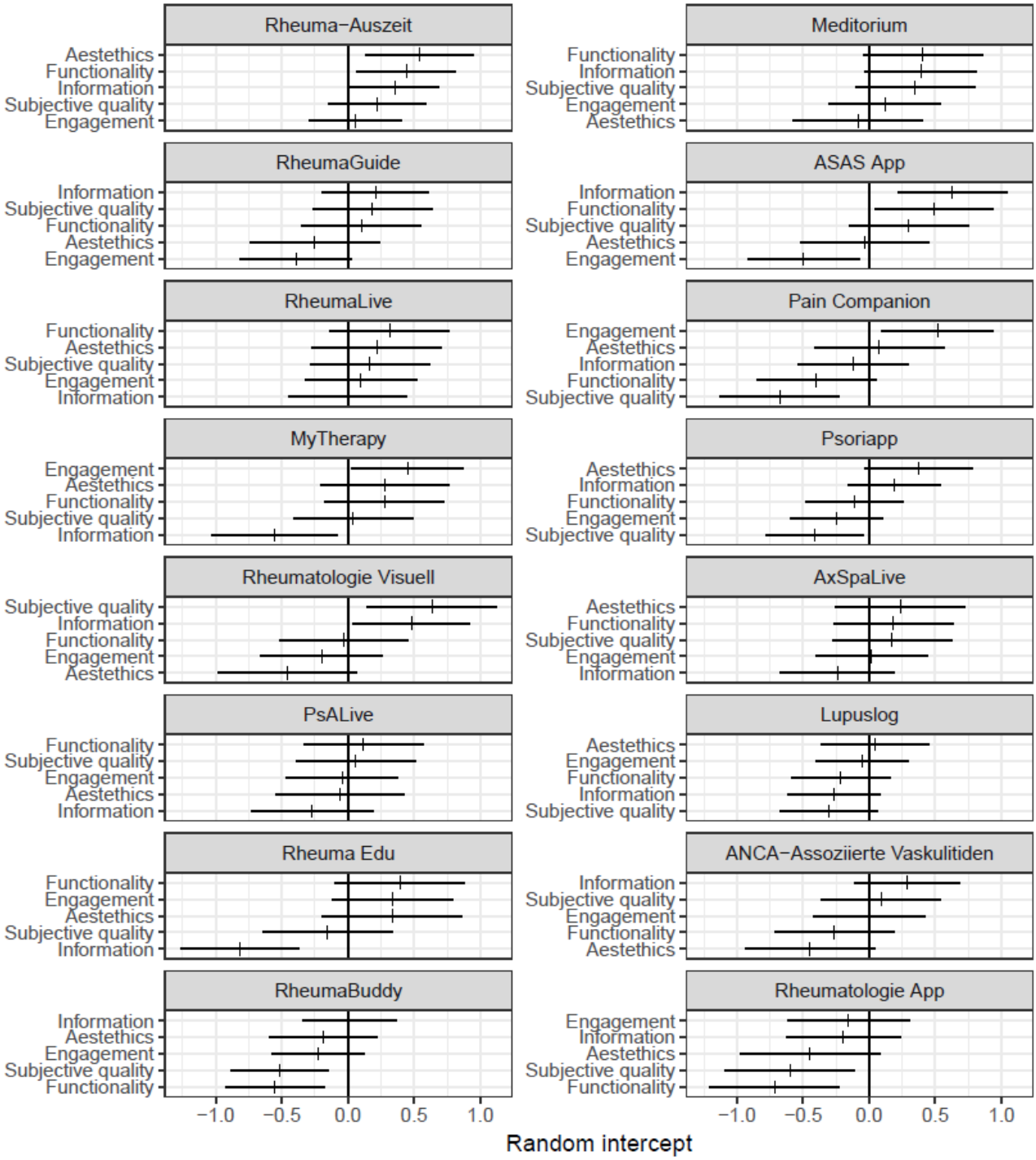
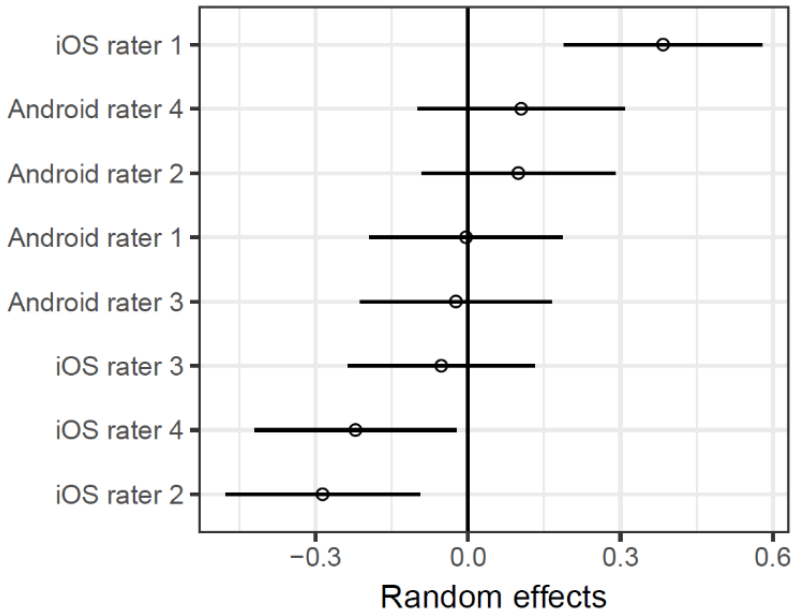


Figure 4. Rater deviations in item scores.



Correlation analysis between overall MARS scores of the apps and their respective store ratings was limited by the availability of store ratings. We did not find a significant correlation between MARS scores and store ratings, whether overall or grouped by operating system.

Discussion

Principal Findings

This analysis of German rheumatology-specific apps showed that most apps were patient focused (9 out of 16), and only a minority of rheumatic diseases were specifically targeted. For several rheumatic diseases, such as systemic sclerosis, Sjogren syndrome, Behcet disease, and familial Mediterranean fever, there were no apps available. The three apps with the highest overall MARS score included videos, audio files, and images. The inclusion of multimedia content therefore seems to be advisable.

The MARS rating itself is quite subjective, shown by the great interrater differences despite the tutorial video and test phase with discussion of the results. The poor rater agreement is likely also due to a low number of apps and a high number of raters. This result is in line with other app rating studies including more than two raters [13]. We chose not to discuss and adapt conflicting results like others [11,14] because we believe this results in a falsification of data. Studies using MARS and including only two raters often showed good interrater reliability [10]. We used MARS because it is one of the most established app rating tools; however, several tools with different weaknesses and strengths exist [7]. Interestingly, one developer created patient apps that were all certified as CE-labeled medical products for three common rheumatic diseases (RA, PsA, SpA). No other apps were found to be CE certified. However, being CE certified did not guarantee a top ranking. In order to harmonize research and increase trust and transparency, an international task force is needed to create guidelines and accepted quality criteria. These guidelines are desperately needed, as easily available quality indicators such as app store ratings only poorly reflect their true quality.

In order to increase acceptance and use among patients and health care professionals, clinical studies are also urgently needed. Only one app referred to a clinical study; however, these results could not be clearly identified using Google Scholar or PubMed. We believe that it is necessary to shift the current developer status from commercial developers toward universities and independent research institutions including patients as well as physicians. The importance of including patients is highlighted by the fact that the only app mainly developed by a patient organization had the highest MARS score (4.19/5). These findings illustrate current unmet needs hindering the use of eHealth tools despite their great potential.

Patient self-assessments via smartphone strongly correlate with rheumatologist assessments [15] and could cost effectively enhance current tight control strategies. An official, highly customizable app developed by a trusted and independent organization based on a common minimal data set would allow the creation of a holistic repository. This app would ensure maximally efficient use of resources. Due to its large user base, it would provide a powerful passive dataset for research. A role model could be the Swiss Web app, mySCQM [16], developed by Swiss Clinical Quality Management in Rheumatic Diseases. This app allows patients to enter data in between visits and share these data with their doctor and the national Swiss registry. To increase acceptance of such an app among patients, it seems advisable to also allow safe communication and file exchange between patients and their physicians. Further research to identify key components [17,18] and stakeholder preferences [5,19-21] is needed.

Limitations

This study has some limitations. First, we only looked at app stores, and no systematic literature search was performed. For future projects, it would be time saving to use an automated process to identify and filter apps as proposed by Albrecht et

al [22]; however, this process is restricted to native mobile apps. Web-based apps that are not featured in app stores are therefore not included. Due to the growing popularity of Web-based apps [23], we believe it is crucial to include these apps in future research projects.

Due to the app store approach using a limited amount of German search terms, useful rheumatology apps may have been overlooked. To facilitate app evaluation, app randomization, and data analysis, we only included apps that were available in both app stores. Another limitation is the fact that only physicians performed the app rating, although most of the apps were created for patient use. As already suggested by others [13,24], future research should include patients.

Due to lack of expertise and resources, we only checked the data security very briefly in terms of presence of a data policy statement, password protection, and log-in requirement. A professional in-depth security check should be applied to identify any risks, as mobile apps often do not follow data protection laws [25] and could be potentially harmful for the end user [26]. Finally, it should be mentioned that due to the rapid speed of mobile app development, this review might already be out of date once published. A main limitation of the MARS score itself is that to our knowledge there is no clear definition of a high-quality app, and its meaning often varies [11,27].

Comparison With Prior Work

A major strength of this study lies in its ability to guide recommendations of apps by rheumatologists for their patients. To our knowledge, no review and analysis of mobile apps in rheumatology available in German app stores has been carried out yet. In contrast to other studies [13,28] using MARS, we identified apps targeting patients as well as physicians. Furthermore, apps were tested on iOS and Android platforms to identify usability differences. To our knowledge, no previous study using MARS had as many raters as this study did. We tried to include many raters to better represent different subjective perspectives and pick up any interrater rating weaknesses of the MARS.

The low number of recommendable rheumatology-specific apps found in the previous survey [5] can now largely be explained by the lack of German rheumatology-specific apps in general (16 identified apps in total) and their heterogeneous quality. However, in contrast to a previous rheumatology app review [10], we identified one patient and one physician app with promising overall MARS scores (>4/5).

This study exposes the lack of reliable studies for mobile apps in general [29] and specifically in rheumatology [10]. Similarly, in this work we observed a wide range of MARS scores reflecting heterogeneous quality. Grainger et al [10] reported that 6 out of 11 patient-targeted apps allowed data sharing. In our analysis, this was the case for 8 out of 9 patient-targeted apps.

The lack of academic app developers reported by Salazar et al [11] is supported by our work. In accordance with prior publications [11,27], there was no strong correlation of app store ratings and MARS ratings. App store ratings therefore seem to be a poor quality indicator. In our analysis, 89% (8/9) of the patient apps had a symptom tracking function. In a previous publication [17] focusing on rheumatoid arthritis, this was only the case for 50% of the apps.

Noticeably, Rheuma Auszeit, the only app mainly developed by a patient organization, had the best MARS score and lowest rater standard deviation. This highlights the importance and success of including patients in the app development process, as stressed by Grainger et al [10]. Interestingly, this was the only patient app not containing a diary function. The name of the app is translated as rheuma timeout, implying that the goal of this app is exactly the opposite of tracking pain. This could be a main cause of adherence problems with patient apps, as patients are constantly reminded of their disease and limitations.

Based on our study findings, we established 10 recommendations (Multimedia Appendix 4) that might direct developers to create better apps that maximize patient and physician satisfaction.

Conclusion

To our knowledge, this is the first study to systematically identify and evaluate mobile apps in rheumatology for patients and physicians available in German app stores. App quality, origin, and amount of evidence was heterogeneous. Brief descriptions and recommendations were compiled to provide ready-to-use, useful information for potential users and developers.

We recommend continual evaluation of mobile apps based on automatic crawling techniques; quality evaluations by users (patients and physicians); and supporting cost-effectiveness studies to enhance awareness, use, and potential benefit. Furthermore, we would like to emphasize the importance of research institutes and academics as data recipients and partners in app development. Only then can truly powerful data analysis and insights be collected and used for scientific research.

To maximize the great eHealth potential in rheumatology, a close collaboration of patients, rheumatologists, developers, and industry is needed. To avoid redundant work and save time, international and national eHealth consortiums and collaborations are needed to create guidelines and recommendations.

Acknowledgments

The authors thank the German Society for Rheumatology (Working Group Young Rheumatology) for their support of this project. Thanks to Antoine Murray for proofreading the manuscript. External funding was obtained by Pfizer GmbH, Germany, to conduct the presented work.

Authors' Contributions

J Knitza1, KT, MK, and ER wrote the draft manuscript. J Knitza1, MM, DV, AP, PB, J Kittler, DS, and MK rated apps. KT did the statistical analysis. All authors reviewed the draft and provided comments for changes. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Rater gender, hardware and software.

[PDF File (Adobe PDF File), 20KB - mhealth_v7i8e14991_app1.pdf]

Multimedia Appendix 2

Individual Mobile Application Rating Scale section ratings of included apps.

[PNG File, 233KB - mhealth_v7i8e14991_app2.png]

Multimedia Appendix 3

Mobile Application Rating Scale and app store ratings of included apps.

[PDF File (Adobe PDF File), 24KB - mhealth_v7i8e14991_app3.pdf]

Multimedia Appendix 4

Recommendations for developers of mobile rheumatology apps.

[PNG File, 133KB - mhealth_v7i8e14991_app4.png]

References

1. Kataria S, Ravindran V. Digital health: a new dimension in rheumatology patient care. Rheumatol Int 2018 Nov;38(11):1949-1957. [doi: 10.1007/s00296-018-4037-x] [Medline: 29713795]
2. Burmester G. Rheumatology 4.0: big data, wearables and diagnosis by computer. Ann Rheum Dis 2018 Jul;77(7):963-965 [FREE Full text] [doi: 10.1136/annrheumdis-2017-212888] [Medline: 29802224]
3. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 2019 Dec 21;14(1):69 [FREE Full text] [doi: 10.1186/s13023-019-1040-6] [Medline: 30898118]
4. Gossec L, Guyard F, Leroy D, Lafargue T, Seiler M, Jacquemin C, et al. Detection of flares by decrease in physical activity, collected using wearable activity trackers, in rheumatoid arthritis or axial spondyloarthritis: an application of machine-learning analyses in rheumatology. Arthritis Care Res (Hoboken) 2018 Sep 22. [doi: 10.1002/acr.23768] [Medline: 30242992]
5. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, et al. [Use of medical apps and online platforms among German rheumatologists: results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs]. Z Rheumatol 2018 Dec 12. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
6. Albrecht U, Hillebrand U, von Jan U. Relevance of trust marks and CE labels in German-language store descriptions of health apps: analysis. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e10394 [FREE Full text] [doi: 10.2196/10394] [Medline: 29695374]
7. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
8. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
9. 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]
10. Grainger R, Townsley H, White B, Langlotz T, Taylor W. Apps for people with rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]

11. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
13. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
14. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
15. Walker UA, Mueller RB, Jaeger VK, Theiler R, Forster A, Dufner P, et al. Disease activity dynamics in rheumatoid arthritis: patients' self-assessment of disease activity via WebApp. Rheumatology (Oxford) 2017 Dec 01;56(10):1707-1712 [FREE Full text] [doi: 10.1093/rheumatology/kex229] [Medline: 28957553]
16. Swiss Clinical Quality Management in Rheumatic Diseases. mySCQM URL: https://www.scqm.ch/patienten/patienten-apps [accessed 2019-04-11] [WebCite Cache ID 77Y7ytK1H]
17. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019 Apr;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
18. Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, Van Rooyen V, Swallow V. Young people's, parents', and professionals' views on required components of mobile apps to support self-management of juvenile arthritis: qualitative study. JMIR Mhealth Uhealth 2018 Jan 19;6(1):e25 [FREE Full text] [doi: 10.2196/mhealth.9179] [Medline: 29351898]
19. van der Vaart R, Drossaert CHC, Taal E, van de Laar MA. Patient preferences for a hospital-based rheumatology Interactive Health Communication Application and factors associated with these preferences. Rheumatology (Oxford) 2011 Sep;50(9):1618-1626 [FREE Full text] [doi: 10.1093/rheumatology/ker161] [Medline: 21551221]
20. Mollard E, Michaud K. A mobile app with optical imaging for the self-management of hand rheumatoid arthritis: pilot study. JMIR Mhealth Uhealth 2018 Oct 29;6(10):e12221 [FREE Full text] [doi: 10.2196/12221] [Medline: 30373732]
21. Revenäs A, Opava CH, Martin C, Demmelmaier I, Keller C, Åsenlöf P. Development of a web-based and mobile app to support physical activity in individuals with rheumatoid arthritis: results from the second step of a co-design process. JMIR Res Protoc 2015 Feb;4(1):e22 [FREE Full text] [doi: 10.2196/resprot.3795] [Medline: 25665589]
22. Albrecht U, Hasenfuß G, von Jan U. Description of cardiological apps from the German app store: semiautomated retrospective app store analysis. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11753 [FREE Full text] [doi: 10.2196/11753] [Medline: 30459140]
23. Turner-McGrievy G, Hales S, Schoffman D, Valafar H, Brazendale K, Weaver R, et al. Choosing between responsive-design websites versus mobile apps for your mobile behavioral intervention: presenting four case studies. Transl Behav Med 2017 Dec;7(2):224-232 [FREE Full text] [doi: 10.1007/s13142-016-0448-y] [Medline: 27812798]
24. Baptista S, Oldenburg B, O'Neil A. Response to "Development and validation of the user version of the Mobile Application Rating Scale (uMARS)". JMIR Mhealth Uhealth 2017 Jun 09;5(6):e16 [FREE Full text] [doi: 10.2196/mhealth.6419] [Medline: 28600277]
25. 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]
26. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
27. Terhorst Y, Rathner E, Baumeister H, Sander L. ["Hilfe aus dem App-Store?": Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen]. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
28. Santo K, Richtering S, Chalmers J, Thiagalingam A, Chow C, 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]
29. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. npj Digital Med 2018 May 9;1(1):12. [doi: 10.1038/s41746-018-0021-9]

Abbreviations

ANCA: antineutrophil cytoplasmic antibodies
ICC2k: two-way random, average measures, absolute agreement
JIA: juvenile idiopathic arthritis
MARS: Mobile Application Rating Scale
PsA: psoriatic arthritis
RA: rheumatoid arthritis

SLE: systemic lupus erythematosus
SpA: ankylosing spondylitis

Edited by G Eysenbach; submitted 10.06.19; peer-reviewed by H de Sola, G Peters, A Salazar; comments to author 12.07.19; revised version received 17.07.19; accepted 19.07.19; published 05.08.19.

Please cite as:
Knitza J, Tascilar K, Messner EM, Meyer M, Vossen D, Pulla A, Bosch P, Kittler J, Kleyer A, Sewerin P, Mucke J, Haase I, Simon D, Krusche M
German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS)
JMIR Mhealth Uhealth 2019;7(8):e14991
URL: https://mhealth.jmir.org/2019/8/e14991/
doi:10.2196/14991
PMID:31381501

Original Paper

Suicide Prevention Mobile Apps: Descriptive Analysis of Apps from the Most Popular Virtual Stores

Gema Castillo-Sánchez[1*], MSc; Ismael Camargo-Henríquez[2*], MSc; Juan Luis Muñoz-Sánchez[3,4*], PhD; Manuel Franco-Martín[3,4*], PhD, MD; Isabel de la Torre-Díez[1*], PhD

[1]Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Valladolid, Spain

[2]Higher Technical School of Computer Engineers, Polytechnic University of Madrid, Madrid, Spain

[3]Psychiatry Service, Rio Hortega University Hospital, Valladolid, Spain

[4]Psychiatry Service, Zamora Hospital, Valladolid, Spain

[*]all authors contributed equally

Corresponding Author:
Isabel de la Torre-Díez, PhD
Department of Signal Theory and Communications, and Telematics Engineering
University of Valladolid
Paseo de Belén 15
Valladolid, 47011
Spain
Phone: 34 983423000 ext 3703
Fax: 34 983423666
Email: isator@tel.uva.es

Abstract

Background: Provision of follow-up and care during treatment of people with suicidal intentions is a challenge for health professionals and experts in information and communications technology (ICT). Therefore, health professionals and ICT experts are making efforts to carry out these activities in collaboration by using mobile apps as a technological resource.

Objective: This study aimed to descriptively analyze mobile apps aimed at suicide prevention and to determine relevant factors in their design and development. In addition, it sought to analyze their impact on the support of treatment for patients at risk for suicide.

Methods: We considered 20 apps previously listed in the article “Mobile Apps for Suicide Prevention: Review of Virtual Stores and Literature” (de la Torre et al, JMIR mHealth uHealth 2017;5[10]:e130). To find the apps in this list, the most popular app stores (Android and iOS) were searched using the keyword “suicide prevention.” The research focused on publicly available app information: language, platform, and user ratings. The results obtained were statistically evaluated using 16 parameters that establish various factors that may affect the choice of the user, and the consequent support that the app can offer to a person at risk for suicide.

Results: Of the 20 mobile apps, 4 no longer appeared in the app stores and were therefore excluded. Analysis of the remaining 16 apps sampled showed the following: (1) a high percentage of the apps analyzed in the study (n=13, 82%) are provided in English language; (2) the sampled apps were last updated in 2017, when only 45% of them were updated, but the constant and progressive update of treatments should be reflected in the apps; and (3) the technical quality of these apps cannot be determined on the basis of the distribution of scores, because their popularity indices can be subjective (according to the users). User preference for a particular operating system would require further, more specific research, including study of the differences in the technical and usability aspects between both platforms and the design of medical apps.

Conclusions: Although there are positive approaches to the use of apps for suicide prevention and follow-up, the technical and human aspects are yet to be explored and defined. For example, the design and development of apps that support suicide prevention should be strongly supported by health personnel to humanize these apps, so that the effectiveness of the treatments supported by them can be improved.

(JMIR Mhealth Uhealth 2019;7(8):e13885) doi:10.2196/13885

KEYWORDS

apps; prevention; suicide; virtual store; analysis

Introduction

The recent significant increase in the use of mobile devices can be highlighted as a high-impact technological revolution. The recent increase in health-oriented mobile technologies and apps (mobile health) developed to construct a new modality for assistance and treatment of patients is evident, and the data on people's health collected by these apps are valuable [2].

Technically, the primary objective of mobile apps is to offer, and facilitate access to, information and knowledge without restrictions of time or physical space. This possibility of technological ubiquity creates opportunities for new forms of use and communication from these resources [3-5]. Thus, mobile apps become a possible and real alternative for the treatment of mental health problems, especially for people at risk for suicide in follow-up and treatment [6]. As Larsen et al [7] pointed out, "Apps can be especially useful for suicide prevention interventions, because of their ability to deliver support and interventions in situ and at times of crisis."

In recent years, statistics related to acts and attempts of suicide or self-injury seem to maintain a concerning growing trend. The World Health Organization estimates that more than 800,000 people worldwide die of suicide each year; in addition, suicide is the second leading cause of death among young people aged 15-29 years [8].

Several studies [9,10] have identified the positive effects of new technologies in combating and preventing suicidal behavior. For example, the academic works of Saulsberry et al [1], Carew et al [11], and Hetrick et al [12] indicated different technological resources that can help people in these situations. Thus, discussion and help forums, online treatments, or search for information on the internet has created an interesting positive impact. Similarly, mobile apps have found their niche as technological resources to support this problem.

Some studies on the subject (eg, [13]) suggest that apps that include patient-generated information can be used to optimize outcomes and reduce risks in specific health problems such as suicide. They are also able to identify and understand the technical determinants (eg, design, perceived usefulness, perceived ease of use, and autonomy) that promote the adoption of apps as a habit, which contributes to the patients' well-being or may lead to the prevention of a mental condition through its previous manifestations (behaviors) or symptoms.

A randomized trial of the effectiveness of the *MoodHacker* app showed interesting results regarding depression symptoms among adult patients who used the app compared with patients who used websites [14]; in the trial with a 6-week follow-up, the app showed significant effects on the symptoms of depression, with η^2=0.021 among employed adult patients who used the app compared with subjects with the same condition but with access to websites instead of the app. The app also had stronger effects on individuals with access to an employee assistance program (η^2=0.093). For all users, the *MoodHacker* program showed an improvement in work absence and the mediating factors of behavioral activation, negative thoughts, and knowledge of depression self-care. The study also reported that through this app and the guidance of clinical advisors, the effectiveness of the follow-up for patients who seek support for their treatment, without disconnecting from the human bond, can be significantly improved. Thus, mobile apps that are effectively designed and implemented for mental health treatments should follow particular guidelines to attain the objectives set for the patient.

This study aimed to analyze mobile apps for suicide prevention in order to identify and characterize the main factors, variables, or parameters used for the design and development of these types of mobile apps. In addition, we aimed to determine whether these apps support potential suicide-prone patients in their functionality and thus provide a series of reflections that reinforce previous contributions established by other studies [15,16,17]. This goal may be accomplished by establishment of a simple and clear methodology for identifying the characterizations of these apps and the effect they may have on suicide prevention. This study is an extension of a previous publication [18] and complements it by adding a statistical and descriptive analysis of the characteristics of the 20 previously investigated apps.

Methods

Overview

On the basis of the experience of mental health professionals from Hospital de Zamora, Castilla y León (Spain), and information and communications technology (ICT) professionals, the following variables were selected to generally describe the publicly available mobile apps in app stores. These variables or parameters were selected per the study's objective to quantitatively analyze a set of apps aimed at suicide prevention. The research focused on publicly available app information: language, platform, and user ratings.

These apps were originally listed in a previous publication in 2017 [18] and initially consisted of a list of 20 apps. However, in 2018, the apps were reverified, as four of them no longer appeared in the app stores; this difference was due to the variability of the duration for which the apps are available in the app store. For iOS, apps remain valid for more than 9 months but for Android, they remain valid for a shorter time [17]. The apps *Not Even One*, *Suicide Helplines in India*, *Suicide Thoughts*, and *Suicide or Survive* were excluded, because they did not appear in the app search in 2018. Thus, this section presents the analysis of 16 mobile apps in app stores.

Selection Criteria

Methodologically, a total of 18 variables were established to characterize the technical and functional properties of the reviewed apps. These variables generally correspond with the metadata that can be obtained from the revised app stores and offer a global statistical perspective on the apps they manage.

Table 1. List of the *Features of Interest*.

Variable name	Variable type	Description
Name	Text	Name of the app
Developer	Text	Developer information
Language	Text	Main language *by default*, in which the functional features of the app are offered to the user (interfaces, guides, help)
Total downloads	Number	Total of downloads
Overall rating	Number	Average rating in the store
Total reviews	Number	Total of reviews
Category	Text	App category according to the store
Screenshots	Categorical	Screenshots available from the app
Current version	Number	Current app version
Last update	Date	Latest app update
In-app purchases	Categorical	Possibility to buy things within the app
In-app ads	Categorical	Advertisements within the app
Download size	Number	Application download size (MB)
Developer country	Text	Developer country information
Access	Categorical	Free/premium app
Platform	Categorical	Platform type (Android / iOS)
User journey	Categorical	User guidance on first use
Description	Text	As given by the developer

These variables are listed, defined, and described in Table 1 and are used to define the set of criteria to include, classify, and organize the analyzed apps. This series of properties of the apps selected is called *Features of Interest*.

The variables, or *Features of Interest*, including the name, developer, and total downloads (Table 1) help identify the app and compile a series of descriptive data to distinguish and evaluate them. The variables user journey, screenshots, in-app purchases, and in-app ads allowed us to evaluate characteristics of the apps that could influence their download and use by people in treatment and follow-up.

In addition, the rest of the *Features of Interest* presented here are available in app stores and allow us to determine, for example, the group of people who can use them when estimating the Language variable. This is helpful because these apps can be customized to the language of the population to which suicide attention or prevention can be provided. Thus, after selecting and determining the variables to be used, the data collected are normalized and cleaned (eg, formatting, spellings, duplication, and extra spaces) to provide a consistent basis for analysis. To subsequently construct the final set of data (data that will be used with the analysis tools), we started by using the initial raw data acquired from the search in the app stores in the following process: selecting the data captured in tables, organizing the records, and reviewing the attributes of these data according to the variables.

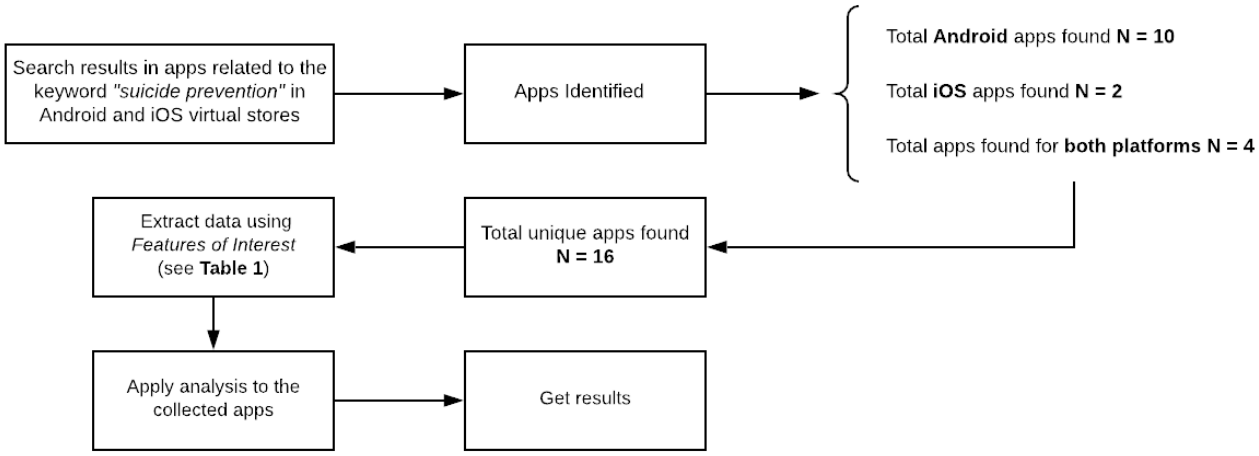
Search Strategy, Data Synthesis, and Analysis

We applied the *Features of Interest* to a data search and analysis strategy proposed in Figure 1. This figure describes the process of data acquisition through app searches based on keywords, classification, and organization of the intermediate results obtained from the virtual stores (Android and iOS); selectivity and analysis of data through the *Features of Interest*; and the results obtained and presented in this paper.

As shown in Figure 1, we searched the app stores using the keyword “suicide prevention” and identified 10 Android apps, 2 iOS apps, and 4 apps for both platforms. The searches were conducted and results were obtained in October 2018, leading to identification of 16 apps presented in this study. The series of parameters described in Table 1 were applied to this set as the data extraction criteria. Our search is similar to those used in previous studies (Android [19], iOS [20], both [18]). In our first study [18], an exploratory and descriptive investigation was carried out, which allowed us to derive basic aspects that are useful for this study, of which the *Features of Interest* was the most important.

As a preliminary result, we obtained a list of apps that matched the search criteria: 10 mobile apps in the Android platform, 2 mobile apps in the iOS platform, and 4 apps found could be run on both platforms. From this list, individual review of each app was performed by evaluating and recording the values that each app contributed as per the previously defined *Features of Interest* (Table 1). The completeness of the recording of these values for each app allowed us to analyze the general characteristics of the apps, with reasoning based on the quantification of the values obtained for each feature of interest. Finally, the results produced were organized and presented statistically.

Figure 1. Data search and analysis strategy.



Results

Features of Interest

This section briefly describes the results obtained from the analysis of the 16 selected apps. The general considerations provided in Tables 2 and 3 correspond to the most relevant *Features of Interest* previously established.

Table 2. *Features of Interest* of the apps (N=16).

Feature	Values, n (%)
Language	
English	13 (82)
Dutch	1 (6)
Spanish	1 (6)
Multi-language	1 (6)
Developer country	
Canada	1 (6)
Scotland	1 (6)
Spain	1 (6)
United Kingdom	2 (13)
United States	6 (38)
Not specified	5 (31)
Screenshots	
Yes	15 (94)
No	1 (6)
In-app ads	
Yes	1 (6)
No	15 (94)
Platform	
Android	10 (63)
iOS	2 (13)
Both	4 (25)
User journey	
Yes	12 (75)
No	4 (25)
Last update	
2013	1 (6)
2015	1 (6)
2016	4 (25)
2017	4 (25)
2018	6 (38)

Table 3. App download, rating, and review data (part of the *Features of Interest* list).

App name	Overall rating[a]	Total reviews, n	Total downloads, n
DMHS[b] Suicide Prevention	3.80	4	100
Suicide Prevention Help Squads	3.70	14	1000
Prevent Suicide - NE Scotland	5.00	27	1000
HELP Prevent Suicide	4.00	5	500
Ask & Prevent Suicide	4.60	13	1000
Suicide? Help!	2.90	72	10000
Prevensuic	4.30	49	1000
Stay Alive	3.80	201	10000
Suicide Preventive	3.20	28	5000
Suicide Lifeguard	3.10	40	1000
Suicide Safety Plan	3.90	179	10000
A Friend Asks	4.20	17	5000
Virtual Hope Box	4.20	611	10000
Operation Reach Out	4.00	24	1000
We Care	5.00	4	1000
R U Suicidal	2.70	30	1000

[a]1=poorly rated, 5=highly rated.

[b]DMHS: Durham Mental Health Services

Language

A high percentage of the apps analyzed in the study (82%) are presented in English. The remaining 12% are offered to their users in Spanish or Dutch, and only 6% of the apps are presented with multilingual capability. Notably, the proportion of apps in a particular language (English) could limit the app's access to patients who do not know that language, reducing the impact and use of these apps for patients with a regional language other than those offered by the apps analyzed.

Developer's Country

The predominance of English language in the analyzed apps was correlated with the countries of origin in which they were developed: 63% of the apps reviewed were developed collectively in the United States (38%), the United Kingdom (13%), Canada (6%), and Scotland (6%). On the contrary, only one of the apps was developed in Spain (6%). Finally, the origin could not be identified for 31% of the apps.

Total Downloads

Table 3 lists the most outstanding apps that, at the time of the analysis, had an average of 10,000 downloads. The following four apps were identified in this range of downloads: *Suicide? Help*, *Stay Alive*, *Suicide Safety Plan*, and *Virtual Hope Box*. The following were apps with 5000 downloads: *Suicide Preventive* and *A Friend Asks*. The apps *HELP Prevent Suicide* and *DMHS Suicide Prevention* had the lowest download average numbers of 500 and 100, respectively. Finally, the remaining apps in the evaluated set followed the trend of at least 1000 downloads. A possible indicator that can influence download averages could be the popularity of such apps among users, personal (or clinical) recommendations, and ranking within the app store. However, these are not determinants to qualify the relative effectiveness of the apps as tools for suicide prevention or support treatments. The total number of downloads was obtained from the apps available on the Android platform.

Overall Rating

During analysis, we found that use of Overall Rating as a comparative indicator in relation to the total downloads of the apps reinforces the argument that the download index or popularity of an app is not directly related to its effectiveness as a tool to support treatments and suicide prevention. Table 3 shows that apps that obtained a high index of downloads (*Suicide? Help!*, *Stay Alive*, *Suicide Safety Plan*, and *Virtual Hope Box*) now present average ratings of 2.90, 3.80, 3.90, and 4.20 (on a maximum rating of 5.00), respectively. However, they do not exceed the rating of 5.00, similar to apps with lower downloads, such as *Prevent Suicide - NE Scotland* or *We Care*, that had an average of 500 and 1000 downloads, respectively.

Total Reviews

Table 3 presents a new comparison variable. The apps most commented on or reviewed by the public (users and relatives) present heterogeneous values, highlighting only three apps with high values: *Stay Alive* (201 reviews), *Suicide Safety Plan* (179 reviews), and *Virtual Hope Box* (611 reviews). The total number of reviews could be associated with the popularity of the apps due to the influence of the total downloads on the rating given to each app. However, the rating could continue to be a subjective metric due to its reliance on opinions with a strong component of feelings (negative or positive) toward the app.

Screenshots

Most of the evaluated apps (n=15, 94%) presented screenshots showcasing the app's features to potential users before they downloaded the app; such screenshots offer a brief overview of the app's esthetic aspects and most relevant functionalities.

Last Update

An aspect valued in the evaluation and analysis of apps is the date of updates or the frequency in which their developers include new features in the apps. For example, many of the apps (n=6) were updated until 2018 (Table 2). However, these values are relative and, to a great extent, depend on the continuity of the app by its developers and their affiliation with private or public organizations that collaborate in the app's development and growth (through funding, inclusion of new treatments, helplines, and groups).

In-App Ads

We observed a trend of noninclusion of ads within the content of the apps. This feature is useful, as it avoids detracting the patient's concentration or overwhelming the patient with unwanted advertising that would compromise the objectives of the app (treatment, help, and follow-up) and thus discourage app use due to disinterest.

Platform

Two platforms (Android and iOS) and their corresponding app stores were evaluated. We observed a marked trend of development of apps oriented toward the Android platform (63%), which can be associated with the popularity and assortment of mobile devices (cell phones and tablets) for this platform. In addition, notably, there was an orientation toward apps with a duality of operation between platforms (Android ↔ iOS). In our analysis, 25% (n=4) of the reviewed apps showed this duality by providing access to users of one or the other platform. Finally, 13% (n=2) of all apps under study run on the iOS platform natively.

User Journey

An important feature of this study was consideration of the user journey as a teaching resource within the evaluated apps; 12 apps reviewed (75%) present this resource as a part of their design, with 4 apps not using this component. Designers seem to support the good use of the app by the patient in a guided manner, which promotes their care and supports the objectives of the app itself in relation to the patient.

Discussion

Principal Findings

Our findings suggest that *Virtual Hope Box* has many factors that fulfill the investigated parameters. First, it is the only app among the apps analyzed that offered six languages, which allows it to reach a larger audience in different regions. This app had the highest numbers of downloads (n=10,000) and reviews (n=611) from among the 16 apps, possibly due to this reason. In addition, it had an average rating of 4.20 and no in-app advertisements or purchase options, which may be discouraging to the users. Another important and significant aspect is that the app is updated frequently; its last update was in October 2018. Thus, it maintains constant support. Finally, the users' perception of the app with regard to the omission of the word "suicide" as part of the app's name is notable. The word "suicide" is often associated with negative emotions, and use of positive words such as "hope" as part of the app name is a better option for the prevention of such incidents.

The *Virtual Hope Box* was developed at the National Center for Tele-Health and Technology (US Department of Defense) and is based on the analog version of the hope box concept, a therapeutic tool used by clinicians to help patients with depression or suicidal tendencies to redirect their negative thoughts toward reflecting on reasons for living instead [21]. This app is particularly interesting because, based on the evaluated results, the technical elements of its use and design could be used in the development of similar apps.

Limitations

That there were some limitations to the set of apps sampled, and these limitations condition the presentation of the analyzed data. First, combining the Android and iOS apps into one row has resulted in a major technical flaw. For example, in case the app is available for both Android and iOS, the data recorded for downloads, ratings, and all other variables do not clearly indicate whether the data are for the Android or iOS platform. Each app's data should have been recorded separately. Second, as the number of apps (N=16) was very low, we could not prove statistical significance for any of our results. For example, there were only two iOS apps; therefore, no result could be validated. Third, this study did not examine the characteristics associated with the app content, clinical appropriateness, compliance with suicide prevention policy guidelines, or the scientific evidence base for the included apps, which limits the extent to which this analysis can evaluate such apps beyond providing an overview of details such as current usage and user ratings.

Conclusions

Establishing parameters to identify mobile apps for monitoring, prevention of, and provision of attention to suicide is not a trivial task. However, this joint effort between ICT experts and mental health professionals has made it possible to generally describe the metadata provided in app stores in terms of the apps offered for these purposes. These data allow us to observe certain behaviors of interest related to downloads, determine the predominant language, use technological resources, and confirm previously established characteristics. Analysis of the 20 mobile apps sampled showed the following main findings: (1) a high percentage of the apps analyzed in the study (82%) are provided in English language; (2) the sampled apps were last updated in 2017, when only 45% of them received an update, but the constant and progressive update of treatments should be reflected in the apps; and (3) it is impossible to accurately determine the technical quality of these apps based on the distribution of scores (Table 3) because the popularity indices shown in Table 3 can be subjective (according to the users). Therefore, it would be necessary to obtain information on the design protocols, development, verification of applications, technical resources for the platforms, tools, engineering

techniques used, etc. Thus, further evaluation is needed to determine this accurately.

Finally, the apps designed for suicide prevention and follow-up for patients at risk for suicide could be more successful if they were actively supported in their design by qualified mental health professionals, who could provide humanization by following up or treating people who, in these critical situations, need help using an app as a communication resource. Thus, the accompaniment of specialized medical personnel becomes a determining factor in suicide prevention because the components that cause it are complex and as variable as the different circumstances faced by people at risk. Facilitating treatment with this type of technology, with an adequate human approach, could save lives, but it will still be a challenge to implement for health professionals working in combination with ICT experts.

Future Work

Research will continue in this area and will be taken into account for the design and development of a mobile app with the possibility for it to be tested in patients with a suicidal risk.

Acknowledgments

This research has been partially supported by the European Commission and the Ministry of Industry, Energy, and Tourism under the project AAL-20125036 named *Wetake Care: ICT-based Solution for (Self-) Management of Daily Living*.

Conflicts of Interest

None declared.

References

1. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized clinical trial of a primary care internet-based intervention to prevent adolescent depression: one-year outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
2. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza J, et al. Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed Eng Online 2015;14(Suppl 2):S6 [FREE Full text] [doi: 10.1186/1475-925X-14-S2-S6] [Medline: 26329639]
3. Boulos MN, Brewer A, Karimkhani C, Buller D, Dellavalle R. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
4. Keengwe J, Bhargava M. Mobile learning and integration of mobile technologies in education. Educ Inf Technol 2013 Jan 26;19(4):737-746. [doi: 10.1007/s10639-013-9250-3]
5. Clay CA. Exploring the use of mobile technologies for the acquisition of clinical skills. Nurse Educ Today 2011;31:6. [doi: 10.1016/j.nedt.2010.10.011]
6. Muñoz-Sánchez JL, Sánchez-Gómez MC, Martín-Cilleros MV, Parra-Vidales E, de Leo D, Franco-Martín MA. Addressing suicide risk according to different healthcare professionals in Spain: a qualitative study. Int J Environ Res Public Health 2018 Dec 26;15(10):2117 [FREE Full text] [doi: 10.3390/ijerph15102117] [Medline: 30261622]
7. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
8. WHO. World Health Organization. 2016. Suicide Across The World (2016) URL: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/
9. Luxton DD, June J, Kinn J. Technology-based suicide prevention: current applications and future directions. Telemed J E Health 2011;17(1):50-54. [doi: 10.1089/tmj.2010.0091] [Medline: 21214382]
10. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
11. Carew C, Kutcher S, Wei Y, McLuckie A. Using digital and social media metrics to develop mental health approaches for youth. Adolesc Psychiatry 2014 Jul 31;4(2):116-121. [doi: 10.2174/2210676604020140709122825]
12. Hetrick SE, Dellosa M, Simmons M, Phillips L. Development and pilot testing of an online monitoring tool of depression symptoms and side effects for young people being treated for depression. Early Interv Psychiatry 2015 Feb;9(1):66-69. [doi: 10.1111/eip.12127] [Medline: 24612591]
13. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
14. Birney AJ, Gunn R, Russell J, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
15. Nuij C, van Ballegooijen W, Ruwaard J, de Beurs D, Mokkenstorm J, van Duijn E, et al. Smartphone-based safety planning and self-monitoring for suicidal patients: rationale and study protocol of the CASPAR (Continuous Assessment for Suicide

Prevention And Research) study. Internet Interv 2018 Sep;13:16-23 [FREE Full text] [doi: 10.1016/j.invent.2018.04.005] [Medline: 30206514]
16. Aguirre R, McCoy M, Roan M. Development guidelines from a study of suicide prevention mobile applications (apps). J Technol Hum Serv 2013 Jul;31(3):269-293. [doi: 10.1080/15228835.2013.814750]
17. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 9;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
18. de la Torre I, Castillo G, Arambarri J, López-Coronado M, Franco M. Mobile apps for suicide prevention: review of virtual stores and literature. JMIR Mhealth Uhealth 2017 Oct 10;5(10):e130 [FREE Full text] [doi: 10.2196/mhealth.8036] [Medline: 29017992]
19. Google Play Store. 2018. Android applications URL: https://play.google.com/store/apps [accessed 2018-08-30]
20. Apple Store. 2018. App Store - Find the apps URL: https://www.apple.com/ios/app-store/ [accessed 2018-08-27]
21. Luxton D, June J, Chalker S. Mobile health technologies for suicide prevention: feature review and recommendations for use in clinical care. Curr Treat Options Psych 2015 Sep 26;2(4):349-362. [doi: 10.1007/s40501-015-0057-2]

Abbreviations

ICT: information and communications technology

Edited by G Eysenbach; submitted 07.03.19; peer-reviewed by JL Fernández-Alemán, M Larsen, R Melia; comments to author 28.04.19; revised version received 24.05.19; accepted 29.05.19; published 13.08.19.

Please cite as:
Castillo-Sánchez G, Camargo-Henríquez I, Muñoz-Sánchez JL, Franco-Martín M, de la Torre-Díez I
Suicide Prevention Mobile Apps: Descriptive Analysis of Apps from the Most Popular Virtual Stores
JMIR Mhealth Uhealth 2019;7(8):e13885
URL: http://mhealth.jmir.org/2019/8/e13885/
doi:10.2196/13885
PMID:31411144

Original Paper

Behavior Change Content, Understandability, and Actionability of Chronic Condition Self-Management Apps Available in France: Systematic Search and Evaluation

Luiza Siqueira do Prado[1], MSc; Camille Carpentier[2], MSc; Marie Preau[2], PhD; Anne-Marie Schott[1,3], MD, PhD; Alexandra Lelia Dima[1], PhD

[1]Équipe d'Accueil 7425 Health Services and Performance Research, Université de Lyon, Lyon, France

[2]Équipe d'Accueil 4163 Groupe de Recherche en Psychologie Sociale, Psychology Institute, Université Lumière Lyon 2, Bron, France

[3]Centre Hospitalier Lyon Sud, Pôle de Santé Publique, Hospices Civils de Lyon, Lyon, France

Corresponding Author:
Luiza Siqueira do Prado, MSc
Équipe d'Accueil 7425 Health Services and Performance Research
Université de Lyon
8 avenue Rockefeller
Lyon, 69373
France
Phone: 33 4 26 68 82 23
Email: luiza.siqueira-do-prado@univ-lyon1.fr

Abstract

Background: The quality of life of people living with chronic conditions is highly dependent on self-management behaviors. Mobile health (mHealth) apps could facilitate self-management and thus help improve population health. To achieve their potential, apps need to target specific behaviors with appropriate techniques that support change and do so in a way that allows users to understand and act upon the content with which they interact.

Objective: Our objective was to identify apps targeted toward the self-management of chronic conditions and that are available in France. We aimed to examine what target behaviors and behavior change techniques (BCTs) they include, their level of understandability and actionability, and the associations between these characteristics.

Methods: We extracted data from the Google Play store on apps labelled as *Top* in the *Medicine* category. We also extracted data on apps that were found through 12 popular terms (ie, keywords) for the four most common chronic condition groups—cardiovascular diseases, cancers, respiratory diseases, and diabetes—along with apps identified through a literature search. We selected and downloaded native Android apps available in French for the self-management of any chronic condition in one of the four groups and extracted background characteristics (eg, stars and number of ratings), coded the presence of target behaviors and BCTs using the BCT taxonomy, and coded the understandability and actionability of apps using the Patient Education Material Assessment Tool for audiovisual materials (PEMAT-A/V). We performed descriptive statistics and bivariate statistical tests.

Results: A total of 44 distinct native apps were available for download in France and in French: 39 (89%) were found via the Google Play store and 5 (11%) were found via literature search. A total of 19 (43%) apps were for diabetes, 10 for cardiovascular diseases (23%), 8 for more than one condition in the four groups (18%), 6 for respiratory diseases (14%), and 1 for cancer (2%). The median number of target behaviors per app was 2 (range 0-7) and of BCTs per app was 3 (range 0-12). The most common BCT was *self-monitoring of outcome(s) of behavior* (31 apps), while the most common target behavior was *tracking symptoms* (30 apps). The median level of understandability was 42% and of actionability was 0%. Apps with more target behaviors and more BCTs were also more understandable (ρ=.31, *P*=.04 and ρ=.35, *P*=.02, respectively), but were not significantly more actionable (ρ=.24, *P*=.12 and ρ=.29, *P*=.054, respectively).

Conclusions: These apps target few behaviors and include few BCTs, limiting their potential for behavior change. While content is moderately understandable, clear instructions on when and how to act are uncommon. Developers need to work closely with health professionals, users, and behavior change experts to improve content and format so apps can better support patients in coping with chronic conditions. Developers may use these criteria for assessing content and format to guide app development and evaluation of app performance.

Trial Registration: PROSPERO CRD42018094012; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=94012

(JMIR Mhealth Uhealth 2019;7(8):e13494) doi:10.2196/13494

KEYWORDS

mHealth; mobile phone; app; self-management; chronic conditions; target behaviors; behavior change techniques; understandability; actionability

Introduction

Chronic conditions are the main cause of disability and premature death worldwide, representing the highest number of disability-adjusted life years in the Global Burden of Disease project [1]. More specifically, four groups of diseases—cardiovascular diseases, cancers, respiratory diseases, and diabetes—cause 80% of premature deaths related to chronic conditions [2]. The rise in prevalence of such conditions, while being determined by multiple causes, is highly related to unhealthy lifestyles and population aging. Treatment requires a long-term and multidisciplinary approach, including therapeutic education and lifestyle changes, to prevent further aggravation and/or premature death; treatment focuses on modifiable behavioral risk factors, such as insufficient physical activity or inadequate diet [2]. Achieving and maintaining satisfactory quality of life is strongly dependent on the patient's ability to reduce behavioral risks and to regularly perform specific self-care activities defined together with health care providers (ie, self-management behaviors). This process of active engagement in obtaining skills and taking part in health-related decisions is also called patient empowerment [3]; this may be mediated by technology that facilitates self-awareness and understanding how and when to take action regarding measured physiological parameters, to prevent or react to deterioration in health status.

Currently, mobile health (mHealth) mobile phone apps can support patients in performing behaviors such as symptom monitoring and medication intake, among others [4]; therefore, they have the potential to help improve population health and reduce health care costs. By the end of 2017, mobile broadband subscriptions were expected to reach 4.3 billion worldwide [5]. In 2012, one in five mobile phone users had at least one health-related mobile app on his or her phone [6]; in 2015, over half of that population had downloaded at least once a health-related app [7]. In addition, numbers of mHealth apps downloaded are similar between individuals with and without chronic conditions [8]. The availability of mHealth apps and people's tendency to carry their devices with them at all times mean they can also be used for delivering behavioral interventions to large populations [9]. Yet, despite the increasing number of studies and reviews on the use of such apps on health outcomes, evidence for effectiveness is still unclear [10,11]. Furthermore, in the fast-paced technology culture, few mHealth interventions are designed in collaboration with patients, clinicians, or behavioral scientists or are subject to rigorous testing [4,12]. The result is a wide and heterogeneous range of offerings that differ in their objective, content, and user experience [13].

To assess the potential of self-management apps to improve individual and population health, it is useful to consider them as technology-mediated health behavior change interventions. For such interventions to be effective, they need to intervene on the causal behavioral pathways relevant to the health of their user group (ie, to target specific behaviors causally linked to the desired outcomes). Moreover, from a psychological perspective, they need to include behavior change techniques (BCTs), which are the *active ingredients* of behavior change interventions: reproducible and irreducible components of these interventions that can trigger change in the psychological determinants of these behaviors and, consequently, improve health [14]. In recent years, health behavior theorists and intervention developers have been building consensus on methods to identify BCTs present in existing interventions, which resulted in a 93-BCT taxonomy that is currently used as a shared framework for intervention evaluation and development [15]. The presence of behavioral change content has been shown to increase the effectiveness of mHealth and Internet-based interventions [16,17]. Examining the mHealth app offerings in terms of occurrence of target behaviors and BCTs can be informative regarding the current state-of-the-art on behavior change; it can also highlight opportunities for improvement [18], for example, by studying links between usage patterns of individual behavior change content and changes in health outcomes.

The presence of relevant behavior change content does not by itself guarantee that users will be able to interact with it and potentially change their behaviors. These tools need to present information in an accurate, comprehensible, and actionable way; they must also consider different communication competencies, styles, and health literacy levels to optimize their reach and enhance health decision making [19]. Evidence shows that most educational materials are too complex for patients with low health literacy [20]. To assess the suitability of mHealth apps for diverse audiences, it is useful to consider them as health-related materials. The Patient Education Material Assessment Tool for audiovisual materials (PEMAT-A/V) is a commonly used method in this domain; this tool evaluates the extent to which health-related materials are understandable (ie, understandability) and give clear instructions regarding actions that users may take to apply the information presented (ie, actionability) [20]. Presenting relevant behavior change content in a suitable manner is therefore important for ensuring that the intended users achieve the goals the app is supposed to facilitate. Apps with richer behavioral content may also present it in a more understandable and actionable way, and this can be seen as an indicator of app quality and of the level of expertise of the developing team. Yet, to our knowledge, no examination of both content and suitability of apps was performed to date. Understanding the links between content and format in the

current app offerings in a specific territory may provide insights into the rapidly evolving app development phenomenon and recommendations for improvement.

As mHealth develops worldwide, evaluations of app content and format are increasingly common and necessary to inform policy discussions on achieving the potential of mHealth to improve public health [21]. For example, BCTs were shown not to be widely implemented in top-ranked physical activity apps in the United States [22]. In New Zealand, in physical activity and dietary apps, BCTs associated with increased effectiveness in modifying these behaviors were more common in paid apps [23]. In Canada, theory-based cognitive-behavioral content was found to be present in only 10% of apps for depression [24]. Although app use is a global phenomenon, research is normally limited to apps in English, while the available offerings in commercial marketplaces are restricted to geographical regions. The French Health National Strategy 2018-2022 has forecasted a generalization of digital services in health care and put special interest in promoting favorable health behaviors and fighting social inequities in access to health care [25]. It is thus important to assess the current mHealth app offerings in France and in French, especially for self-management of chronic conditions. To our knowledge, no review with these characteristics has been published to date worldwide, and the potential of these tools to support behavior change for self-management of chronic conditions has not yet been fully examined. Such evaluation is instrumental for orienting the development of this expanding field in a way that best serves the interests of all stakeholders, including patients, health care professionals, app developers, payers, and the health care system.

Therefore, this systematic review of mHealth apps for chronic condition self-management in France aimed to answer the following questions: (1) What behaviors are targeted in these apps, and by which BCTs?; (2) What levels of understandability and actionability characterize these apps?; and (3) Are apps with more behavioral change content also easier to understand and act upon?

Methods

Overview

We developed a systematic review protocol based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines, an evidence-based minimum set of items for reporting systematic reviews [26]. We registered the protocol with the International Prospective Register of Systematic Reviews (PROSPERO), an international database of prospectively registered systematic reviews (registration number: CRD42018094012). The PRISMA checklist [26] is available in Multimedia Appendix 1.

Apps were identified through two different approaches: (1) a search of peer-reviewed articles reporting on development or validation of mHealth apps for self-management of chronic conditions and (2) a search in the Android commercial marketplace for mobile phones. We used the Android marketplace Google Play store as it represents 88% of the global mobile phone market [27] and most apps are developed for both operating systems (ie, Android and iOS).

A systematic search of PubMed (ie, MEDLINE), IEEE, and Web of Science electronic bibliographic databases was conducted; all search terms are available in Multimedia Appendix 2. We searched for peer-reviewed articles and conference papers published between 2012 and 2018 concerning mHealth self-management interventions for the previously stated chronic conditions. Articles and papers had to report on empirical research on the development or validation of mHealth tools, pilot studies, or randomized controlled trials, both protocols and reports of study results. Articles and papers were assessed independently by two investigators (LSdP and CC) based on title and abstract, followed by full-text examination to identify available apps in France, in French, and for Android from the Google Play store.

Subsequently, a list of the first 500 free apps labelled as *Top* in the Google Play store in the *Medicine* category and the first 55 paid apps available in the same category was extracted (n=555). The number of paid apps was limited by the marketplace. Another search was performed using 12 keywords in French related to the four groups of diseases: cardiovascular diseases, cancers, respiratory diseases, and diabetes. The keywords were as follows: “maladie cardiaque,” “maladie coeur,” “AVC accident vascular cérébral,” “infarctus,” “maladie pulmon,” “asthme,” “BPCO,” “maladie respiratoire,” “cancer,” “diabète,” “diabète type 1,” and “diabète type 2.” For each keyword, the first 20 apps shown in the marketplace (n=240) were extracted. The complete list (n=795) was assessed independently by two investigators. After screening, selected apps were divided into five groups: four according to the conditions they targeted—*cancers*, *diabetes*, *respiratory diseases*, or *cardiovascular diseases*— and one generic group (ie, *other*) for apps that targeted behaviors like medication adherence and physical activity, irrespective of medical condition (ie, they did not specifically reference a disease within the previous four groups).

All searches, data extraction, and coding were done between March and April 2018. Coding was performed a second time in October 2018 by a different investigator, along with interrater reliability, reconciliation, and coding review. A Samsung J7 mobile phone with Android, version 6.0.1, was used for downloading and evaluating all selected apps.

Eligibility

All native apps available for download in France and in French at the Google Play store designed for patients for the self-management of cardiovascular diseases, cancers, respiratory diseases, and diabetes were eligible. There was no restriction considering price or the source of for-profit or not-for-profit funding. The following apps were excluded: apps that were clearly not for chronic conditions (ie, apps that did not state their main purpose or were designed for other users or purposes, such as training health professionals or students, hospitals, or medical laboratories; making medical appointments; reaching emergency services; or geographical localization of defibrillators or pharmacies); apps for chronic conditions other than those in the four groups studied in this work; apps for chronic conditions in the four groups that were not for self-management (ie,

offering general health information or risk assessment); apps that required additional hardware, like special glucometers; apps with descriptions in French but content in English or with different names and same content; and, finally, apps that were no longer available in October 2018.

Eligible apps may describe their objective as chronic disease self-management or target only specific behaviors relevant for these conditions, such as medication adherence, trigger management, exacerbation management, physical activity, dietary behaviors, and archiving health information. Apps that targeted only preventive behaviors, such as physical activity and diet, with no reference to chronic condition management were also excluded.

Screening and Selection

The reference management software, Zotero (Corporation for Digital Scholarship), was used to identify and remove duplicate records in the literature search. Titles and abstracts of remaining records were screened by two independent reviewers to establish eligibility. If two reviewers recommended study or app inclusion, the full text of the study or the app availability was sourced for review and appraisal in order to determine eligibility for this study. If reviewer discordance arose, consensus was reached through discussion and arbitration with a third investigator.

Data Extraction and Analysis

The following app characteristics were extracted from the Google Play store, along with app's name and available description: information on the presence of sales on the app (ie, paid app or free app with or without paid features), number of downloads (ie, from 10+ to 10,000,000+), user ranking (ie, from 1 to 5 stars), number of ratings, version, last update, and developer information. Developers were then categorized into three groups: (1) *private company*, comprising single-app developers and dedicated app-developing companies; (2) *nonprivate*, comprising nongovernmental organizations, public institutions, or European projects; and (3) *pharmaceutical and medical device companies*, comprising bigger players in the market, such as pharmaceutical laboratories and other medical technology companies.

Target Behaviors and Behavior Change Content

Target behaviors were coded following detailed examination of app content and further functions, like sending notifications to users to perform tasks such as drinking water, exercising, etc. Each target behavior present was coded once per app; one app could contain several target behaviors. Behavior change content was coded by a trained investigator (LSdP) using the BCT taxonomy [15]. This taxonomy represents a consensus of hierarchically structured techniques developed to specify behavioral interventions. Each BCT present was coded once for each app; one app could contain multiple BCTs. A second trained coder (ALD) evaluated a subset of 8 selected apps out of 44 (18%); interrater reliability was computed with bias-adjusted kappa.

Understandability and Actionability Assessment

Material is understandable and actionable when users of different health literacy levels can "process and explain key messages" and "identify what they can do based on the information presented" [20]. Understandability and actionability levels were evaluated using the PEMAT-A/V [20], which is a systematic method to evaluate understandability and actionability of patient education materials. It includes 13 items in five topics—Content, Word Choice & Style, Organization, Layout & Design, and Use of Visual Aids—to evaluate understandability (eg, "The material makes its purpose completely evident," "The material uses common, everyday language," etc) and four items for actionability (eg, "The material clearly identifies at least one action the user can take"). Each item was rated with 0 (If Disagree) or 1 (If Agree), while items that were not applicable received N/A. Item scores were added and divided by the maximum score possible, excluding items that were not applicable, and the result multiplied by 100 to get a percentage score. A second coder assessed understandability and actionability scores for the same subset of apps used for BCT coding; interrater reliability was computed using the intraclass correlation coefficient.

Data Analysis

Coding was done using Microsoft Excel and all statistical analyses were performed using RStudio, version 1.1.383. We examined app characteristics, behavioral content, and PEMAT-A/V scores via descriptive statistics. We performed nonparametric tests to compare groups and to investigate the correlation between the variables of interest, given their distribution properties. Associations between behavioral content and PEMAT-A/V scores were investigated via bivariate correlations (the Spearman rank correlation coefficient, ρ). Additional exploratory analyses regarding relationships between app characteristics and these content and format properties are reported in Multimedia Appendices 3-6 for interested readers.

Results

Search

Of the total 704 unique apps identified in the Google Play store, 167 (23.7%) had descriptions in languages other than French; 104 apps (14.8%) were considered as targeting chronic conditions in general and 50 apps (7.1%) focused on chronic conditions in the four groups of diseases included in this review. Other chronic conditions, not considered in this work, were back pain, migraine and headache, sleep apnea, and depression, among others. For app selection, agreement between reviewers was substantial (κ=.62). Reconciliation was done by a third reviewer when agreement was not reached after discussion between the first two. A total of 50 apps out of 704 (7.1%) met the inclusion criteria. Nonetheless, 3 of the selected apps required other connected objects (eg, connected glucometer, blood glucose sensor, or smart watch), 3 had descriptions in French but app content was entirely in English, and 2 pairs of apps had different names but the same content, so 1 app from each pair was removed; the number of excluded apps was 8. When the second round of coding of BCT content and understandability and actionability scores was performed in

October 2018 for calculation of interrater reliability and score revision, 3 apps were no longer available on the marketplace and were removed from the analysis.

The literature search yielded 1344 abstracts, and 234 manuscripts were assessed based on full text. The kappa indicating interrater reliability was .33 (ie, fair agreement), and a third reviewer did reconciliation by checking all records disagreed upon by the first two reviewers. Many of the manuscripts found through the literature search were reviews, as well as reviews of reviews (n=282). No peer-reviewed or conference papers from French institutions were selected for full-text screening. A total of 7 [28-34] out of 234 (3.0%) manuscripts selected mentioned at least one native app for Android available in France and in French; these were included in our review. Of the 7 apps, one was present in two different articles and one required a glucometer connected to the mobile phone, which prevented app use. This resulted in 5 distinct apps downloaded from the Google Play store. Apps found through the literature search were not present in the list of apps extracted previously from the Google Play store. Finally, we analyzed a list of 44 unique native apps: 5 from the literature search and 39 from the marketplace search (see Figure 1).

Sample Characteristics of Apps

A total of 44 apps were downloaded and analyzed, most of them targeting diabetes (19/44, 43%). The least-represented category was cancer (1/44, 2%). Table 1 shows the main characteristics of the apps analyzed.

Figure 1. Flowchart of the screening process.

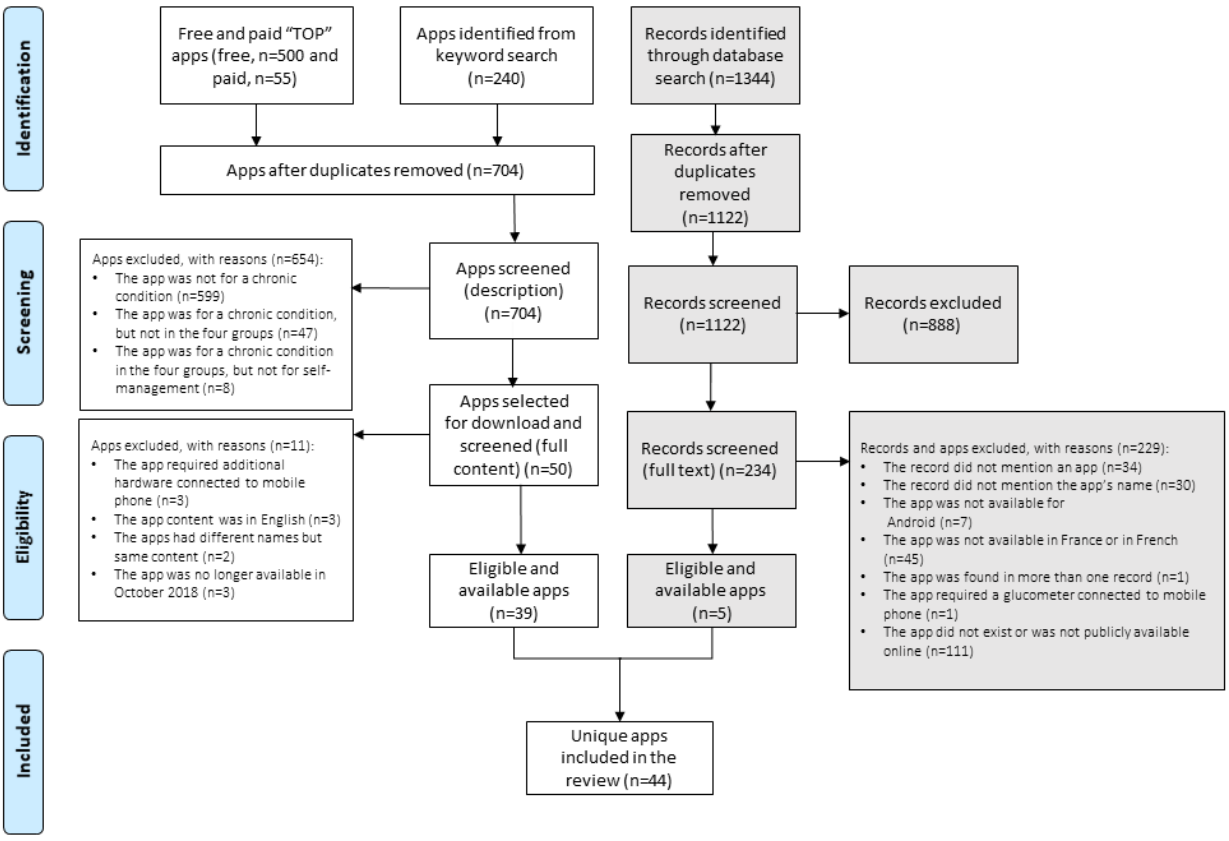


Table 1. Characteristics of apps sample.

Characteristics	Apps (N=44)
Disease category, n (%)	
Cancers	1 (2)
Respiratory diseases	6 (14)
Cardiovascular diseases	10 (23)
Diabetes	19 (43)
Other	8 (18)
Number of stars, mean (SD)	4.18 (0.48)
Number of stars, range	3-5
Number of user ratings, mean (SD)	16,140.1 (61,401.6)
Number of user ratings, range	2-374,462
Downloads, n (%)	
50+	1 (2)
100+	3 (7)
500+	4 (9)
1000+	4 (9)
5000+	4 (9)
10,000+	6 (14)
50,000+	3 (7)
100,000+	10 (23)
500,000+	3 (7)
1,000,000+	3 (7)
5,000,000+	2 (5)
10,000,000+	1 (2)
Gratuity, n (%)	
Paid app	1 (2)
With paid features	14 (32)
Without paid features	29 (66)
Developer, n (%)	
Nonprivate organization[a]	4 (9)
Pharmaceutical or medical device company	13 (30)
Private app-development company	27 (61)

[a]Nongovernmental organizations, public institutions, or European projects.

Target Behavior and Behavior Change Technique Characteristics

We identified the presence of 10 target behaviors and 21 BCTs in our app sample (see Table 2). The maximum number of target behaviors observed per app was 7 in 2 apps; the median number was 2. A total of 5 apps did not present any target behaviors. A total of 4 apps had no BCTs present and another 4 had only 1 BCT, while only 1 app had more than 10 (n=12). The median number of BCTs per app was 3 (range 0-12).

The kappa for interrater reliability in BCT coding was .68 (ie, substantial agreement); the reconciliation process was used to revise coding in the whole app sample.

Table 2. Prevalence of target behaviors and behavior change techniques (BCTs) in the app sample.

Target behaviors and BCTs	Occurrence in apps (N=44), n (%)
Target behavior	
Tracking symptoms	26 (59)
Medication adherence	13 (30)
Tracking diet	12 (27)
Tracking weight	11 (25)
Archiving health information	9 (20)
Physical activity	6 (14)
Attending medical appointments	3 (7)
Tracking emotional symptoms	3 (7)
Drinking water	2 (5)
Tracking sleep	2 (5)
BCT	
Self-monitoring of outcome(s) of behavior	31 (70)
Feedback on outcome(s) of behavior	25 (57)
Self-monitoring of behavior	19 (43)
Prompts and cues	17 (39)
Information about health consequences	14 (32)
Goal setting (outcome)	11 (25)
Graded tasks	5 (11)
Action planning	4 (9)
Biofeedback	3 (7)
Feedback on behavior	3 (7)
Social support (practical)	2 (5)
Social comparison	2 (5)
Instruction on how to perform the behavior	1 (2)
Demonstration of the behavior	1 (2)
Credible source	1 (2)
Monitoring of emotional consequences	1 (2)
Social reward	1 (2)
Goal setting (behavior)	1 (2)
Social support (unspecified)	1 (2)
Social support (emotional)	1 (2)

The most common target behaviors were *tracking symptoms* (eg, in the case of apps for hypertension, these included measuring blood pressure and recording the values in the app journal) (26/44, 59%); *medication adherence* (eg, recording medication name and dosage and setting alarms to remember taking them) (13/44, 30%); *tracking diet* (eg, in the case of diabetes apps, these included noting food quantities in an app journal) (12/44, 27%); *tracking weight* (11/44, 25%); and *archiving health information* (eg, recording clinical test results in an app journal) (9/44, 20%). For BCTs, the most common were *self-monitoring of outcome(s) of behavior* (31/44, 70%), followed by *feedback on outcome(s) of behavior* (25/44, 57%), *self-monitoring of behavior* (19/44, 43%), *prompts and cues* (17/44, 39%), *information about health consequences* (14/44, 32%), and *goal setting (outcome)* (11/44, 25%). All target behaviors and BCTs mentioned above were encountered in more than 20% of analyzed apps.

Figure 2 shows examples of target behaviors and BCTs. The left-hand screenshot shows a blood glucose journal, corresponding to the target behavior *tracking symptoms* and the BCT *self-monitoring of outcome(s) of behavior*. The middle screenshot shows a graph with blood glucose level variation through a period of one week along with the blood glucose target range defined by the user and his or her health care

provider, corresponding to the BCTs *feedback on outcome(s) of behavior* and *goal setting (outcome)*. The right-hand screenshot shows a food journal corresponding to the target behavior *tracking diet* and the BCT *self-monitoring of behavior*. This specific app is designed for people with type 1 diabetes and knowing the amount of carbohydrates in meals is essential for adjusting insulin dosage.

Understandability and Actionability Scores

Out of the 44 apps, 2 (5%) had an actionability score of 100%. The mean understandability score was 43.50% (SD 22.24), the median was 42% (interquartile range [IQR] 28), and the values ranged from 9% to 92%. For actionability, the mean score was 23.50% (SD 36.86), the median was 0% (IQR 50), and values ranged from 0% to 100%. A total of 30 apps out of 44 (68%) had null actionability (ie, they had no clearly stated actions the user could take regarding the self-management behaviors the app targeted). Figure 3 shows the co-occurrence of understandability and actionability scores in the sample.

The kappa for interrater reliability for understandability scores was .65 (ie, substantial agreement) and for actionability scores was .02 (ie, poor agreement). Both coders rated actionability as low for most apps. Differences were mostly related to an interpretation ambiguity in the first item of the actionability assessment—"The material clearly identifies at least one action the user can take"—upon which all three other actionability items were dependent. The fact that there were only four items to evaluate also influenced the different scores. Disagreements were discussed and the process of reconciliation led to revising the scores for the other apps in the sample.

Understandability and actionability scores were positively correlated (ρ=.67, $P<.001$) and so were the number of BCTs and target behaviors per app (ρ=.62, $P<.001$). Understandability had a positive correlation to the number of BCTs per app (ρ=.35, $P=.02$) and number of target behaviors per app (ρ=.31, $P=.04$). This may suggest that apps with more target behaviors and BCTs also tended to present this content in a way that is easier to understand. Actionability had moderate positive correlations to target behaviors per app (ρ=.24, $P=.12$) and BCTs per app (ρ=.29, $P=.054$), which were not statistically significant.

Figure 2. Screenshots from the Gluci-Check app from the Google Play store. The screenshots show examples of target behaviors and behavior change techniques (BCTs) used in apps. The left-hand screenshot corresponds to the target behavior *tracking symptoms* and the BCT *self-monitoring of outcome(s) of behavior*; the middle screenshot corresponds to the BCTs *feedback on outcome(s) of behavior* and *goal setting (outcome)*; the right-hand screenshot corresponds to the target behavior *tracking diet* and the BCT *self-monitoring of behavior*.

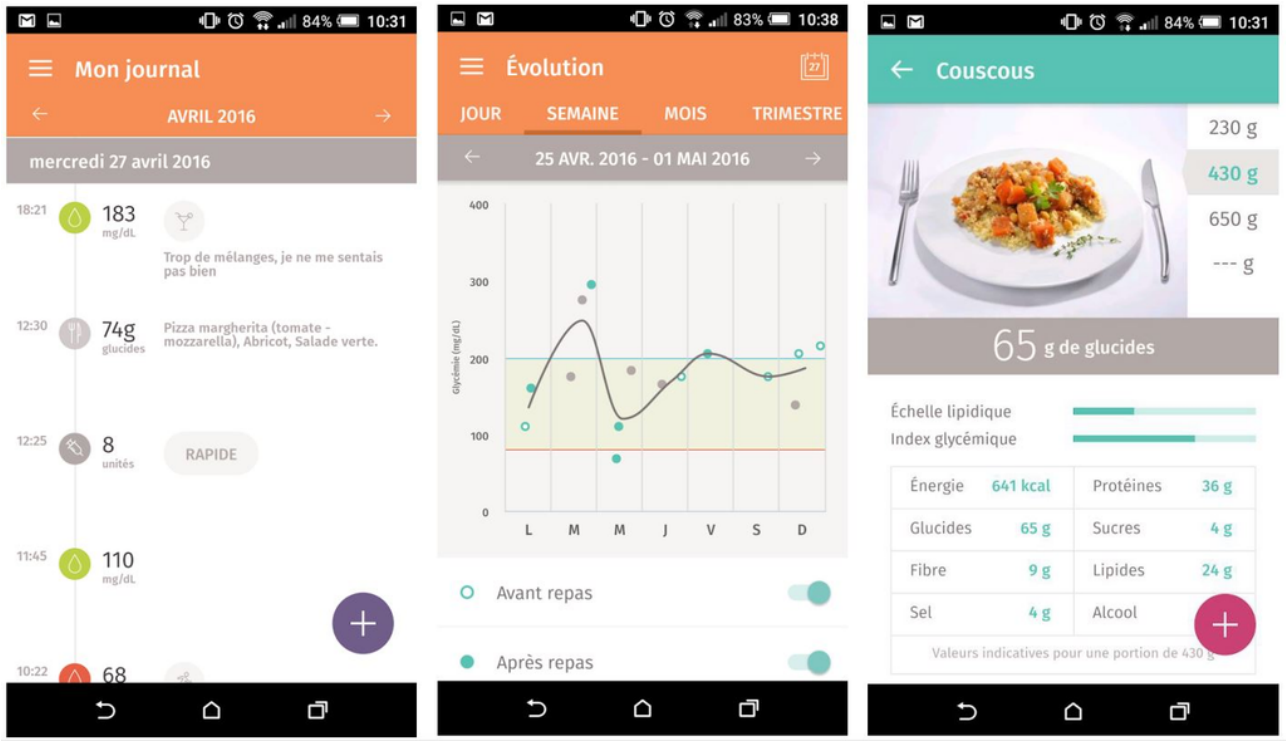
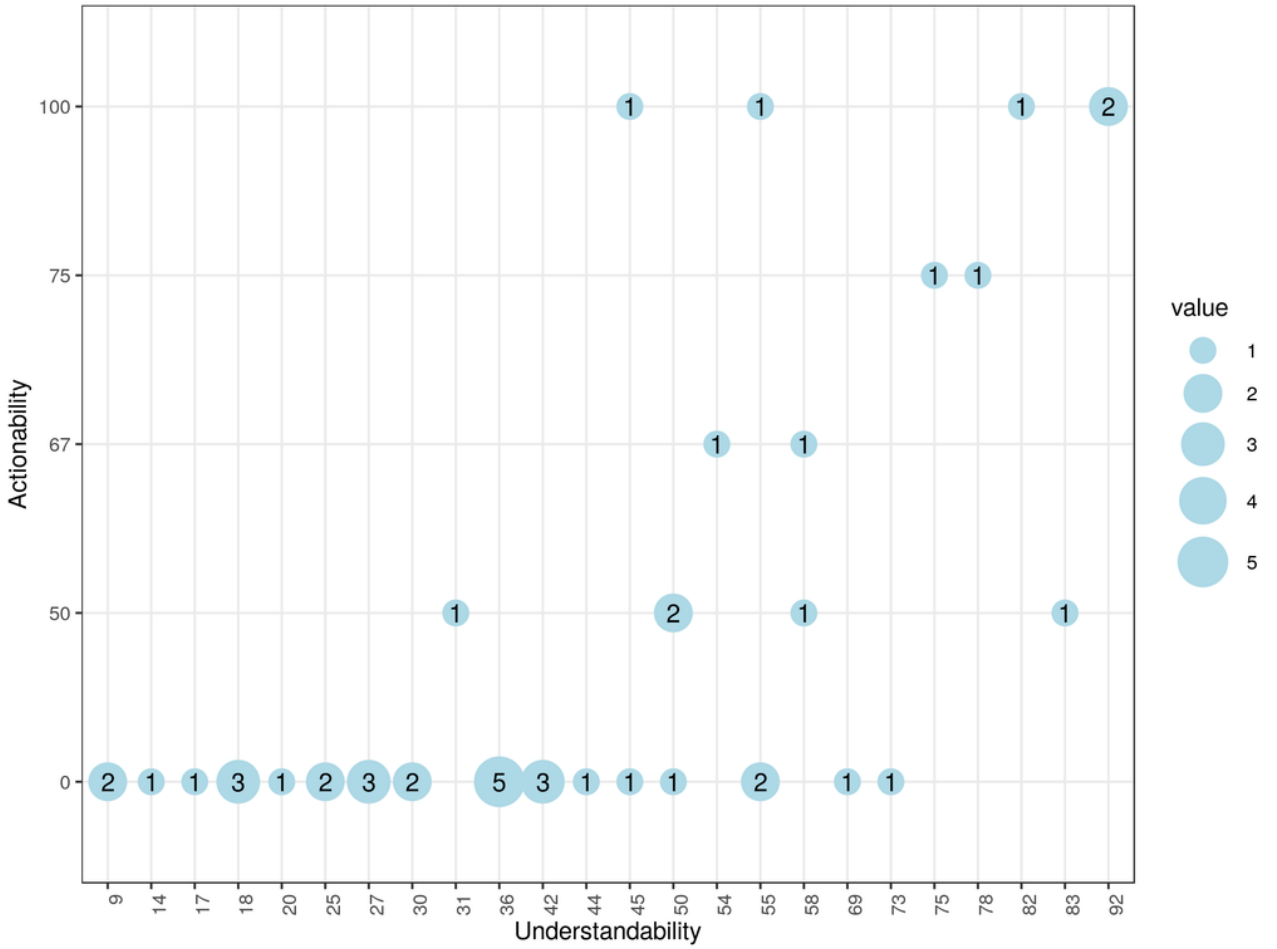


Figure 3. Understandability versus actionability; the circle size and label indicate the number of apps with the corresponding two scores.



Discussion

Principal Findings

The use of mHealth apps for supporting health-related behavior change and patient empowerment is being widely discussed as a solution to health care challenges worldwide, particularly for chronic conditions. This review showed that in 2018 in France, the potential of mHealth is far from being achieved. The apps available for download on Google Play had relatively limited behavior change content and, although moderately easy to understand for diverse audiences, they did not commonly point to clear actions users may take to self-manage the condition targeted. To better support patients with chronic conditions, apps can be improved by building on more solid behavior change research and applying it in ways that are easier to understand and act upon.

By searching directly on Google Play, among 704 apps with different purposes and languages, we found only 39 apps available in French targeting self-management in the four groups of chronic conditions with the highest mortality rates—cardiovascular diseases, cancers, respiratory diseases, and diabetes. This suggests that, for a patient or health care professional interested in using apps to manage a chronic condition, finding an appropriate app can be challenging. Moreover, our literature search identified only 5 apps available in France and in French. In another review with similar methodology performed in Canada with mHealth apps in English, of the total of 107 apps for depression analyzed, 48 were found through a literature search [24]. This highlights the variability of mHealth apps offerings worldwide, in terms of both availability and research, and the importance of studying the evolution of this domain in different countries and languages.

Targeting specific key behaviors in interventions and combining active components (ie, BCTs) that are potentially effective for behavior change in the context of chronic conditions is crucial to achieve the intended change and, consequently, improve and maintain quality of life [35,36]. In our sample, the most common target behavior was *Tracking symptoms*, and an important prevention behavior, *Physical activity*, was present in only 6 apps. The median number of target behaviors per app was 2, ranging from 0 to 5. The modifiable behaviors described as priorities by the World Health Organization as risk factors increasing mortality rates, such as tobacco use, alcohol consumption, and excessive salt intake, were not addressed in any of the analyzed apps. There was substantial variation in the number of BCTs present and the majority of the apps focused on self-monitoring, confirming the focus on monitoring behaviors as previously shown in the literature. We observed 20 BCTs in our sample and the median number of BCTs per app was 3, ranging from 0 to 12. Conroy et al [22] observed 26 BCTs in total and the most top-ranked apps for physical activity incorporated less than 4 BCTs, ranging from 1 to 13. Direito et al [23] found 26 BCTs in total, with an average of 8.1 BCTs

per app, ranging from 2 to 18; free apps, such as most of the apps evaluated in our study, presented a slightly smaller average (ie, 6.6, ranging from 3 to 14). In both studies, *provide instruction* was the most common BCT, which was not observed in the case of chronic condition mHealth apps in our study. Similar to the results of Martinez-Pérez et al [37] regarding mHealth apps for the most prevalent conditions by the World Health Organization, we found more assistive and monitoring characteristics in the apps in our sample than informative and educational ones, also reinforced by the most common target behavior *tracking symptoms*. *Demonstration of the behavior* was an uncommon BCT, and videos or illustrations to clarify the use of measurement equipment, such as blood glucose or pressure monitors, were rare. Furthermore, only a third of apps presented *information about health consequences*, which is indispensable for understanding complications related to the chronic conditions discussed here. *Goal setting (outcome)* was also frequent (>60%) in physical activity apps [22], while in our work, it was present in less than one-fourth of apps. Goal management BCTs (ie, *goal setting* and *goal review*) were found to be effective in physical activity and dietary behavior change interventions [38-41]. Also, the combination of self-monitoring techniques with at least one other self-regulation technique (ie, intention formation, feedback on performance, specific goal setting, and review of behavioral goals) is shown to be more effective than other interventions [40]. Moreover, *action planning*, which is highly related to actionability and overcoming emergencies, was present in only 5 apps. We were thus able to identify limited behavioral change content and a focus on monitoring rather than goal management or education.

The apps in our study were to a large extent not suitable for low literacy audiences. The median understandability and actionability scores were 42% and 0%, respectively. In a previous study that applied the PEMAT-A/V to 43 apps intended for parent education (ie, parenting, child health, or infant health) [42], 30 apps had understandability scores between 76% and 100%, while for actionability, 19 apps had scores in this range. We found most apps had hard-to-read text (ie, small font and too much text). In previous work with a different methodology, Meppelink et al [43] showed that almost 80% of Dutch health information websites were over the recommended B1 reading level: B1 reading level means 95% of the population can understand the information. In addition, the predominantly low actionability of the apps in our study shows that we are still far from fulfilling the potential of mHealth tools to increase patient autonomy. More than half of the analyzed apps did not present any clearly stated action and they did not have any suggestions concerning the data recorded by users on health-related events. For mHealth apps to fully achieve their potential to support chronic condition treatment, clearly indicating actions is imperative (eg, diabetes apps need to indicate that patients need to intervene immediately if high or low blood glucose is recorded and give concrete physical activity suggestions to users). In our study, apps with more target behaviors and BCTs were also more understandable, indicating that developers who consider behavioral content may also be more careful with making sure apps are comprehensible for users; levels of actionability were low irrespective of behavioral content. We therefore highlight actionability as a priority to address in app development: stating actions users can take, addressing users directly when describing actions, presenting actions in short explicit steps, and explaining how to use data visualization to take action [20].

Strengths and Limitations

First, our study used a three-pronged search strategy to identify apps relevant for our research questions: two strategies likely employed by users to identify apps in the marketplace (ie, top-ranked mHealth apps and active keyword search) and one strategy to identify apps that have been subject to scientific research (ie, a literature search). However, only the Android app marketplace was examined in this study and, although Google has the largest portion of the mobile app market and most apps are present in both marketplaces, not considering the second-most popular app marketplace (ie, Apple App Store) can lead to omission of relevant apps. Nonetheless, we believe our search strategies enabled us to obtain a representative sample to describe the current state-of-the-art in mHealth self-management support. Second, we only considered peer-reviewed papers and conference articles published in English, even though we were looking for apps available in France and in French. A future study of the iOS marketplace and French databases may be useful to complement our findings. We have also downloaded and assessed both behavioral content as well as understandability and actionability by two independent coders interacting directly with the apps, not only the descriptions available in the commercial marketplace. Thus, we were able to obtain a comprehensive assessment of the properties examined and reflect also on the assessment tools used.

We identified several issues for further improvement. First, while the BCT taxonomy enables systematic coding with good intercoder reliability, there is no consensus to date on classifying target behaviors apart from broad domains [44]. We have followed commonly used terminology to describe target behaviors in this study, yet our descriptions would have certainly benefited from standardized labels. Target behavior definition and selection is a key step in behavior change [45] and working toward a consensus on target behavior classification would further facilitate evidence synthesis. Second, since PEMAT-A/V was developed for educational materials using audio and video resources, we encountered a few difficulties when applying its criteria to apps. For example, app names and descriptions are commonly less informative in apps than what is expected for other health-related educational materials, names may be unrelated to the condition, and descriptions do not necessarily contain all app features. These characteristics may be interpreted as low understandability but may also be due to different design conventions in apps, which may have to be considered as an underestimation of understandability in our sample. PEMAT-A/V was selected after careful review of several tools, as it was considered best suited to app assessment by the research team. However, we would support a future adaptation of PEMAT-A/V for apps, which could aim to reconcile the usual brevity of the app medium with the requirements of effective communication for different audiences. Third, our study focused on the content and format of apps and excluded other criteria for judging app quality, from user engagement

and functionality [46] to data security and ethical and legal standards [47]. A comprehensive evaluation was beyond the scope of our review and would need to consider multiple dimensions.

Conclusions

Our findings suggest that mHealth apps available in France could be improved in terms of content and format. They also illustrate how two readily available tools—the BCT taxonomy and the PEMAT-A/V—can provide useful insights into the potential of an app to support patient empowerment. These tools can be used by different stakeholders in app development or to assess the existing offerings to ensure an effective contribution of apps to patient care; we would recommend their inclusion in broader app development and evaluation guidelines. Given the prevalence of the chronic conditions considered here, it is essential to make sure different levels of health literacy are considered when developing health-related materials. Also, the development of mHealth apps should involve users and consider their behavioral support needs and be accompanied by research on whether their content and use are able to effectively change behavior. Apps could also benefit from integrating more instructions for intended users on actions to perform, more modifiable behavioral risk factors, and more behavior change content, especially BCTs associated with increased effectiveness in modifying target behaviors.

Acknowledgments

LSdP was supported by a master's internship grant from the Chaire Prevent'Horizon and PhD funding (2018-2021) from an IDEXLYON Fellowship grant (grant number: 16-IDEX-0005). AD was supported by a Marie Curie Individual Fellowship from the European Commission (fellowship number: MCRA-IF n°706028) during the conduct of this review and manuscript preparation.We thank Thibaud Socquet-Clerc for his help with app selection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 67KB - mhealth_v7i8e13494_app1.pdf]

Multimedia Appendix 2
Search strategy in the Google Play store and in PubMed (ie, MEDLINE), IEEE, and Web of Science databases.
[PDF File (Adobe PDF File), 44KB - mhealth_v7i8e13494_app2.pdf]

Multimedia Appendix 3
App sample with data extracted from the Google Play store.
[XLSX File (Microsoft Excel File), 12KB - mhealth_v7i8e13494_app3.xlsx]

Multimedia Appendix 4
Dataset used to perform analyses.
[XLSX File (Microsoft Excel File), 17KB - mhealth_v7i8e13494_app4.xlsx]

Multimedia Appendix 5
R markdown script.
[TXT File, 32KB - mhealth_v7i8e13494_app5.txt]

Multimedia Appendix 6
R markdown report with descriptive statistics and inferential statistical and exploratory analyses.
[PDF File (Adobe PDF File), 235KB - mhealth_v7i8e13494_app6.pdf]

References

1. Institute for Health Metrics and Evaluation. Global Burden of Disease (GBD) URL: http://www.healthdata.org/gbd [accessed 2019-01-17]
2. World Health Organization. 2018 Jun 01. Noncommunicable diseases URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2019-01-22] [WebCite Cache ID 75cir0jNf]
3. Division of Health Promotion, Education and Communications (HPR), Health Education and Health Promotion Unit (HEP). Health Promotion Glossary. Geneva, Switzerland: World Health Organization; 1998. URL: https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf [accessed 2019-08-18]

4. Klasnja P, Pratt W. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
5. ICT Data and Statistics Division within the Telecommunication Development Bureau of ITU. Measuring the Information Society Report 2017. Volume 1. Geneva, Switzerland: International Telecommunication Union; 2017. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2017/MISR2017_Volume1.pdf [accessed 2019-08-18]
6. Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2012 Nov 08. Mobile health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2019-01-17] [WebCite Cache ID 75cjEvVn8]
7. Krebs P, Duncan DT. Health app use among US mobile phone owners: A national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
8. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: Analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
9. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: Qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
10. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: Systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
11. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: Systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
12. Breland JY, Yeh VM, Yu J. Adherence to evidence-based guidelines among diabetes self-management apps. Transl Behav Med 2013 Sep;3(3):277-286 [FREE Full text] [doi: 10.1007/s13142-013-0205-4] [Medline: 24073179]
13. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
14. Hardeman W, Sutton S, Griffin S, Johnston M, White A, Wareham NJ, et al. A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation. Health Educ Res 2005 Dec;20(6):676-687. [doi: 10.1093/her/cyh022] [Medline: 15781446]
15. 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]
16. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
17. Webb T, 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]
18. de Korte E, Wiezer N, Bakhuys Roozeboom M, Vink P, Kraaij W. Behavior change techniques in mHealth apps for the mental and physical health of employees: Systematic assessment. JMIR Mhealth Uhealth 2018 Oct 03;6(10):e167 [FREE Full text] [doi: 10.2196/mhealth.6363] [Medline: 30282621]
19. Kreps GL. The relevance of health literacy to mHealth. Stud Health Technol Inform 2017;240:347-355. [Medline: 28972527]
20. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403 [FREE Full text] [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
21. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: A systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
22. 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]
23. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
24. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
25. Ministère des Solidarités et de la Santé. 2017 Dec 20. La stratégie nationale de santé 2018-2022 URL: https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/strategie-nationale-de-sante/article/la-strategie-nationale-de-sante-2018-2022 [accessed 2019-01-08]

26. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
27. Bhattacharya A. Quartz. 2016 Nov 03. Android just hit a record 88% market share of all smartphones URL: https://qz.com/826672/android-goog-just-hit-a-record-88-market-share-of-all-smartphones/ [accessed 2019-01-22] [WebCite Cache ID 75cjKkaJ3]
28. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: The SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
29. Kitsiou S, Thomas M, Marai G, Maglaveras N, Kondos G, Arena R, et al. Development of an innovative mHealth platform for remote physical activity monitoring and health coaching of cardiac rehabilitation patients. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17); February 16-19, 2017; Orlando, FL p. 133. [doi: 10.1109/BHI.2017.7897223]
30. Kumar S, Moseson H, Uppal J, Juusola JL. A diabetes mobile app with in-app coaching from a certified diabetes educator reduces A1C for individuals with type 2 diabetes. Diabetes Educ 2018 Jun;44(3):226-236. [doi: 10.1177/0145721718765650] [Medline: 29575982]
31. Mora P, Buskirk A, Lyden M, Parkin CG, Borsa L, Petersen B. Use of a novel, remotely connected diabetes management system is associated with increased treatment satisfaction, reduced diabetes distress, and improved glycemic control in individuals with insulin-treated diabetes: First results from the Personal Diabetes Management study. Diabetes Technol Ther 2017 Dec;19(12):715-722 [FREE Full text] [doi: 10.1089/dia.2017.0206] [Medline: 29027812]
32. Morawski K, Ghazinouri R, Krumme A, McDonough J, Durfee E, Oley L, et al. Rationale and design of the Medication adherence Improvement Support App For Engagement-Blood Pressure (MedISAFE-BP) trial. Am Heart J 2017 Apr;186:40-47. [doi: 10.1016/j.ahj.2016.11.007] [Medline: 28454831]
33. Osborn CY, van Ginkel JR, Marrero DG, Rodbard D, Huddleston B, Dachis J. One Drop | Mobile on iPhone and Apple Watch: An evaluation of HbA1c improvement associated with tracking self-care. JMIR Mhealth Uhealth 2017 Nov 29;5(11):e179 [FREE Full text] [doi: 10.2196/mhealth.8781] [Medline: 29187344]
34. Knight B, McIntyre H, Hickman I, Noud M. Qualitative assessment of user experiences of a novel smart phone application designed to support flexible intensive insulin therapy in type 1 diabetes. BMC Med Inform Decis Mak 2016 Sep 15;16:119 [FREE Full text] [doi: 10.1186/s12911-016-0356-6] [Medline: 27629774]
35. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: A scoping review of methods used. Transl Behav Med 2018 Mar 01;8(2):212-224. [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
36. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
37. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: Review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
38. Hankonen N, Sutton S, Prevost AT, Simmons RK, Griffin SJ, Kinmonth AL, et al. Which behavior change techniques are associated with changes in physical activity, diet and body mass index in people with recently diagnosed diabetes? Ann Behav Med 2015 Feb;49(1):7-17 [FREE Full text] [doi: 10.1007/s12160-014-9624-9] [Medline: 24806469]
39. Avery L, Flynn D, van Wersch A, Sniehotta FF, Trenell MI. Changing physical activity behavior in type 2 diabetes: A systematic review and meta-analysis of behavioral interventions. Diabetes Care 2012 Dec;35(12):2681-2689 [FREE Full text] [doi: 10.2337/dc11-2452] [Medline: 23173137]
40. 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]
41. 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]
42. Davis DW, Logsdon MC, Vogt K, Rushton J, Myers J, Lauf A, et al. Parent education is changing: A review of smartphone apps. MCN Am J Matern Child Nurs 2017;42(5):248-256. [doi: 10.1097/NMC.0000000000000353] [Medline: 28639997]
43. Meppelink CS, van Weert JCM, Brosius A, Smit EG. Dutch health websites and their ability to inform people with low health literacy. Patient Educ Couns 2017 Nov;100(11):2012-2019. [doi: 10.1016/j.pec.2017.06.012] [Medline: 28624261]
44. 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]

45. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
46. Stoyanov S, Hides L, Kavanagh D, 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]
47. Dimitropoulos L, Patel V, Scheffler SA, Posnack S. Public attitudes toward health information exchange: Perceived benefits and concerns. Am J Manag Care 2011 Dec;17(12 Spec No):SP111-SP116 [FREE Full text] [Medline: 22216769]

Abbreviations

BCT: behavior change technique
IQR: interquartile range
mHealth: mobile health
PEMAT-A/V: Patient Education Material Assessment Tool for audiovisual materials
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols
PROSPERO: International Prospective Register of Systematic Reviews

Edited by G Eysenbach; submitted 28.01.19; peer-reviewed by MS Aslam, K Stasiak, B Price; comments to author 04.04.19; revised version received 14.05.19; accepted 21.05.19; published 26.08.19.

Please cite as:
Siqueira do Prado L, Carpentier C, Preau M, Schott AM, 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;7(8):e13494
URL: http://mhealth.jmir.org/2019/8/e13494/
doi:10.2196/13494
PMID:31452522

Original Paper

A Comparison of Functional Features in Chinese and US Mobile Apps for Diabetes Self-Management: A Systematic Search in App Stores and Content Analysis

Yuan Wu[1,2*], MD; Yiling Zhou[1,3*], MRes; Xuan Wang[3,4], PhD; Qi Zhang[5], MB; Xun Yao[6], MD; Xiaodan Li[7,8], MD; Jianshu Li[9], PhD; Haoming Tian[1], MM; Sheyu Li[1,3], MD

[1]Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, Chengdu, China
[2]Xi An Jiao Tong University, The First Affiliated Hospital of Xi An Jiao Tong University, Xi An, China
[3]Department of Population Health & Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom
[4]Science for Life Laboratory, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden
[5]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China
[6]Department of Academic Affairs, West China School of Medicine, Sichuan University, Chengdu, China
[7]Department of Gastroenterology, West China Hospital, Sichuan University, Chengdu, China
[8]Medicines Monitoring Unit, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom
[9]Department of Biomedical Polymer and Artificial Organs, College of Polymer Science and Engineering, Sichuan University, Chengdu, China
[*]these authors contributed equally

Corresponding Author:
Sheyu Li, MD
Department of Endocrinology and Metabolism
West China Hospital
Sichuan University
37# Guoxue Road
Wuhou District
Chengdu, 610041
China
Phone: 86 13194874843
Fax: 86 2885422982
Email: lisheyu@gmail.com

Abstract

Background: Mobile health interventions are widely used for self-management of diabetes, which is one of the most burdensome noncommunicable chronic diseases worldwide. However, little is known about the distribution of characteristics and functions of in-store mobile apps for diabetes.

Objective: This study aimed to investigate the distribution of characteristics and functions of the in-store mobile apps for self-management of diabetes in the United States and China using a predefined functional taxonomy, which was developed and published in our previous study.

Methods: We identified apps by searching diabetes in English or Chinese in the Apple iTunes Store and Android Markets (both in the United States and China) and included apps for diabetes self-management. We examined the validity and reliability of the predefined functional taxonomy with 3 dimensions: clinical module, functional module, and potential risk. We then classified all functions in the included apps according to the predefined taxonomy and compared the differences in the features of these apps between the United States and China.

Results: We included 171 mobile diabetes apps, with 133 from the United States and 38 from China. Apps from both countries faced the challenges of evidence-based information, proper risk assessment, and declaration, especially Chinese apps. More Chinese apps provide app-based communication functions (general communication: Chinese vs US apps, 39%, 15/38 vs 18.0%, 24/133; *P*=.006 and patient-clinician communication: Chinese vs US apps, 68%, 26/38 vs 6.0%, 8/133; *P*<.001), whereas more US apps provide the decision-making module (Chinese vs US apps, 0%, 0/38 vs 23.3%, 31/133; *P*=.001), which is a high-risk

module. Both complication prevention (Chinese vs US apps, 8%, 3/38 vs 3.8%, 5/133; *P*=.50) and psychological care (Chinese vs US apps, 0%, 0/38 vs 0.8%, 1/133; *P*>.99) are neglected by the 2 countries.

Conclusions: The distribution of characteristics and functions of in-store mobile apps for diabetes self-management in the United States was different from China. The design of in-store diabetes apps needs to be monitored closely.

(JMIR Mhealth Uhealth 2019;7(8):e13971) doi:10.2196/13971

KEYWORDS

diabetes mellitus; self-management; mobile apps; risk assessment; prevalence; China; United States

Introduction

Background

Diabetes mellitus is one of the most burdensome noncommunicable chronic diseases in China and the western world [1-3]. It represented 8.4% of all-cause deaths worldwide among adults aged 20 to 79 years in 2013 [4] and incurred a cost of US $1.31 trillion in 2015 [5].

Diabetes requires long-term care that is highly individualized to suit the needs of each patient [6]. The care includes lifestyle modification (eg, diet, physical activity, and weight management), glucose monitoring, prevention of complications, and multiple medication regimens [7,8]. Barriers to the effectiveness of diabetes care include insufficient knowledge and training that patients received [9], requiring permanent behavior change [10], financial burden [11], and complication-specific treatments [12]. Patient education and support by a multidisciplinary team of professionals (eg, physicians, nurses, dietitians, and psychotherapists) may help improve the quality of care [13], although they could be costly and unavailable in some developing countries [14].

Mobile apps developed for diabetes self-management are new and become more and more popular [15-19]. Although randomized trials and systematic reviews suggest that mobile apps, in general, are effective in improving the glucose control for patients with type 2 diabetes [20], the apps available in app stores are highly varied in function, design, and quality [21] and have not been necessarily assessed rigorously by effective randomized control trials. Our previous study [8] developed a taxonomy with 3 dimensions (ie, clinical module, functional module, and potential risk) to provide a functional classification and risk assessment for mobile apps for diabetes self-management and suggested that different functions and combinations may contribute differently to the effectiveness of diabetes management. The clinical module of this taxonomy consists of monitoring, medication management, lifestyle modification, complication prevention, and psychosocial care; the functional module includes logs, structured display, general education, personalized feedback, and communication; potential risk includes 3 levels: high, medium, and low risk. The taxonomy serves as a novel and reliable tool to classify and evaluate the content and functions of diabetes self-management apps in the market.

Objectives

This study aimed to describe and compare the characteristics and functions of mobile apps for diabetes self-management in 2 of the largest app markets, China and the United States, to provide suggestions for future development and usage of mobile apps for diabetes self-management using our predefined taxonomy [8].

Methods

Data Source

In an electronic search, conducted on December 5, 2016, using *diabetes* as the keyword in both English and Chinese languages, we identified apps in English or Chinese from Apple iTunes store (China and the United States), Google Android Play (the United States), Tencent Android Market (Tencent Holdings Limited), Baidu Android Market (Baidu, Inc), and 360 Android Market (Qihoo 360 Technology Co Ltd).

Selection Criteria

We included mobile apps for diabetes self-management, which was defined as supporting the interactive self-monitoring of blood glucose. The first 500 mobile apps in the initial search list were included in this study. This is because users tend to choose the top mobile apps that are sorted based on customer reviews and download count. Mobile apps outside the range are less likely to be selected and downloaded by users.

The exclusion criteria were (1) duplicated apps (apps with same name and same producer are defined as duplication regardless of different versions), (2) apps without any meaningful introduction or instruction in the app store, (3) apps designed only for the health professionals, (4) apps without Chinese or English interfaces, and (5) apps with no update since January 1, 2014.

App Selection and Data Extraction

In total, 2 investigators (YW and QZ) independently selected the apps for inclusion according to the selection criteria. The investigators extracted the following data from each included study: the app name, developer, specifications (medical, health, and fitness or unavailable), acquisition cost (free, paid, or in-app purchase), downloading fee, the latest update date, target users (type 1 diabetes, type 2 diabetes, gestational diabetes, prediabetes, all types, or unavailable), safety statement (potential risks or use under guidance), supporting evidence (descriptive study, observational study, or randomized controlled trials), and source of information (clinical guidelines). The inconsistency of the study inclusion and data extraction was solved by a discussion between the 2 investigators.

Validity and Reliability of the Developed Taxonomy

The list of function was summarized according to our previously developed taxonomy, which was described in detail in previous publication [8]. First, the preliminary taxonomy was coded [8]. Then 2 reviewers independently classified functions of all included apps according to the coded preliminary taxonomy and slightly modified the preliminary taxonomy. Furthermore, 2 new functions, namely recording insulin injection and reminder to record the medication, were added to the slot of recording used medication and side effect and to the slot of reminder to take the medication of the taxonomy, respectively (Multimedia Appendix 1). The function of off-target alert and setting targets were moved to the complication prevention slot because it is more than a monitoring function but a strategy for complication prevention. Diabetes process and treatment options were removed from the slot of using medications safely and effectively. This process guaranteed that our predefined taxonomy could be used to classify all current in-store diabetes apps. A panel of 5 reviewers used the predefined taxonomy to classify functions of 10 Chinese apps and 10 US apps, which covered all coded functions to assess the reliability of taxonomy with Krippendorff alpha.

Assessment of the App Functions

Functional features of included apps were specified by crossing of the functional and clinical module with a risk assessment for each slot [8]. The potential risk of each function was assessed using methods based on the taxonomy, which was developed according to the US Food and Drug Administration (FDA) risk-based framework [8,22]. The risk level of a mobile app was determined by the highest risk level of any of its functions.

Statistics Analysis

The baseline characteristics were summarized using SPSS (version 21), and the comparison between the Chinese and US apps was evaluated using a chi-square test or Fisher exact test with a significance level of .05 using Open Epi (version 3.01). The frequencies and percentages of the modules were calculated using SPSS (version 21). The functional differences between the Chinese and US apps were examined by a chi-square test with a significance level of .05 by Open Epi (version 3.01).

Results

Basic Characteristics

After searching in the Apple iTunes store, Google Android store, Baidu Android Store, 360 Android Store, and Tencent Android Store, 1667 apps were found. After screening, 171 apps were finally included in this study, 38 from China and 133 from the United States. The process of app selection is shown in the flowchart in Figure 1.

The characteristics of all included apps were summarized in Table 1. Among Chinese mobile apps, 21% (8/38) were not adequately categorized ($P<.001$), whereas all the 133 US mobile apps were categorized as either *health and fitness* or *medical* mobile apps. Regarding the acquisition costs, there were more free apps available in China than in the United States (92% [35/38] in China vs 75.2% [100/133] in the United States; $P=.04$). US mobile apps had more specific target audiences than Chinese mobile apps (less than 3% [1/38] in China vs 21.1% [28/133] in the United States; $P=.01$). None of the Chinese mobile apps provided a clear safety declaration or supporting evidence, whereas 14.3% (19/133) of US mobile apps had a clear safety declaration, and 2.2% (3/133) of US mobile apps supplied supporting evidence.

Figure 1. The flowchart of app selection.

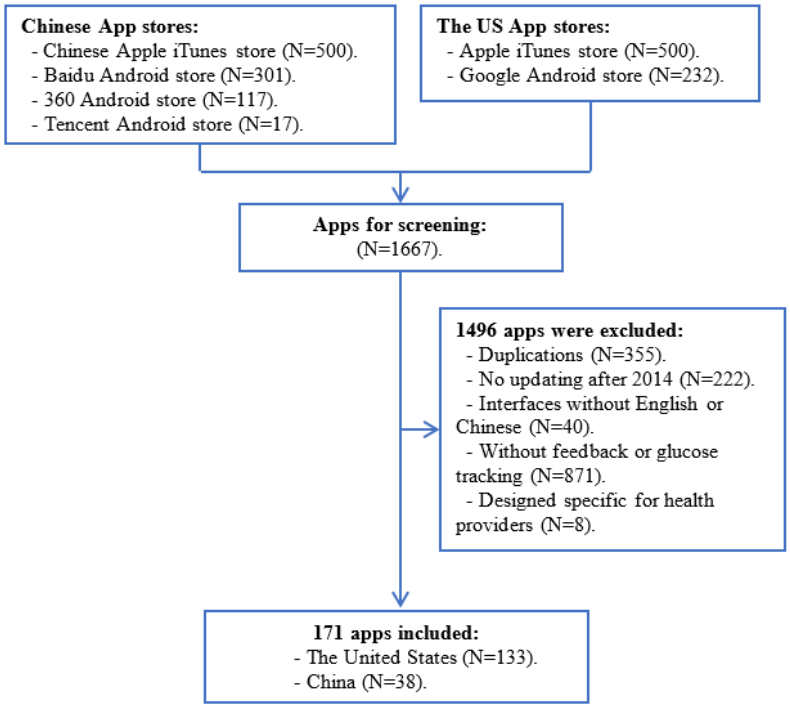


Table 1. Metadata of in-store mobile apps for diabetes self-management between the United States and China.

Category	China (N=38)	United States (N=133)	*P* value
Specifications, n (%)			
Medical	4 (11)	90 (67.7)	<.001[a]
Health fitness	26 (68)	43 (32.3)	<.001[a]
N/A[b]	8 (21)	0	<.001[a]
Acquisition costs, n (%)			
Free	35 (92)	100 (75.2)	.04[a]
In-app purchase	0	22 (16.5)	.04[a]
Paid	3 (8)	11 (8.3)	.04[a]
Statement on target users, n (%)			
With clear statement			
T1DM[c]	0	2 (1.5)	.01[a,d]
T2DM[e]	0	6 (4.5)	.01[a,d]
GDM[f]	1 (3)	1 (0.8)	.01[a,d]
All types	0	19 (14.3)	.01[a,d]
Without clear statement	37 (97)	105 (78.9)	.01[a,d]
Safety statement, n (%)			
With	0	19 (14.3)	.03[g]
Without	38	114 (85.7)	.03[g]
Supporting evidence, n (%)			
With	0	3 (2.2)	.99[g]
Without	38	130 (97.8)	.99[g]

[a]Chi-square test.

[b]N/A: not available.

[c]T1DM: type 1 diabetes.

[d]Compare apps with a clear statement on target users between China and the United States.

[e]T2DM: type 2 diabetes.

[f]GDM: gestational diabetes mellitus.

[g]Fisher exact test.

Validity and Reliability of Taxonomy

The standardized classification code was completed, which supported the validity of the taxonomy. The inconsistency of coding by each coding investigator was presented as the standardized rate of coding error (Multimedia Appendix 2). The Krippendorff alpha is .8229, which shows the reliability of the taxonomy is acceptable.

Functions and Modules

The characteristics of each function provided by the Chinese and US apps have been shown in heatmaps (Figures 2 and 3), and the statistics have been shown in Multimedia Appendix 3.

Chinese and US mobile apps had common functions, including structured display (China: 28/38, 74%; and the United States: 111/133, 83.5%); recording activities, diets, and weights (China: 24/38, 63%; and the United States: 103/133, 77.4%); and recording used medication and side effects (China: 19/38, 50%; and the United States: 83/133, 62.4%).

However, mobile apps between the 2 countries have differences in less common functions. In the 38 Chinese mobile apps, recording complication-related status and appointments with doctors (1/38, less than 3%), preventing complications (3/38, 8%), and setting targets (off-target alert, 3/38, 8%) were the least common functions. In the 133 US mobile apps, addressing psychosocial issues (1/133, 0.8%), reminder to take medications (1/133, 0.8%), and instructions for monitoring (3/133, 2.3%) were the least common functions. Furthermore, none of the 38 Chinese mobile apps included the following 7 functions: recording insulin injections site, recording mood, addressing psychosocial issues, instructions for monitoring, clinical decision

making, reminder to record medications, and reminder to quit smoking and visit doctors.

When we took a close look at the distribution of functions in these self-management apps quantitatively, we found that there were differences between apps from China and the United States. Reminder to monitor (China: 5/38, 13% and the United States: 39/133, 29.3%; *P*=.045), setting targets (China: 3/38, 8% and the United States: 39/133, 29.3%; *P*=.007), and clinical decision making (China: 0/38, 0% and the United States: 31/133, 23.3%; *P*=.001) were significantly more common functions in US mobile apps than Chinese apps. However, the following 3 functions were more commonly included in the Chinese apps: patient-clinician communication (China: 26/38, 68% and the United States: 8/133, 6.0%; *P*<.001), using medications safely and effectively (China: 17/38, 45% and the United States: 12/133, 9.0%; *P*<.001), and general communication (China: 15/38, 40% and the United States: 24/133, 18.0%; *P*=.006).

Risk Assessment of Mobile Apps for Diabetes Between the United States and China

In the United States, 23.3% (31/133) and 65.4% (87/133) of the mobile apps were assessed as high risk and medium risk, respectively. However, none of the Chinese apps was assessed as high risk (Table 2). The difference in the risk of diabetes self-management apps between the United States and China is significant (*P*=.004; Table 2).

Figure 2. Heatmap of features of 38 Chinese mobile apps for diabetic self-management.

Figure 3. Heatmap of features of 133 US mobile apps for diabetes self-management.

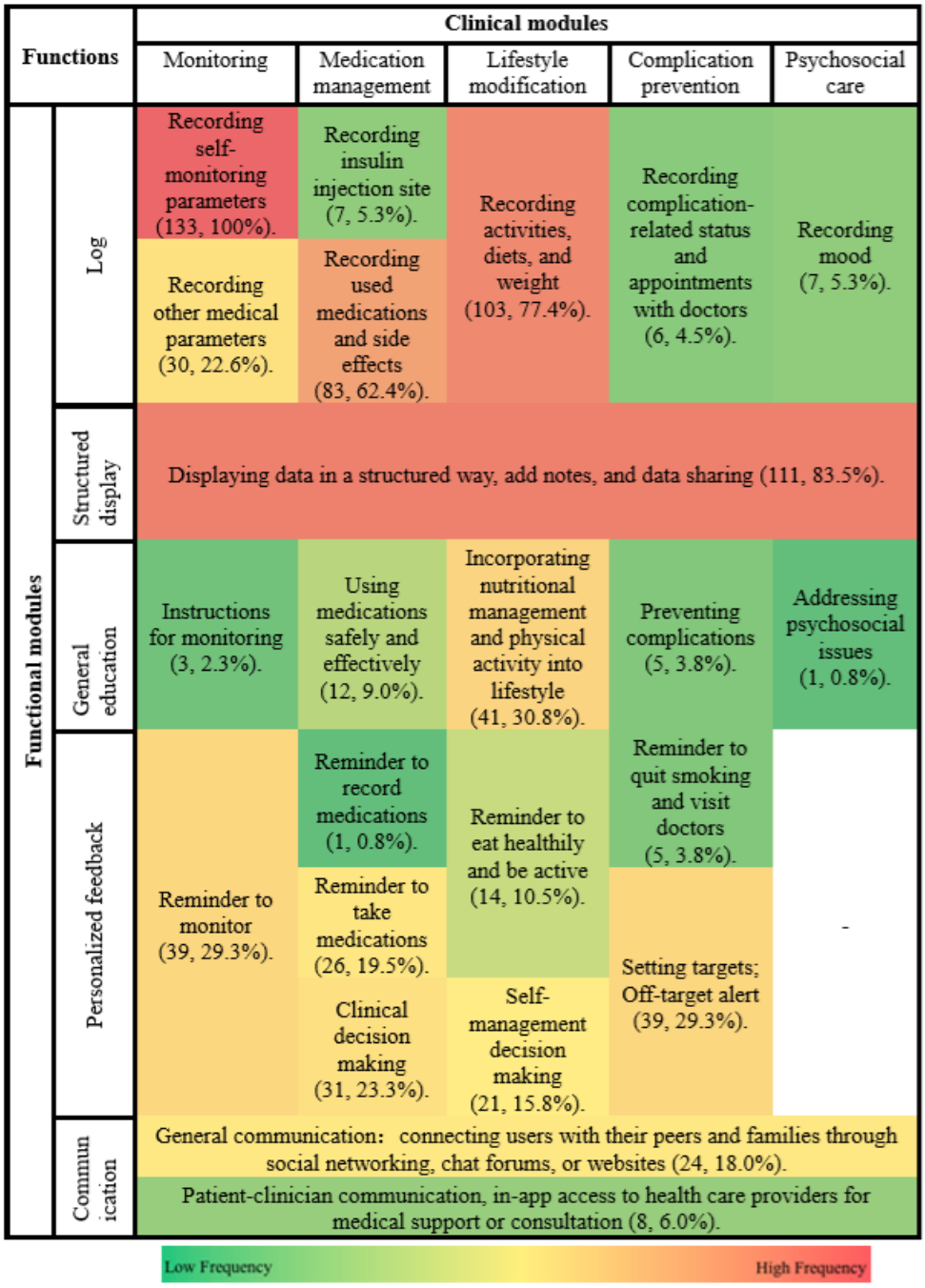


Table 2. The risk assessment of mobile apps for diabetes self-management between the United States and China.

Risk	United States (N=133), n (%)	China (N=38), n (%)	*P* value
Low	15 (11.3)	3 (8)	.004
Medium	87 (65.4)	35 (92)	.004
High	31 (23.3)	0 (0)	.004

Discussion

Principal Findings

This study systematically illustrated the characteristics and functions of in-store mobile apps for diabetes self-management in the United States and China, 2 of the largest app markets, using a predefined taxonomy. It is a useful and reliable tool to categorize functions of these apps despite the country difference [8]. What is more important is that, by using the taxonomy, the study highlights the differences between mobile apps for diabetes self-management in the 2 countries.

Metadata of Apps

All 133 US apps were categorized as either *medical* or *health and fitness* apps, whereas there are still some Chinese apps (21%, 8/38) that have not been categorized. The reason could be that the FDA risk report and policy recommendations for mobile health technologies suggest that all mobile health–related apps should be divided into 3 categories—general management, health management, and medical devices. Another reason could be that the FDA has the jurisdiction for all mobile health technologies [23]. However, there is no similar regulation in China.

According to a previous survey [19], the needs of Chinese patients with type 1 and type 2 diabetes were significantly different from each other. However, only 1 app in China stated their target users, whereas a few more in the US apps suggested that it was important for app developers to differentiate their products to meet the precise needs of different populations.

Functions and Modules

Our results show that for diabetes self-management mobile apps, monitoring, lifestyle modification, and medicine management were the 3 most common clinical modules in both China and the United States. It is not surprising that they are the most essential low-risk functions with low technical barriers. However, complication prevention and psychological care were rarely found in the apps from either country although complication prevention was supposed to be associated with more hemoglobin $A1_c$ reduction [8,24,25]. The complexity in developing these functions and modules may be the key barriers from being widely adopted.

The functions related to personalized feedback (eg, reminder to monitor, clinical decision making, and setting targets) were more common in US apps than Chinese apps. These functions are based on the built-in algorithms (usually predictive modeling using the collected personal data and probably involving advanced techniques such as artificial intelligence) and provide quick and direct solutions for users' problems. They are important functions especially when the off-line health care is inaccessible [26]. However, the clinical decision-making model was at high risk [8] and possibly underdeveloped at the moment [27]. The development of personalized feedback modules should be done carefully and with caution, whereas more attempts using different algorithms should be encouraged at the same time.

The communication modules (ie, general communication and patient-clinician communication) are more common in China than in the United States. It is in line with previous population-based survey in China [19], showing that doctor-patient communication is critical to both health providers and patients.

This study adopted our predefined taxonomy [8], which could be different from other systems, for example, Antonio Martinez-Millana et al [26] developed a taxonomy for patients with type 1 diabetes. It comprised 3 hierarchical levels with 10 areas on the first level.

One similar study focusing on Chinese diabetic mobile apps investigated the risk factors related to app use. They suggested that setting recording insulin therapy and dosage in the app might help the patients with type 1 diabetes. They emphasized the importance of the determination of target users before app development [19]. Our study found that 1 Chinese app included recording insulin injection, and only 1 app stated its target type of diabetes. The development of future mobile apps could consider these aspects to improve the effectiveness of mobile apps and user safety.

Strengths

The strengths of this study are as follows: First, this is the first survey for in-store apps using a predefined taxonomy. Second, the consistency of the predefined taxonomy was strictly tested for its reliability before use. Finally, 2 countries with diverse culture, health care systems, and economic statuses were investigated.

Limitations

This study also has limitations. First, we only screened the introduction of included in-store apps without downloading and using the apps, which may result in missing out on functional information about the mobile app. However, it is highly consistent with the strategy for a new user to choose an app in the app store. Second, we selected the apps at only one time point, which may miss the longitudinal update of app functions. Further research on monitoring the updates of mobile apps is warranted. Finally, the sample size of apps between China and the United States, accounting for less than 33% (133 vs 38), is unbalanced, which may affect the statistical significance.

Conclusions

In summary, the in-store mobile apps for diabetes self-management were different between China and the United States. The apps from both countries faced the challenges of lacking evidence-based information, proper risk assessment, and declaration, especially Chinese apps. More Chinese apps included in-app communication modules, whereas more US apps included the clinical decision-making module, which is with high risk. Both complication prevention and psychological care were neglected by the 2 countries. The design of app functions in both countries needs to be optimized, and deep interaction between the app developers and users is recommended. Appropriate surveillance is required to monitor the quality and performance of in-store apps.

Acknowledgments

The authors thank Ms Amanda Jia En Koh and Mr Brandon Size Mann Tan from the University of Dundee for reviewing the language of the manuscript. The corresponding author was supported by grants from the National Natural Science Foundation of China (grant number 81400811 and 21534008), Sichuan Science and Technology Program (No 2019YFH0150), Cholesterol

Fund by China Cardiovascular Foundation and China Heart House and the International Visiting Program for Excellent Young Scholars of Sichuan University. XL was supported by the Talents Cultivation Fund of West China Hospital.

Authors' Contributions

YW, XY, JL, HT, and SL conceived this study. YW and QZ searched the app store, screened the apps, and abstracted the characteristics and functions of the apps. YW and YZ conducted the statistical analysis. YW, YZ, XW, XL, and SL drafted the manuscript. All authors discussed and interpreted the results and reviewed the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The coded taxonomy.
[PDF File (Adobe PDF File), 76KB - mhealth_v7i8e13971_app1.pdf]

Multimedia Appendix 2
The average of standardized rate of coding error per coder.
[PDF File (Adobe PDF File), 69KB - mhealth_v7i8e13971_app2.pdf]

Multimedia Appendix 3
The comparison of characteristics of functions provided by mobile apps for self-management of diabetes between the United States and China.
[PDF File (Adobe PDF File), 60KB - mhealth_v7i8e13971_app3.pdf]

References

1. Li S, Yu C, Li Y, Li Q, Zhang R, Hou Q, et al. Study design and baseline characteristics of inpatients with diabetes mellitus in a tertiary hospital in China: a database study based on electronic medical records. J Evid Based Med 2018 Aug;11(3):152-157. [doi: 10.1111/jebm.12291] [Medline: 29512333]
2. Liu M, Liu SW, Wang LJ, Bai YM, Zeng XY, Guo HB, et al. Burden of diabetes, hyperglycaemia in China from to 2016: findings from the 1990 to 2016, global burden of disease study. Diabetes Metab 2019 Dec;45(3):286-293. [doi: 10.1016/j.diabet.2018.08.008] [Medline: 30196138]
3. Weng W, Liang Y, Kimball E, Hobbs T, Kong S. Trends in comorbidity burden and treatment patterns in type 2 diabetes: longitudinal data from a US cohort from 2006 to 2014. Diabetes Res Clin Pract 2018 Aug;142:345-352. [doi: 10.1016/j.diabres.2018.05.031] [Medline: 29802955]
4. IDF Diabetes Atlas Group. Update of mortality attributable to diabetes for the IDF Diabetes Atlas: estimates for the year 2013. Diabetes Res Clin Pract 2015 Sep;109(3):461-465. [doi: 10.1016/j.diabres.2015.05.037] [Medline: 26119773]
5. Zhang P, Gregg E. Global economic burden of diabetes and its implications. Lancet Diabetes Endocrinol 2017 Jun;5(6):404-405. [doi: 10.1016/S2213-8587(17)30100-6] [Medline: 28456417]
6. Majumder E, Cogen FR, Monaghan M. Self-management strategies in emerging adults with type 1 diabetes. J Pediatr Health Care 2017;31(1):29-36 [FREE Full text] [doi: 10.1016/j.pedhc.2016.01.003] [Medline: 26861574]
7. Weller SC, Baer R, Nash A, Perez N. Discovering successful strategies for diabetic self-management: a qualitative comparative study. BMJ Open Diabetes Res Care 2017;5(1):e000349 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000349] [Medline: 28761649]
8. 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]
9. Fritz HA. Challenges to developing diabetes self-management skills in a low-income sample in North Carolina, USA. Health Soc Care Community 2017 Dec;25(1):26-34 [FREE Full text] [doi: 10.1111/hsc.12172] [Medline: 25522673]
10. Rossi MG, Bigi S. mHealth for diabetes support: a systematic review of apps available on the Italian market. Mhealth 2017;3:16 [FREE Full text] [doi: 10.21037/mhealth.2017.04.06] [Medline: 28567412]
11. Wong CK, Jiao F, Tang EH, Tong T, Thokala P, Lam CL. Direct medical costs of diabetes mellitus in the year of mortality and year preceding the year of mortality. Diabetes Obes Metab 2018 Jun;20(6):1470-1478. [doi: 10.1111/dom.13253] [Medline: 29430799]
12. Quinn CC, Swasey KK, Crabbe JC, Shardell MD, Terrin ML, Barr EA, et al. The impact of a mobile diabetes health intervention on diabetes distress and depression among adults: secondary analysis of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2017 Dec 7;5(12):e183 [FREE Full text] [doi: 10.2196/mhealth.8910] [Medline: 29217502]

13. NICE | The National Institute for Health and Care Excellence. 2011. Diabetes in Adults URL: https://www.nice.org.uk/guidance/qs6/chapter/quality-statement-3-structured-education-programmes-for-adults-with-type-1-diabetes [accessed 2019-03-10] [WebCite Cache ID 76lohVrzt]
14. Tharkar S, Devarajan A, Kumpatla S, Viswanathan V. The socioeconomics of diabetes from a developing country: a population based cost of illness study. Diabetes Res Clin Pract 2010 Sep;89(3):334-340. [doi: 10.1016/j.diabres.2010.05.009] [Medline: 20538363]
15. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
16. Tack CJ, Lancee GJ, Heeren B, Engelen LJ, Hendriks S, Zimmerman L, et al. Glucose control, disease burden, and educational gaps in people with type 1 diabetes: exploratory study of an integrated mobile diabetes app. JMIR Diabetes 2018 Nov 23;3(4):e17 [FREE Full text] [doi: 10.2196/diabetes.9531] [Medline: 30470680]
17. Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z. Exploration of users' perspectives and needs and design of a type 1 diabetes management mobile app: mixed-methods study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11400 [FREE Full text] [doi: 10.2196/11400] [Medline: 30249580]
18. 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]
19. Zhang Y, Li X, Luo S, Liu C, Xie Y, Guo J, et al. Use, perspectives, and attitudes regarding diabetes management mobile apps among diabetes patients and diabetologists in China: national web-based survey. JMIR Mhealth Uhealth 2019 Feb 8;7(2):e12658 [FREE Full text] [doi: 10.2196/12658] [Medline: 30735147]
20. 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]
21. Jimenez G, Lum E, Car J. Examining diabetes management apps recommended from a Google search: content analysis. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11848 [FREE Full text] [doi: 10.2196/11848] [Medline: 30303485]
22. US Food & Drug Adminstration. 2014. FDASIA Health IT Report Proposed Strategy and Recommendations for a Risk-Based Framework URL: https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf [accessed 2019-03-10] [WebCite Cache ID 76lp9UgEO]
23. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
24. Franco P, Gallardo AM, Urtubey X. Web-based interventions for depression in individuals with diabetes: review and discussion. JMIR Diabetes 2018 Sep 14;3(3):e13 [FREE Full text] [doi: 10.2196/diabetes.9694] [Medline: 30291082]
25. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 5;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
26. 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]
27. Öberg U, Isaksson U, Jutterström L, Orre C, Hörnsten A. Perceptions of persons with type 2 diabetes treated in Swedish primary health care: qualitative study on using ehealth services for self-management support. JMIR Diabetes 2018 Mar 12;3(1):e7 [FREE Full text] [doi: 10.2196/diabetes.9059] [Medline: 30291075]

Abbreviations

FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 11.03.19; peer-reviewed by G Jimenez, A Martinez-Millana; comments to author 11.04.19; revised version received 04.06.19; accepted 07.07.19; published 28.08.19.

Please cite as:
Wu Y, Zhou Y, Wang X, Zhang Q, Yao X, Li X, Li J, Tian H, Li S
A Comparison of Functional Features in Chinese and US Mobile Apps for Diabetes Self-Management: A Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2019;7(8):e13971
URL: http://mhealth.jmir.org/2019/8/e13971/
doi:10.2196/13971
PMID:31464191

Original Paper

A Multidimensional Electronic Hydroxyurea Adherence Intervention for Children With Sickle Cell Disease: Single-Arm Before-After Study

Susan Creary[1], MD, MSc; Deena Chisolm[1], PhD; Joseph Stanek[1], MS; Jane Hankins[2], MD, MS; Sarah H O'Brien[1], MD, MSc

[1]Nationwide Children's Hospital, The Ohio State University, Columbus, OH, United States

[2]St Jude Children's Research Hospital, Memphis, TN, United States

Corresponding Author:
Susan Creary, MD, MSc
Nationwide Children's Hospital
The Ohio State University
FOB 3rd Floor
Columbus, OH, 43205
United States
Phone: 1 614 722 3563
Fax: 1 614 722 3559
Email: susan.creary@nationwidechildrens.org

Abstract

Background: Hydroxyurea is a disease-modifying medication for patients with sickle cell disease (SCD). Despite demonstrated efficacy, hydroxyurea nonadherence in clinical practice is common and results in worse health outcomes for nonadherent patients. Mobile Directly Observed Therapy (Mobile DOT) is a pilot-tested, electronic, multidimensional hydroxyurea adherence intervention for children with SCD. Mobile DOT includes sending daily text message reminders to patients to take hydroxyurea, patients recording and sending daily videos that capture their hydroxyurea administrations for the research team to review and track adherence, providing personalized feedback to patients about their adherence, and providing small monetary incentives to patients if they achieve high hydroxyurea adherence.

Objective: This study aimed to determine if Mobile DOT increases hydroxyurea adherence in children with SCD and to explore its impact on hematologic and clinical outcomes.

Methods: This was a single-arm, 6-month intervention study of patients with SCD on hydroxyurea who were aged ≤19 years and reported having access to an electronic device. Participants' hydroxyurea adherence when they received Mobile DOT was compared with their adherence 6 months before and after receiving Mobile DOT. Participants' medication possession ratio (MPR) was calculated from their pharmacy dispensing records and was used to measure adherence. Laboratory and clinical outcomes were abstracted from participants' electronic medical records. Infrequently hospitalized patients who received at least 160 days of the intervention were considered to be engaged participants.

Results: Of 91 patients who were approached, 55 enrolled and 34 engaged with Mobile DOT. The median age of the engaged participants was 10 years (range 2-18.8 years), and 21 (62%, 21/34) participants were male, 28 (82%, 21/34) had hemoglobin SS SCD, and 19 (56%, 19/34) were prescribed hydroxyurea for at least a year before enrollment. With Mobile DOT, engaged participants' median MPR increased from 61.7% to 84.4% (P<.001) and significantly more (67% vs 30%; P=.002) achieved ≥80% hydroxyurea adherence compared with baseline values. Engaged participants' mean fetal hemoglobin (HgbF) levels and mean corpuscular volumes (MCV) improved significantly after 6 months of Mobile DOT (P=.04 and P=.001, respectively), but their adherence, HgbF levels, and MCV returned to baseline values during the 6 months after the intervention. Hospitalizations and the clinical outcomes that were measured occurred infrequently during the study. Nonengagement was associated with being female and having a recent SCD complication. In addition, having insufficient electronic data, being unable to quickly complete Mobile DOT each day, and not perceiving that Mobile DOT was beneficial may have further decreased engagement.

Conclusions: Mobile DOT shows promise as an effective intervention for some children with SCD. Modifications that may improve recruitment, reduce attrition, and increase engagement were identified and could increase the impact that Mobile DOT has on children with SCD.

Trial Registration: ClinicalTrials.gov NCT02578017; https://clinicaltrials.gov/ct2/show/NCT02578017

(JMIR Mhealth Uhealth 2019;7(8):e13452) doi:10.2196/13452

KEYWORDS

hydroxyurea; children; sickle cell disease; mobile health; mhealth; adherence; textmessaging

Introduction

Sickle cell disease (SCD) is the most common, severe inherited blood disorder. Children with SCD experience substantial morbidity, including frequent vaso-occlusive crisis (VOC) pain and acute chest syndrome (ACS) episodes, stroke, and decreased quality of life [1,2]. Hydroxyurea is a once-daily, Food and Drug Administration–approved, preventative, and disease-modifying medication for children with SCD. Multiple large studies show that hydroxyurea improves quality of life while reducing the frequency of VOC and ACS episodes, hospitalizations, hospital readmissions, and transfusions in children with SCD [3-5]. Despite demonstrated efficacy and safety, hydroxyurea nonadherence is common in clinical practice. Health insurance claims data suggest that only 30% to 50% of children with SCD on hydroxyurea in the United States take it on at least 80% of the days it is prescribed, and this nonadherence is associated with worse clinical outcomes, lower quality of life, and higher costs of care for nonadherent patients [2,6,7].

Strategies to improve adherence to medications for other chronic diseases include interventions such as directly observed therapy for tuberculosis [8], reminder alerts for hyperlipidemia and hypertension [9], counseling or motivational interviewing for human immunodeficiency virus [10], and incentives or contingency management for diabetes [11]. In children and adolescents with chronic diseases, including those with SCD, electronic devices show promise as useful tools to provide these interventions to improve adherence and disease self-management [12-15]. Single-dimension adherence interventions, however, have inherent limitations. They also show varying levels of success at increasing adherence likely because they are not able to address multiple components of the Health Behavior Model (HBM), the health promotion theory frequently used to explain medication adherence behavior [16]. The HBM proposes that perceived disease severity, susceptibility to disease complications, self-efficacy, costs and benefits of being adherent, and receiving reminder cues are factors that influence adherence. For example, in the context of the HBM, hydroxyurea nonadherence could occur if patients and/or their families do not perceive the long-term health benefits of taking hydroxyurea outweigh the potential risk of experiencing side effects or the length of time that they must take hydroxyurea before its benefits begin to manifest. In addition, for young children with SCD who rely on caregivers to be adherent, nonadherence could result if caregivers do not have the self-efficacy skills to (1) remember to administer the hydroxyurea when it is due, (2) accurately measure the correct dose of hydroxyurea, and/or (3) deliver hydroxyurea to their children if they are uncooperative with administration. Finally, for adolescents with SCD, perceived invulnerability or insusceptibility, which are common during this stage of development, could contribute to hydroxyurea nonadherence and negative health outcomes.

Therefore, to address multiple HBM constructs that can influence hydroxyurea adherence behavior using devices that most children with SCD report having access to [17], we pilot tested an electronic, multidimensional adherence strategy called Mobile Directly Observed Therapy (Mobile DOT) and published these results in 2014 [18]. Mobile DOT is a hydroxyurea adherence strategy that includes sending daily text message (short message service [SMS]) to patients to remind them to take hydroxyurea, patients recording and sending daily videos that capture their hydroxyurea administrations to the research team to review and track adherence, electronically providing personalized feedback to patients about their adherence, and providing small monetary incentives to patients if they achieve high hydroxyurea adherence. In our pilot work, we found that Mobile DOT was a feasible and acceptable adherence intervention and could achieve ≥90% hydroxyurea adherence in a small cohort of children with SCD. The objectives of this study were to determine if Mobile DOT increases hydroxyurea adherence and to explore its potential impact on hematologic and clinical outcomes in a larger population of children with SCD.

Methods

Study Design, Recruitment, and Participants

These data were collected as part of an institutional review board–approved study. The details of the Mobile DOT intervention and the protocol for this study were previously published [19]. Briefly, this was a single-arm, cross-over study performed at Nationwide Children's Hospital (NCH), a large, comprehensive pediatric hospital in Ohio. English-speaking patients with SCD (any genotype) who were aged ≤19 years, prescribed hydroxyurea for at least the previous 180 days, not receiving chronic transfusion therapy, planned to receive care at NCH for the following year, and had (or their legal guardian had) personal daily access to an electronic device (eg, smartphone or tablet) that could receive SMS text message alerts and record and send daily videos were eligible. Electronic device access was confirmed, before enrollment, when the patient successfully submitted a video to the secure Mobile DOT website.

Patients were sequentially approached to participate when they presented for clinical care, and follow-up study visits were completed during participants' standard of care hematology visits. Participants' electronic medical records (EMRs) were used to collect demographic information, laboratory and clinical data, hydroxyurea dispensing pharmacies, and duration of hydroxyurea treatment. Hydroxyurea dosing and monitoring

were left to clinicians' discretion and tracked using the EMR, but institutional standards, derived from published hydroxyurea protocols [20], were available to guide this management.

There were 3 sequential study periods: baseline, Mobile DOT, and postintervention that were each approximately 180 days. The baseline period was defined as the 180 days before enrollment, the Mobile DOT period started the day following enrollment, and the postintervention period started immediately after the Mobile DOT period. The Mobile DOT and postintervention periods were designed to be 180 days but could vary slightly in length to coordinate participants' study visits with their standard of care hematology visits. In addition, because hydroxyurea administration that occurs in the hospital setting is highly monitored and includes interventions that are separate from Mobile DOT (eg, all hydroxyurea dose administrations are observed and tracked by a nurse), only participants who completed at least 160 days of the Mobile DOT period and who were hospitalized for fewer than 20 days during this time were considered to be engaged participants. Participants who completed at least 160 days of the postintervention period before their final study visit and were hospitalized for fewer than 20 days during this period were considered to have completed the study.

Mobile Directly Observed Therapy

A custom and secure Mobile DOT website to send the alerts and store participants' videos was built by the research technology team at NCH and run under an Apache Web server (version 2.2.15; Apache Software Foundation), was written in the Perl and JavaScript languages, and was run on a 64-bit CentOS 6.5 Linux operating system. Participants were informed before consenting of the potential risk that videos could potentially be intercepted during transmission to the secure website, but their videos were secure once they were received by the website.

Alerts

Participants identified their preferred hydroxyurea administration time and developed their own personalized SMS text message alerts at enrollment. Using a secure Microsoft Exchange mail server, participants and/or their consenting caregivers (if aged <14 years) were sent up to 4 SMS text message alerts to remind them to take hydroxyurea every day.

Video Directly Observed Therapy

Participants were instructed to self-record (or have their consenting caregiver record if aged<11 years) their daily hydroxyurea administrations with their electronic device and email these videos as an attachment every day to the secure website for the research team to review. Participants were trained to record continuous videos that allowed for easy participant recognition, a view of their hydroxyurea before and during administration, and a view of their empty mouth after ingestion. Participants who were instructed to temporarily discontinue hydroxyurea by their clinical provider (eg, for myelosuppression) were instructed to continue to submit daily videos, stating this information to be considered adherent during this time. VDOT was not required during hospitalizations. Temporary electronic device access lapses (for up to 5 days for each 60-day segment of the intervention period) were allowed, but to be considered adherent during these lapses, participants had to notify the research team (by email, SMS text message, or voicemail) on each of these days to confirm that they had taken their hydroxyurea.

A background script on the secure server retrieved the participants' emails with video attachments. Videos were stored locally on the operating system in a participant directory with restricted access to the research team. Information about the participants, the emails, and videos were stored in a local SQLite database. To ensure that the videos were unique, the research team reviewed all videos to ensure that they represented valid hydroxyurea administration and checked the megabyte size of the videos to ensure variability if there was concern that participants were submitting videos that were not unique. Only 2 research team members who were familiar with the participants were responsible for reviewing and tracking adherence.

Feedback

The research team attempted to contact participants (or their consenting caregiver) through SMS text message, email, or telephone when a video was not received to encourage daily hydroxyurea adherence and provide positive feedback when participants achieved the adherence incentive goal.

Incentives

Participants were mailed a US $30 gift card if their videos confirmed at least 90% adherence over the past 30 days. A research team member spent approximately 1 min per participant per day to track adherence and send messages when needed.

Outcome Measures

Medication Possession Ratio

MPR is the percentage of days during a given period of time that patients have access to a medication. MPR was the primary adherence measure used because the pharmacy dispensing data needed for this calculation could be retrospectively collected. Previous studies suggest if multiple months of pharmacy dispensing records are used, MPR provides an accurate quantitative estimate of patients' maximum possible adherence during that time [21-23]. Hydroxyurea dispensing records were requested after the participant completed their participation in the study and from all pharmacies that participants' self-reported they used and/or were identified in their EMR. Participants' hydroxyurea MPR was calculated for each study period using the following formula: the number of days that dispensing records confirmed access to hydroxyurea in a study period divided by the total number of days in a study period, multiplied by 100. MPR calculations accounted for hydroxyurea doses that were held or received during any hospitalization and for any prescribed dose adjustments documented in the EMR during the study. Participants' with a hydroxyurea MPR of ≥80% during a study period were defined as adherent for that period because improved clinical outcomes were seen at this adherence threshold in a randomized pediatric SCD hydroxyurea clinical trial [3,24].

Video Directly Observed Therapy Adherence

Observed adherence, either in-person or by video, is a reliable and valid measure of patients' medication adherence [21]. Engaged participants' VDOT adherence was calculated using the following formula: the number of days VDOT was completed divided by the total number of days in the Mobile DOT period multiplied by 100. VDOT adherence calculations accounted for days that the research team was notified of device access lapse and for doses received during hospitalizations.

Laboratory Data

Large pediatric clinical trials show that hydroxyurea exposure significantly increases fetal hemoglobin (HgbF) production, total hemoglobin (Hgb), and mean corpuscular volume (MCV) and reduces absolute neutrophil counts (ANC) [3,25-27]. In this study, laboratory studies that were obtained during standard of care hematology visits within 40 days of the end of a study period were analyzed. NCH uses Sebia Zone Electrophoresis to reliably quantify HgbF levels [28] and the Sysmex XN-1000, a standard hematology analyzer, to reliably measure Hgb, MCV, and ANC [29].

Clinical Outcomes

Multiple pediatric clinical trials confirm that hydroxyurea reduces hospitalizations, erythrocyte transfusions, ACS episodes, and VOC episodes requiring hospitalization [3,30]. These outcomes were abstracted from participants' EMR during the study.

Satisfaction Survey

To measure the amount of time that it took participants to complete Mobile DOT each day, their willingness to continue the intervention without payment, and their satisfaction and perceived effectiveness of the intervention, participants (or the consenting caregivers of participants aged<14 years) completed the 5-point Likert scale Mobile DOT satisfaction survey [18] at the end of their Mobile DOT period or at their withdrawal visit. Participants who withdrew were also asked to provide a reason for withdrawal.

Study Withdrawal

Participants who discontinued hydroxyurea moved out of the area or informed the research team that they did not want to continue to participate were immediately withdrawn from the study. The research team also stopped sending SMS text message alerts and attempting to contact participants who did not submit a video or respond to research staff communications for more than 30 consecutive days during the Mobile DOT period. These participants were withdrawn from the study at their next clinic visit.

Statistical Analysis

Descriptive statistics were used to describe sample characteristics and report the clinical outcomes, frequency and percentages for qualitative variables, and median and range or mean and standard error for quantitative variables. Nonparametric tests were used to compare baseline characteristics between engaged and nonengaged participants. McNemar test was used to compare proportion of patients who were adherent at baseline and with Mobile DOT. To compare adherence and hematologic measures across the 3 study periods, linear mixed models with Tukey-adjusted pairwise comparisons were used. All *P* values were 2-sided, and $P<.05$ was considered statistically significant. Statistical analyses were performed using SAS software, version 9.4 (SAS Institute).

Results

Study Participants

Between July 2014 and January 2017, 91 unique prospective participants were approached, 55 enrolled (76% of the planned sample size), 34 engaged with the Mobile DOT intervention, and 29 completed the study (Figure 1). All the participants planned to use a smartphone as their electronic device. None of the patients who declined participation reported privacy concerns as a reason for study refusal. Recruitment ended before enrolling the planned sample size when 97% of the prospective patients had been approached because recruitment would have needed to be extended for multiple months to approach the remaining patients at their standard hematology visits.

Dispensing records were received for 54 (33 engaged) participants, and these participants were included in the MPR analyses. At baseline, 14 (25%, 14/54) enrolled participants were adherent (MPR ≥80%) to hydroxyurea, and 17 (31%, 17/54) had a hydroxyurea MPR <50%.

Table 1 shows a comparison of the baseline characteristics between engaged and nonengaged participants. Of the 17 engaged participants who were aged<10 years, 8 (47%) were adherent to hydroxyurea at baseline and 3 (27%) of the 11 nonengaged participants who were aged<10 years were adherent at baseline. In addition, 2 (13%) of the 17 engaged participants who were aged ≥10 years were adherent to hydroxyurea at baseline, and one (10%) of the 10 nonengaged participants who were aged ≥10 years was adherent to hydroxyurea at baseline. Engaged participants received Mobile DOT for a median of 191 days, and those who completed the study were followed for a median of 181 days during the postintervention period. Engaged participants' median prescribed hydroxyurea dose did not significantly change from baseline to the end of the Mobile DOT period (25.9 mg/kg/day vs 25.8 mg/kg/day; $P>.99$). Of those who completed the study, 21 (72%) were prescribed a dose that was within 5 mg/kg/day of their baseline dose, but their median prescribed hydroxyurea dose was significantly lower at the end of the postintervention period compared with baseline (23.5 mg/kg/day; $P=.004$).

Figure 1. Participant flow diagram (Mobile DOT: Mobile Directly Observed Therapy).

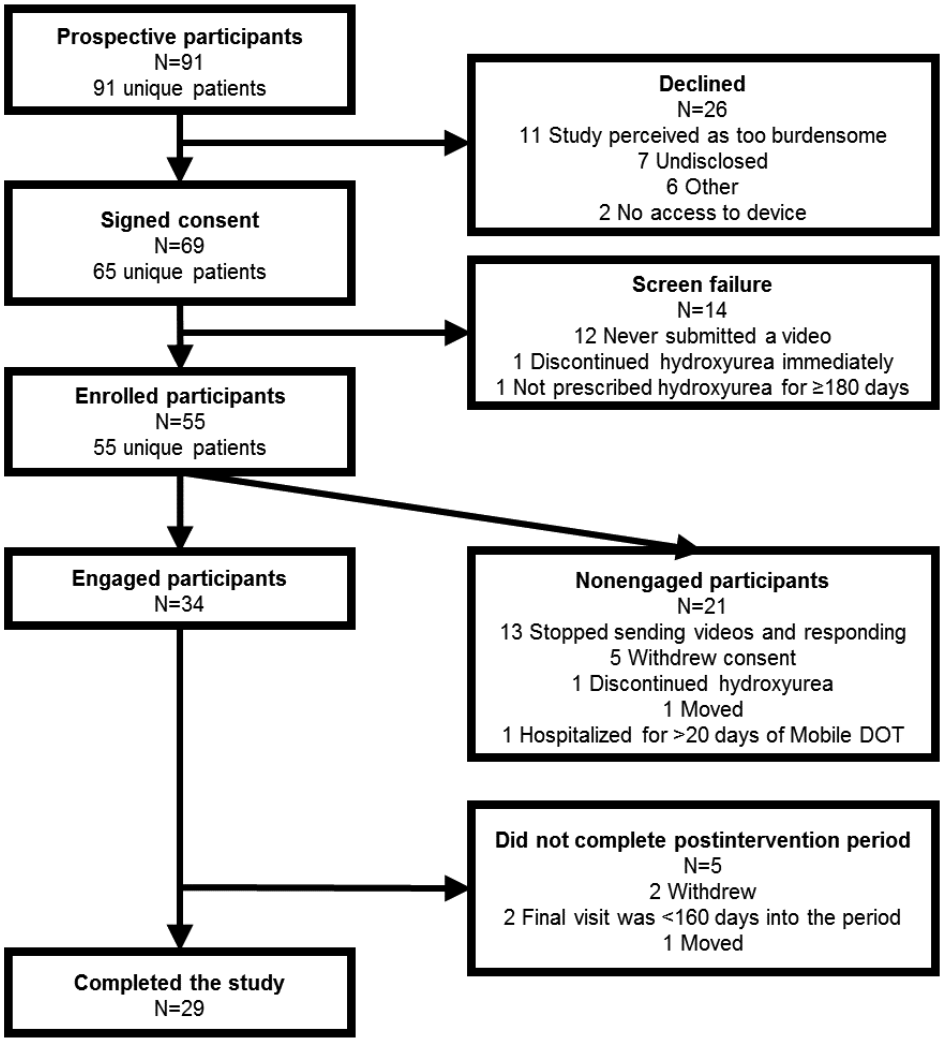


Table 1. Participants' baseline characteristics.

Characteristics	All (n=55)	Nonengaged (n=21)	Engaged (n=34)	*P* value
Median age (years; range)	10.0 (2.0-19.0)	9.4 (2.7-19.0)	10.0 (2.0-18.8)	.95
Age (years), n (%)				.99
<10	28 (51)	11 (52)	17 (50)	
Oct-17	22 (40)	8 (38)	14 (41)	
≥18	5	2	3	
Male, n (%)	26 (47)	5 (234)	21 (62)	.01
Race, n (%)				.52
Black or African American	53 (96)	21 (100)	32 (94)	
Biracial	2	0	2	
Genotype, n (%)				.93
SS	47 (856)	19 (91)	28 (82)	
SC	3 (6)	1 (5)	2 (6)	
SB^0	2 (4)	1 (5)	1 (3)	
SB^+	2 (4)	0 (0)	2 (6)	
Other (SE)	1 (2)	0 (0)	1 (3)	
Duration of hydroxyurea use, n (%)				.8
6-12 months	25 (456)	10 (48)	15 (44)	
>12 months	30 (55)	11 (52)	19 (56)	
Indication for hydroxyurea, n (%)				
Clinical complication	53 (96)	21 (100)	32 (94)	.52
Severe genotype or family request[a]	4 (7)	1 (5)	3 (9)	.99
Median hydroxyurea dose (mg/kg/d)	25.9	25.9	25.9	.41
Hydroxyurea MPR[b]				
Median, %	61.7	66.7	61.1	.86
0%-19%, n	3	1	2	—[c]
20%-39%, n	9	6	3	—
40%-59%, n	11	1	10	—
60%-79%, n	17	9	8	—
≥80%, n	14	4	10	—
Median hematologic studies				
Hemoglobin, g/dL	9.3	8.6	9.4	.14
MCV[d], fl/L	90.9	91.8	88.2	.68
Hemoglobin F, %	21.1	17.8	24.8	.07
Absolute neutrophil count per μL	3690	4770	3415	.003

[a]Initiated hydroxyurea therapy before experiencing a sickle cell disease complication.

[b]Medication possession ratio (MPR) could not be calculated for 1 participant because dispensing records were not received.

[c]Descriptive data only.

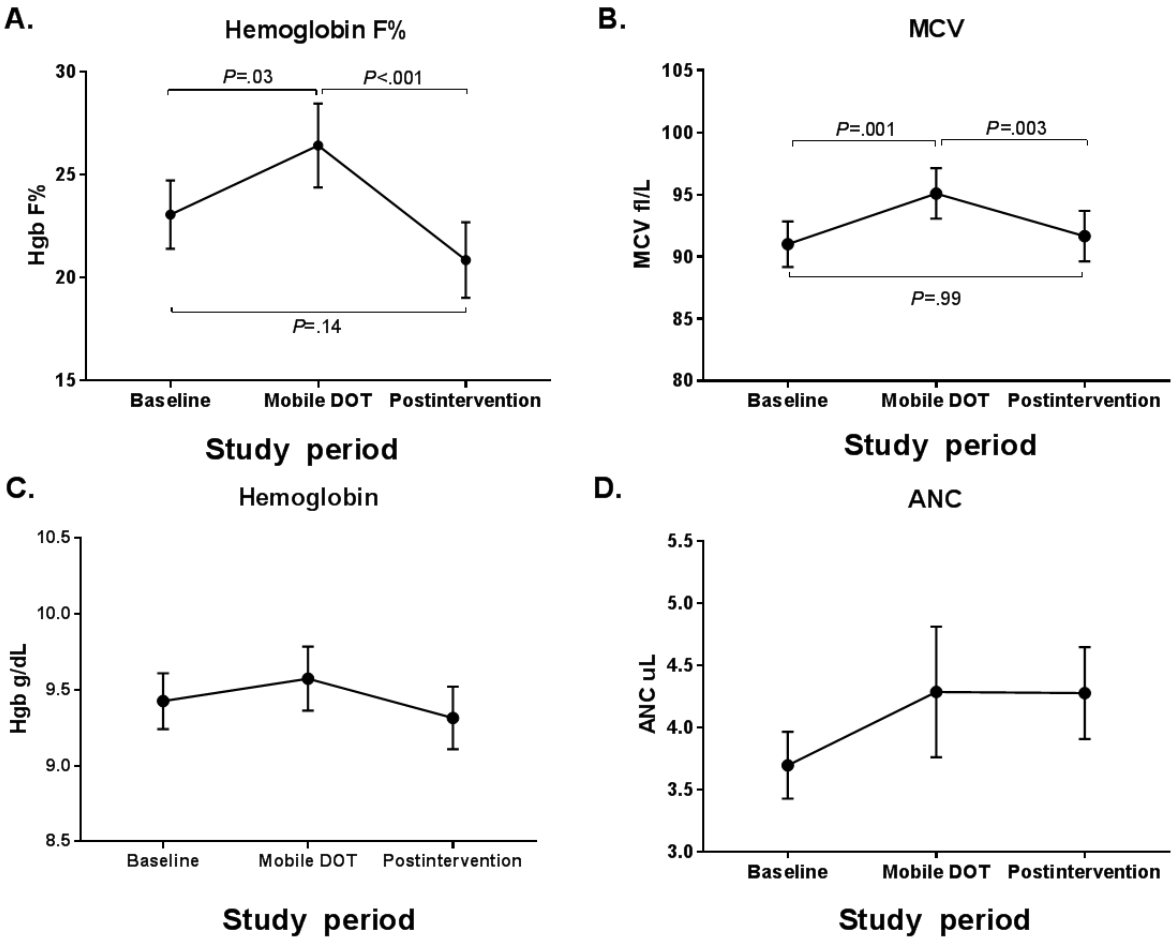
[d]MCV: mean corpuscular volume.

Hydroxyurea Adherence

Participants' median hydroxyurea MPR was significantly higher with Mobile DOT compared with baseline, but median postintervention MPR was not significantly different compared with baseline (Figure 2). The proportion of engaged participants that were adherent (MPR ≥80%) to hydroxyurea significantly increased from 30% at baseline to 67% with Mobile DOT

(*P*=.002). In addition, 20 (61%) of the engaged participants increased their MPR by >15% with Mobile DOT compared with their baseline. Median VDOT adherence (88.1%) and MPR adherence (87.8%) for the engaged participants during the Mobile DOT period were not significantly different (*P*=.99).

Figure 2. Participants' hydroxyurea MPR adherence during each study period (Mobile DOT: Mobile Directly Observed Therapy; MPR: medication possession ratio).

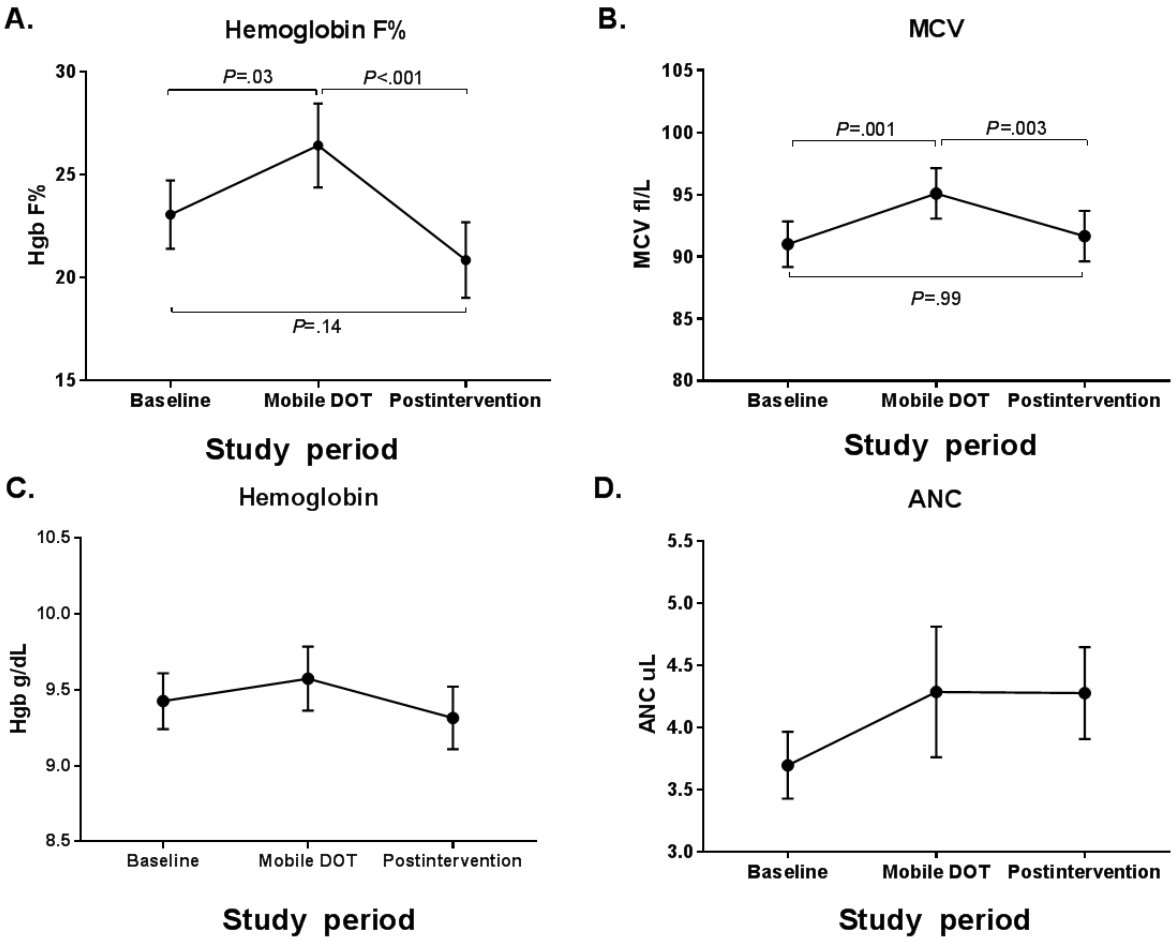


Hematologic and Clinical Outcomes

Engaged participants' mean MCV and HgbF levels at the end of Mobile DOT were significantly higher than baseline but were not significantly different at the end of postintervention period compared with baseline. Engaged participants' mean Hgb and ANC did not significantly change throughout the study (Figure 3).

During the baseline period, 8 participants received at least one erythrocyte transfusion (6 nonengaged vs 2 engaged, *P*=.04), 11 had at least one ACS episode (8 nonengaged vs 3 engaged, *P*=.01), and 12 had at least one VOC hospitalization (7 nonengaged vs 5 engaged, *P*=.18). The cohort experienced a total of 13 ACS episodes and 26 VOC hospitalizations during the baseline period, but 1 participant accounted for 9 of these VOC hospitalizations and was nonengaged during the Mobile DOT period because of continued recurrent VOC hospitalizations. During the Mobile DOT period, 3 engaged participants received at least one erythrocyte transfusion, 2 had at least one ACS episode, and 6 had at least one VOC hospitalization. During the postintervention period, 1 participant required at least one transfusion, 3 had at least one ACS episode, and 3 had at least one VOC hospitalization.

Figure 3. Engaged participants' (n=34) laboratory studies at the end of each study period (ANC: absolute neutrophil count; Hgb: hemoglobin; MCV: mean corpuscular volume; Mobile DOT: Mobile Directly Observed Therapy).



Satisfaction Survey

All the engaged participants and 16 (76%) of the nonengaged participants completed the satisfaction survey. Of the engaged participants, 31 (92%) agreed that they took hydroxyurea more often with Mobile DOT, 34 (100%) agreed that Mobile DOT was easy to use, 30 (88%) agreed that they would be willing to continue Mobile DOT, 19 (56%) agreed that they would be willing to continue Mobile DOT without pay, and 30 (88%) reported they could complete Mobile DOT in less than 3 min each day. Of the nonengaged participants, 9 (56%) agreed that they took hydroxyurea more often with Mobile DOT, 8 (50%) agreed that Mobile DOT was easy to use, 7 (44%) agreed they would be willing to continue Mobile DOT, 5 (31%) agreed that they would be willing to continue Mobile DOT without pay, and 8 (50%) reported that they were able to complete Mobile DOT in less than 3 min each day.

Discussion

Hydroxyurea Adherence

Hydroxyurea nonadherence is common among children with SCD [7,24], and an effective hydroxyurea adherence intervention is needed to improve health outcomes. Overall, our results suggest that Mobile DOT improves hydroxyurea adherence, at least temporarily, in children who engage with the intervention. Of the engaged children, two-thirds achieved 80% or higher MPR adherence with Mobile DOT, which is noteworthy considering many of these children had appreciably lower baseline hydroxyurea adherence. Also, median MPR significantly improved with Mobile DOT, despite including some subjects with less potential to improve their MPR because they had 80% or higher adherence at baseline. Finally, although not all the engaged participants achieved 80% or higher adherence with Mobile DOT, many still incrementally increased their adherence (eg, >15% increase or an additional hydroxyurea dose per week), suggesting that Mobile DOT has the potential to improve adherence for some of the most nonadherent children.

Hematologic Outcomes

Our results also showed that engaged participants' mean MCV and HgbF levels increased during the Mobile DOT period. Given that MCV and HgbF level changes from hydroxyurea exposure typically occur before clinical improvements manifest [30], this could be an early indication that with continued use, Mobile DOT may have potential hematologic and clinical benefits. Although significant Hgb changes were not appreciated, this may be related to the relatively short duration of time that Mobile DOT was provided compared with the time required to see improved Hgb with hydroxyurea [3] and our participants' relatively high baseline Hgb levels. In addition, 15% of our

engaged patients had either Hemoglobin SC, $S\beta^+$, or SE, and past studies suggest that other SCD genotypes may not have the same degree of hematologic response with hydroxyurea compared with those with hemoglobin SS [31,32].

Similar to other medication adherence interventions [33], participants' adherence and laboratory improvements declined once the intervention was discontinued. Survey results showed that many engaged participants were willing to continue Mobile DOT beyond 6 months, suggesting that for some, Mobile DOT may be able to be used longer term to sustain adherence. However, short booster sessions or additional interventions may be needed for others when their adherence declines. Laboratory improvements likely returned to baseline because adherence declined but also the median prescribed hydroxyurea dose decreased by the end of the study. Mean ANC did not change during the study to suggest that severe myelosuppression frequently occurred that would have necessitated this decrease. Most participants were also prescribed doses at the end of the study that were similar to their baseline doses. We, therefore, suspect this decrease was because many participants were growing children and the hydroxyurea protocol [20] and institutional guidelines do not provide guidance about how frequently to increase hydroxyurea to account for weight gain from growth.

Clinical Outcomes

Future studies are needed to definitively determine if Mobile DOT reduces erythrocyte transfusions, hospitalizations, and ACS and VOC episodes because these events occurred infrequently at baseline and participants were followed for a relatively short period of time. To limit underestimating the benefit of effective adherence interventions, future studies should also measure their impact on preventing other patient-reported and costly outcomes, such as VOC episodes, that do not result in hospitalization [34] and stroke [35] that may be ameliorated or prevented with hydroxyurea.

Limitations

Several limitations must be considered when interpreting our results. First, despite approaching almost all the eligible patients at NCH, we were unable to enroll our projected sample size [19]. This occurred because we anticipated that more patients at NCH were going to initiate hydroxyurea and become eligible than actually did and because some patients perceived the intervention as too burdensome. These findings underscore the need for interventions, such as shared decision-making tools [36], to increase hydroxyurea acceptance and add to the existing literature that suggests that the time and burden to participate in clinical studies for patients with SCD are prohibitive [37]. In the future, we plan to recruit patients outside of the clinical care setting to avoid approaching patients when they may be overwhelmed by other aspects of their SCD care. This strategy could also increase the number of nonadherent patients that are recruited because clinic nonadherence is associated with hydroxyurea nonadherence [38].

Second, only two-thirds of the enrolled participants remained engaged with Mobile DOT. As only those who engaged improved their adherence, it is possible that the factors that promoted engagement may have contributed to the observed adherence gains. Therefore, identifying these mediating factors is paramount to increasing the impact that this and other adherence interventions may have on children with SCD. For example, our results suggest that nonengaged participants were more likely to be female and have experienced a recent SCD complication (eg, erythrocyte transfusion and/or ACS episode) compared with engaged participants. The small sample size and small number of clinical complications that occurred during the study limit the ability to definitively determine if these factors influence engagement, if they were related to an unmeasured moderating factor, or if they were merely a random finding observed in this sample. In addition, consistent with the literature that suggests widespread electronic device ownership among African Americans [39], including children with SCD [13], only 3 patients reported that not owning a device prevented them from participating or led to their study withdrawal. However, increasing literature suggests that having an insufficient data plan or inconsistent access to Wi-Fi are common among minority populations [40] and could explain why multiple participants became nonengaged and stopped submitting videos or responding to the research team shortly after they enrolled. Although providing devices and cellular data could potentially increase enrollment and engagement with electronic health interventions and ensure consistent intervention delivery, this could also add significant costs that may limit intervention sustainability and use in clinical practice. Finally, although our study suggests Mobile DOT increases hydroxyurea adherence, it is possible that the Hawthorne effect may have also contributed because participants were aware their adherence was going to be closely monitored.

Participant Feedback

Participants' feedback from this study provides insights into modifications that may make this and other adherence interventions more engaging, reduce attrition, and increase adherence in future studies. First, changing the frequency that medication reminders alerts are provided could reduce alert fatigue. In addition, providing feedback that notifies participants how their adherence compares with their previous adherence and with others' could influence participants' perception of the intervention. Incentives may be needed to reward both high adherence and incremental adherence improvements to encourage the most nonadherent participants to continue to engage with the intervention. Finally, for multidimensional interventions, quantifying each component's impact on adherence could identify if components need to be eliminated or optimized or if additional components, such as gamification, may be needed because they might promote engagement with the entire intervention [41].

MPR was used to measure hydroxyurea adherence in this study. Although the medication event monitor system caps are considered the *gold standard* adherence measure [42], these devices were not used in this study because they are not compatible with liquid hydroxyurea formulations, which are commonly used in young children. It is important to note, however, that MPR only measures access to medication and not actual medication administration and can also be unreliable if dispensing reports are inaccurate or incomplete [21]. Engaged

participants' hydroxyurea adherence by MPR and VDOT were similar in this study, but for patients who may have accumulated a supply of hydroxyurea from past prescriptions, MPR may underestimate adherence. Ultimately, studies that validate hydroxyurea adherence measures are needed to accurately identify nonadherent children who could benefit from an adherence intervention. Furthermore, other hydroxyurea adherence interventions are under study, including an electronic alert intervention [43] and a community health care worker intervention [44], and show feasibility, and valid measures will be needed to determine which intervention is most effective.

Finally, systematic reviews suggest there are insufficient data to determine if mobile health interventions, such as Mobile DOT, are cost-effective [45,46]. Although an increasing number of insurers are already providing financial incentives to their members to encourage healthy behaviors [47], to employ insurers to provide this health behavior intervention in clinical practice, future studies will need to evaluate if it is economically feasible and sustainable. We acknowledge that the cost to improve hydroxyurea adherence may be significant, considering poor adherence is difficult to improve and sustain. However, the cost to provide this intervention could be reduced if technology improvements reduce the time and resources required for implementation and if it is reserved for patients with adherence challenges. Ultimately, we suspect that because SCD is a chronic disease that results in costly complications that are reduced with effective hydroxyurea use [48], even relatively costly adherence interventions may still be implementable, and perhaps cost-effective, in clinical practice.

Conclusions

In conclusion, we demonstrate improved 6-month hydroxyurea adherence, MCV, and HgbF level among children with SCD who engaged with an electronic, multidimensional hydroxyurea adherence intervention. Future studies that determine how to increase patient engagement with hydroxyurea adherence interventions, how to sustain improved adherence, and how these interventions impact the outcomes of children with SCD are warranted. If successful, they may be generalizable to improve medication adherence and outcomes of other chronically ill populations.

Acknowledgments

SC was supported by the National Heart Lung and Blood Institute 5K23HL127303. The content is the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N Engl J Med 1994 Jun 9;330(23):1639-1644. [doi: 10.1056/NEJM199406093302303] [Medline: 7993409]
2. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Health-related quality of life and adherence to hydroxyurea in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2017 Dec;64(6). [doi: 10.1002/pbc.26369] [Medline: 27896936]
3. Wang WC, Ware RE, Miller ST, Iyer RV, Casella JF, Minniti CP, BABY HUG Investigators. Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). Lancet 2011 May 14;377(9778):1663-1672 [FREE Full text] [doi: 10.1016/S0140-6736(11)60355-3] [Medline: 21571150]
4. Thornburg CD, Calatroni A, Panepinto JA. Differences in health-related quality of life in children with sickle cell disease receiving hydroxyurea. J Pediatr Hematol Oncol 2011 May;33(4):251-254 [FREE Full text] [doi: 10.1097/MPH.0b013e3182114c54] [Medline: 21516020]
5. Zhou J, Han J, Nutescu EA, Gordeuk VR, Saraf SL, Calip GS. Hydroxycarbamide adherence and cumulative dose associated with hospital readmission in sickle cell disease: a 6-year population-based cohort study. Br J Haematol 2018 Dec;182(2):259-270 [FREE Full text] [doi: 10.1111/bjh.15396] [Medline: 29767446]
6. Thornburg CD, Calatroni A, Telen M, Kemper AR. Adherence to hydroxyurea therapy in children with sickle cell anemia. J Pediatr 2010 Mar;156(3):415-419 [FREE Full text] [doi: 10.1016/j.jpeds.2009.09.044] [Medline: 19880135]
7. Candrilli SD, O'Brien SH, Ware RE, Nahata MC, Seiber EE, Balkrishnan R. Hydroxyurea adherence and associated outcomes among Medicaid enrollees with sickle cell disease. Am J Hematol 2011 Mar;86(3):273-277 [FREE Full text] [doi: 10.1002/ajh.21968] [Medline: 21328441]
8. Iseman MD, Cohn DL, Sbarbaro JA. Directly observed treatment of tuberculosis. We can't afford not to try it. N Engl J Med 1993 Feb 25;328(8):576-578. [doi: 10.1056/NEJM199302253280811] [Medline: 8426627]
9. Chapman RH, Ferrufino CP, Kowal SL, Classi P, Roberts CS. The cost and effectiveness of adherence-improving interventions for antihypertensive and lipid-lowering drugs*. Int J Clin Pract 2010 Jan;64(2):169-181. [doi: 10.1111/j.1742-1241.2009.02196.x] [Medline: 20089007]

10. Binford MC, Kahana SY, Altice FL. A systematic review of antiretroviral adherence interventions for HIV-infected people who use drugs. Curr HIV/AIDS Rep 2012 Dec;9(4):287-312 [FREE Full text] [doi: 10.1007/s11904-012-0134-8] [Medline: 22936463]
11. Stanger C, Ryan SR, Delhey LM, Thrailkill K, Li Z, Li Z, et al. A multicomponent motivational intervention to improve adherence among adolescents with poorly controlled type 1 diabetes: a pilot study. J Pediatr Psychol 2013 Jul;38(6):629-637 [FREE Full text] [doi: 10.1093/jpepsy/jst032] [Medline: 23699750]
12. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
13. 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 Dec 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
14. Thakkar J, Kurup R, Laba TL, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
15. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
16. Clark NM, Houle CR. Theoretical models and strategies for improving adherence and disease management. In: Shumaker SA, Schron EB, Ockene JK, McBee WL, editors. The Handbook of Health Behavior Change. New York: Springer Publishing; 1998:5-32.
17. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
18. Creary SE, Gladwin MT, Byrne M, Hildesheim M, Krishnamurti L. A pilot study of electronic directly observed therapy to improve hydroxyurea adherence in pediatric patients with sickle-cell disease. Pediatr Blood Cancer 2014 Jun;61(6):1068-1073. [doi: 10.1002/pbc.24931] [Medline: 24436121]
19. Creary S, Chisolm DJ, O'Brien SH. ENHANCE-(Electronic Hydroxyurea Adherence): a protocol to increase hydroxyurea adherence in patients with sickle cell disease. JMIR Res Protoc 2016 Oct 3;5(4):e193 [FREE Full text] [doi: 10.2196/resprot.6403] [Medline: 27697749]
20. Yawn BP, Buchanan GR, Afenyi-Annan AN, Ballas SK, Hassell KL, James AH, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. J Am Med Assoc 2014 Sep 10;312(10):1033-1048. [doi: 10.1001/jama.2014.10517] [Medline: 25203083]
21. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
22. Johnston JM, Pennypacker HS. Strategies and Tactics of Behavioral Research. New York: Routledge; 2009.
23. Andrade SE, Kahler KH, Frech F, Chan KA. Methods for evaluation of medication adherence and persistence using automated databases. Pharmacoepidemiol Drug Saf 2006 Aug;15(8):565-74; discussion 575. [doi: 10.1002/pds.1230] [Medline: 16514590]
24. Thornburg CD, Rogers ZR, Jeng MR, Rana SR, Iyer RV, Faughnan L, BABY HUG Investigators. Adherence to study medication and visits: data from the BABY HUG trial. Pediatr Blood Cancer 2010 Feb;54(2):260-264 [FREE Full text] [doi: 10.1002/pbc.22324] [Medline: 19856395]
25. Ware RE, Eggleston B, Redding-Lallinger R, Wang WC, Smith-Whitley K, Daeschner C, et al. Predictors of fetal hemoglobin response in children with sickle cell anemia receiving hydroxyurea therapy. Blood 2002 Jan 1;99(1):10-14 [FREE Full text] [doi: 10.1182/blood.v99.1.10] [Medline: 11756146]
26. Cokic VP, Smith RD, Beleslin-Cokic BB, Njoroge JM, Miller JL, Gladwin MT, et al. Hydroxyurea induces fetal hemoglobin by the nitric oxide-dependent activation of soluble guanylyl cyclase. J Clin Invest 2003 Jan;111(2):231-239 [FREE Full text] [doi: 10.1172/JCI16672] [Medline: 12531879]
27. Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the multicenter study of hydroxyurea in sickle cell anemia. N Engl J Med 1995 May 18;332(20):1317-1322. [doi: 10.1056/NEJM199505183322001] [Medline: 7715639]
28. Greene DN, Pyle AL, Chang JS, Hoke C, Lorey T. Comparison of Sebia capillarys flex capillary electrophoresis with the BioRad variant II high pressure liquid chromatography in the evaluation of hemoglobinopathies. Clin Chim Acta 2012 Aug 16;413(15-16):1232-1238. [doi: 10.1016/j.cca.2012.03.027] [Medline: 22515960]
29. Becker P, Fenneteau O, da Costa L. Performance evaluation of the Sysmex XN-1000 hematology analyzer in assessment of the white blood cell count differential in pediatric specimens. Int J Lab Hematol 2016 Feb;38(1):54-63. [doi: 10.1111/ijlh.12436] [Medline: 26407903]

30. Estepp JH, Smeltzer MP, Kang G, Li C, Wang WC, Abrams C, et al. A clinically meaningful fetal hemoglobin threshold for children with sickle cell anemia during hydroxyurea therapy. Am J Hematol 2017 Dec;92(12):1333-1339 [FREE Full text] [doi: 10.1002/ajh.24906] [Medline: 28913922]
31. Zimmerman SA, Schultz WH, Davis JS, Pickens CV, Mortier NA, Howard TA, et al. Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. Blood 2004 Mar 15;103(6):2039-2045 [FREE Full text] [doi: 10.1182/blood-2003-07-2475] [Medline: 14630791]
32. Luchtman-Jones L, Pressel S, Hilliard L, Brown RC, Smith MG, Thompson AA, et al. Effects of hydroxyurea treatment for patients with hemoglobin SC disease. Am J Hematol 2016 Feb;91(2):238-242 [FREE Full text] [doi: 10.1002/ajh.24255] [Medline: 26615793]
33. Berg KM, Litwin AH, Li X, Heo M, Arnsten JH. Lack of sustained improvement in adherence or viral load following a directly observed antiretroviral therapy intervention. Clin Infect Dis 2011 Nov;53(9):936-943 [FREE Full text] [doi: 10.1093/cid/cir537] [Medline: 21890753]
34. Smith WR, Ballas SK, McCarthy WF, Bauserman RL, Swerdlow PS, Steinberg MH, Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. The association between hydroxyurea treatment and pain intensity, analgesic use, and utilization in ambulatory sickle cell anemia patients. Pain Med 2011 May;12(5):697-705. [doi: 10.1111/j.1526-4637.2011.01096.x] [Medline: 21481164]
35. Ware RE, Davis BR, Schultz WH, Brown RC, Aygun B, Sarnaik S, et al. Hydroxycarbamide versus chronic transfusion for maintenance of transcranial doppler flow velocities in children with sickle cell anaemia-TCD with transfusions changing to hydroxyurea (TWiTCH): a multicentre, open-label, phase 3, non-inferiority trial. Lancet 2016 Feb 13;387(10019):661-670 [FREE Full text] [doi: 10.1016/S0140-6736(15)01041-7] [Medline: 26670617]
36. Crosby LE, Shook LM, Ware RE, Brinkman WB. Shared decision making for hydroxyurea treatment initiation in children with sickle cell anemia. Pediatr Blood Cancer 2015 Dec;62(2):184-185 [FREE Full text] [doi: 10.1002/pbc.25124] [Medline: 25308571]
37. Lebensburger JD, Sidonio RF, Debaun MR, Safford MM, Howard TH, Scarinci IC. Exploring barriers and facilitators to clinical trial enrollment in the context of sickle cell anemia and hydroxyurea. Pediatr Blood Cancer 2013 Aug;60(8):1333-1337 [FREE Full text] [doi: 10.1002/pbc.24486] [Medline: 23418000]
38. Creary SE, O'Brien SH, Stanek J, Hankins JS, Garee A, Ball KR, et al. A retrospective review to determine if children with sickle cell disease receive hydroxyurea monitoring. Pediatr Qual Saf 2017;2(3):e024 [FREE Full text] [Medline: 29399659]
39. Horrigan J. Pew Research Center. 2009. Wireless Internet Use URL: http://pewinternet.org/Reports/2009/12-Wireless-Internet-Use.aspx
40. Smith A. Pew Research Center. 2014. African Americans and Technology Use: A Demographic Portrait URL: http://www.pewresearch.org/wp-content/uploads/sites/9/2014/01/African-Americans-and-Technology-Use.pdf
41. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
42. Ingerski LM, Hente EA, Modi AC, Hommel KA. Electronic measurement of medication adherence in pediatric chronic illness: a review of measures. J Pediatr 2011 Oct;159(4):528-534 [FREE Full text] [doi: 10.1016/j.jpeds.2011.05.018] [Medline: 21722917]
43. Estepp JH, Winter B, Johnson M, Smeltzer MP, Howard SC, Hankins JS. Improved hydroxyurea effect with the use of text messaging in children with sickle cell anemia. Pediatr Blood Cancer 2014 Nov;61(11):2031-2036. [doi: 10.1002/pbc.25177] [Medline: 25132074]
44. Green NS, Manwani D, Matos S, Hicks A, Soto L, Castillo Y, et al. Randomized feasibility trial to improve hydroxyurea adherence in youth ages 10-18 years through community health workers: the HABIT study. Pediatr Blood Cancer 2017 Dec;64(12) [FREE Full text] [doi: 10.1002/pbc.26689] [Medline: 28643377]
45. 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]
46. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
47. Loewenstein G, Asch DA, Volpp KG. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood) 2013 Jul;32(7):1244-1250. [doi: 10.1377/hlthaff.2012.1163] [Medline: 23836740]
48. Wang WC, Oyeku SO, Luo Z, Boulet SL, Miller ST, Casella JF, BABY HUG Investigators. Hydroxyurea is associated with lower costs of care of young children with sickle cell anemia. Pediatrics 2013 Oct;132(4):677-683 [FREE Full text] [doi: 10.1542/peds.2013-0333] [Medline: 23999955]

Abbreviations

ACS: acute chest syndrome

ANC: absolute neutrophil count
EMR: electronic medical record
HgbF: fetal hemoglobin
Hgb: total hemoglobin
HBM: Health Behavior Model
MCV: mean corpuscular volume
MPR: medication possession ratio
Mobile DOT: Mobile Directly Observed Therapy
NCH: Nationwide Children's Hospital
SCD: sickle cell disease
SMS: short message service
VOC: vaso-occlusive crisis

Edited by G Eysenbach; submitted 21.01.19; peer-reviewed by N Shah, NA Duncan, S Badawy; comments to author 14.03.19; revised version received 19.04.19; accepted 23.04.19; published 08.08.19.

Please cite as:
Creary S, Chisolm D, Stanek J, Hankins J, O'Brien SH
A Multidimensional Electronic Hydroxyurea Adherence Intervention for Children With Sickle Cell Disease: Single-Arm Before-After Study
JMIR Mhealth Uhealth 2019;7(8):e13452
URL: https://mhealth.jmir.org/2019/8/e13452/
doi:10.2196/13452
PMID:31397291

Original Paper

Association Between User Engagement of a Mobile Health App for Gout and Improvements in Self-Care Behaviors: Randomized Controlled Trial

Anna Serlachius[1], PhD; Kiralee Schache[1], MSc; Anel Kieser[1], MSc; Bruce Arroll[2], MD; Keith Petrie[1], PhD; Nicola Dalbeth[3], MD, FRACP

[1]Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[2]General Practice and Primary Healthcare, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[3]Bone and Joint Research Group, Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

Corresponding Author:
Anna Serlachius, PhD
Department of Psychological Medicine
Faculty of Medical and Health Sciences
University of Auckland
Private Bag 92019
Victoria Street West
Auckland, 1142
New Zealand
Phone: 64 9 9233073
Email: a.serlachius@auckland.ac.nz

Abstract

Background: Mobile health (mHealth) apps represent a promising approach for improving health outcomes in patients with chronic illness, but surprisingly few mHealth interventions have investigated the association between user engagement and health outcomes. We aimed to examine the efficacy of a recommended, commercially available gout self-management app for improving self-care behaviors and to assess self-reported user engagement of the app in a sample of adults with gout.

Objective: Our objective was to examine differences in self-reported user engagement between a recommended gout app (treatment group) and a dietary app (active control group) over 2 weeks as well as to examine any differences in self-care behaviors and illness perceptions.

Methods: Seventy-two adults with gout were recruited from the community and three primary and secondary clinics. Participants were randomized to use either Gout Central (n=36), a self-management app, or the Dietary Approaches to Stop Hypertension Diet Plan (n=36), an app based on a diet developed for hypertension, for 2 weeks. The user version of the Mobile Application Rating Scale (uMARS, scale: 1 to 5) was used after the 2 weeks to assess self-reported user engagement, which included an open-ended question. Participants also completed a self-report questionnaire on self-care behaviors (scale: 1-5 for medication adherence and diet and 0-7 for exercise) and illness perceptions (scale: 0-10) at baseline and after the 2-week trial. Independent samples t tests and analysis of covariance were used to examine differences between groups at baseline and postintervention.

Results: Participants rated the gout app as more engaging (mean difference –0.58, 95% CI –0.96 to –0.21) and more informative (mean difference –0.34, 95% CI –0.67 to –0.01) than the dietary app at the 2-week follow-up. The gout app group also reported a higher awareness of the importance of gout (mean difference –0.64, 95% CI –1.27 to –0.003) and higher knowledge/understanding of gout (mean difference –0.70, 95% CI –1.30 to –0.09) than the diet app group at follow-up. There were no significant differences in self-care behaviors between the two groups postintervention. The gout app group also demonstrated stronger negative beliefs regarding the impact of gout (mean difference –2.43, 95% CI –3.68 to –1.18), stronger beliefs regarding the severity of symptoms (mean difference –1.97, 95% CI –3.12 to –0.82), and a stronger emotional response to gout (mean difference –2.38, 95% CI –3.85 to –0.90) at follow-up. Participant feedback highlighted the importance of tracking health-related information, customizing to the target group/individual, providing more interactive features, and simplifying information.

Conclusions: Participants found the commercially available gout app more engaging. However, these findings did not translate into differences in self-care behaviors. The gout app group also demonstrated stronger negative illness perceptions at the follow-up.

Overall, these findings suggest that the development of gout apps would benefit from a user-centered approach with a focus on daily, long-term self-care behaviors as well as modifying illness beliefs.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12617001052325; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=373217.

(JMIR Mhealth Uhealth 2019;7(8):e15021) doi:10.2196/15021

KEYWORDS

mobile apps; mHealth; gout; chronic disease; user engagement; illness perceptions

Introduction

The potential to harness mobile technology for improving self-management in chronic disease is substantial, as reflected by the increase in the use of mobile health and medical apps (apps to promote health and manage illness) and their popularity [1,2]. Health apps are currently one of the fastest growing app categories, with over 100,000 apps available for Android and iOS platforms in 2015 [3]. With increasing industry and government investment in medical and health apps [3,4] as well as growing research into their efficacy [5,6], health and medical apps represent an exciting opportunity to improve health outcomes among people with chronic health conditions.

There are surprisingly few evidence-based apps that have been developed for patients with gout [7], even though gout affects approximately 4% of US adults and its rates are increasing worldwide [8,9]. Gout is a treatable condition wherein monosodium urate crystals deposit in the joints and periarticular tissues, causing inflammation and painful flares. Aotearoa/New Zealand has one of the highest rates of gout worldwide [9]; Maori (indigenous New Zealanders) and Pacific adults are disproportionally affected, with a prevalence of more than 8% [10]. Although treatable, nonadherence to effective urate-lowering therapies (ULT) has been reported to be as high as >50% [11-13]. The reasons for nonadherence are complex [13], but previous literature suggests that gout medications are often viewed unfavorably, with a preference for nondrug solutions including dietary strategies [14]. Previous literature also suggests that illness perceptions play a role in self-management behaviors in gout [15,16].

Mobile health (mHealth) interventions represent a promising approach to reducing barriers to care and potentially improving adherence to effective treatments for gout. In 2016, Nguyen and colleagues [7] reviewed commercially available apps available on iOS for managing gout and found that only one app met the recommendations set in patient-focused gout management guidelines. In 2017, we expanded this search to include Android smartphones and found a similar result [17].

In addition to reviewing the content and quality of apps for gout, it is important to examine users' experiences and engagement of apps. User engagement encompasses both how often and for how long people use apps as well as the user's experience of the technology as a whole [18]. Engagement is therefore believed to be closely tied to effectiveness of the intervention [19]. It is estimated that approximately 23% of app users delete an app after its first use [20]; however, user engagement is often not reported as part of mHealth interventions [21] and surprisingly few mHealth interventions examine the association between user engagement and health outcomes [19,22].

Due to these gaps in the literature, the goal of our study was to examine the experiences of using a commercially available gout self-management app compared to a dietary app among patients with gout, with a focus on assessing differences in user engagement, self-care behaviors, and illness beliefs. Using a randomized controlled trial design, we compared the gout self-management app identified in the previous reviews of gout apps [7,17] and a dietary app based on the Dietary Approaches to Stop Hypertension (DASH) diet, as the DASH diet is associated with a lower risk of gout [23]. We predicted that the participants allocated to the gout self-management app would demonstrate higher user engagement, self-care behaviors, and more adaptive illness beliefs at the 2-week follow-up compared to the active control group.

Methods

Design

Participants were randomized to either Gout Central (treatment group), a commercially available self-management app for gout, or the DASH Diet Plan (active control group), a commercially available app based on a diet for hypertension. Seventy-two adults with gout were recruited in Auckland and randomly allocated to one of the app groups between August 2017 and May 2018. Ethics approval was granted by the Health and Disability Ethics Committee (HDEC) in New Zealand (reference number 17/NTA/38), and all participants provided written informed consent. The trial was prospectively registered with the Australian New Zealand Clinical Trials Registry (registration number ACTRN12617001052325).

Participants and Randomization

Participants were recruited through posters in the community and from three primary and secondary clinics in Auckland, New Zealand. Inclusion/exclusion criteria included (1) a diagnosis of gout as defined by the 2015 ACR-EULAR Gout Classification Criteria [24], (2) age>18 years, (3) ability to complete forms in English and provide informed consent, and (4) ownership of or access to an android or IOS smartphone device capable of downloading apps. There were no restrictions regarding ethnic groups; however, a greater emphasis was placed on ensuring that Maori and Pacific peoples were recruited and retained in the study, as they are disproportionally affected by gout. The baseline assessment occurred face-to-face (using hard-copy questionnaires), and the follow-up assessments/questionnaires

were either completed as hard-copy questionnaires and returned by post or completed online.

App group allocations were generated by a biostatistician at the School of Medicine, independent of the intervention delivery. No stratification was used. Randomization occurred via sealed envelopes labelled with sequential study numbers. A research assistant recruited participants and assigned participants to interventions.

Intervention

After randomization, a research assistant helped participants download the app on their phones. The free versions of both apps were used in this study. Participants did not know which app was the intervention of interest/intervention group and which app was the control group.

Gout Central

The participants who were randomized to the treatment group were allocated to use the "Gout Central" app developed by the National Kidney Foundation for 2 weeks. This app includes information about gout and its causes, lifestyle tips for preventing gout flares, and treatment options and identifies common triggers that may cause flares. In addition, this app includes a series of health trackers such as the serum urate tracker and gout flare tracker in which the user can enter their details and track changes across time. Furthermore, this app allows users to enter their doctor appointments, log questions for their health care providers, log their medications and supplements used to manage gout and other conditions, and link them to online resources.

Participants in both app groups were advised to use all the functions they found helpful and were encouraged to use the app on a daily basis. They were also sent two text message reminders (two times during the 2-week trial period) to remind them to use the app on a daily basis. After completing the follow-up questionnaires, the participants were sent a voucher of NZ $50 to thank them for their time.

Dietary Approaches to Stop Hypertension Diet Plan

The active control group used the DASH diet plan app developed/sold by Chelin Apps (Android) and Diego Correa Bonini (iOS). This app provides information about the DASH diet eating plan, which has been shown to be effective in managing various health conditions such as gout [23]. The app educates users about the DASH diet, provides information to allow users to create a diet action plan, and informs users about what foods are beneficial and which ones should be avoided. This app also provides various recipes and meal ideas for breakfast, lunch, dinner, dessert, and snacks.

Measures

Demographic Data

Demographic data including sex, age, and ethnicity were collected at baseline via self-report questionnaires. Current alcohol use ("Do you drink alcohol?" Response: yes/no), smoking status ("Do you smoke?" Response: yes/no), and physical activity ("Do you get at least 30 minutes of physical activity per day, eg, brisk walking?" Response: yes/no) were also assessed at baseline. Disease duration, frequency of gout flares, comorbidities, serum urate, serum creatinine, and current gout treatments were collected via self-report or from participants' medical records. Previous and current app use (including use of health apps) was assessed in the baseline questionnaire. All the self-report questionnaires were administered online or as hard-copy questionnaires.

User Engagement

Our primary outcome measure was user engagement, measured using a modified user version of the Mobile Application Rating Scale (uMARS) [25] and administered after the 2-week trial. The original MARS was designed to allow app developers and health professionals to rate the quality of health apps, while the adapted uMARS scale was developed to allow users to rate health apps. The uMARS can be used to derive three separate scores: the objective app quality score, the subjective app quality score, and the perceived impact score. The objective app quality score is derived using four subscales: engagement, functionality, esthetics, and information quality. All items are rated on a 1-5 Likert scale, where 1 indicates that the app is inadequate in that domain and 5 indicates that the app is excellent in that domain. A total score is calculated by averaging across the four domains. In this study, only two of the subscales were used (engagement and information) to reduce participant burden. Therefore, only individual domain scores are presented for the two subscales without a total objective app quality score.

The subjective app quality score is derived from four items that examine overall user experience: "Would you recommend this app to people who might benefit from it?" "How many times do you think you would use this app in the next 12 months if it was relevant to you?" "Would you pay for the app?" and "What is your overall (star) rating of the app?" Lastly, the perceived impact score consists of six items that measure the impact of using the app on knowledge ("this app has increased my knowledge/understanding of the health behaviour"), attitude ("the app has changed my attitudes towards improving this health behaviour"), and behavior ("use of this app will increase/decrease the health behaviour"). The six items are reported as individual items and measured on a Likert scale ranging from 1 to 5 points (1="strongly disagree," 5="strongly agree"). Lastly, the uMARS has an open-ended question, "Do you have any further comments about the app?"

The uMARS demonstrates good internal reliability for both the instrument overall and the individual subscales within the instrument [25]. The Cronbach alpha coefficient for the engagement subscale score was 0.80, for the information subscale score was 0.60, and for the app subjective quality score was 0.84.

In addition to the uMARS, two additional questions were included to assess how often and for how long participants used the apps: "In the last 14 days, on how many days did you use the app?" and "on the days that you used the app, approximately how many minutes did you spend using the app?"

Self-Care Behaviors

Adherence to gout self-management guidelines and self-care behaviors were assessed with a self-report questionnaire that

covers behaviors related to gout management over the past 7 days. These items were adapted from the diabetes-specific Multidimensional Diabetes Questionnaire [26] and have been used previously to assess self-care behaviors in gout [27]. The items were individually scored and included behaviors such as medication (eg, "Over the last 7 days, how many of your prescribed number of gout pills [eg, allopurinol, probenecid, febuxostat or benzbromarone] did you take?"), rated on a 5-point Likert scale, from 1="none of them" to 5 ="all of them"; exercise (eg, "On how many of the last 7 days did you participate in at least 20 minutes of exercise?"), rated on an 8-point Likert scale, from 0-7 days; and diet-related activities (eg, "How often did you follow your recommended diet over the last 7 days?"), rated on a 5-point Likert scale, from 1="never" to 5="always."

Illness Perceptions

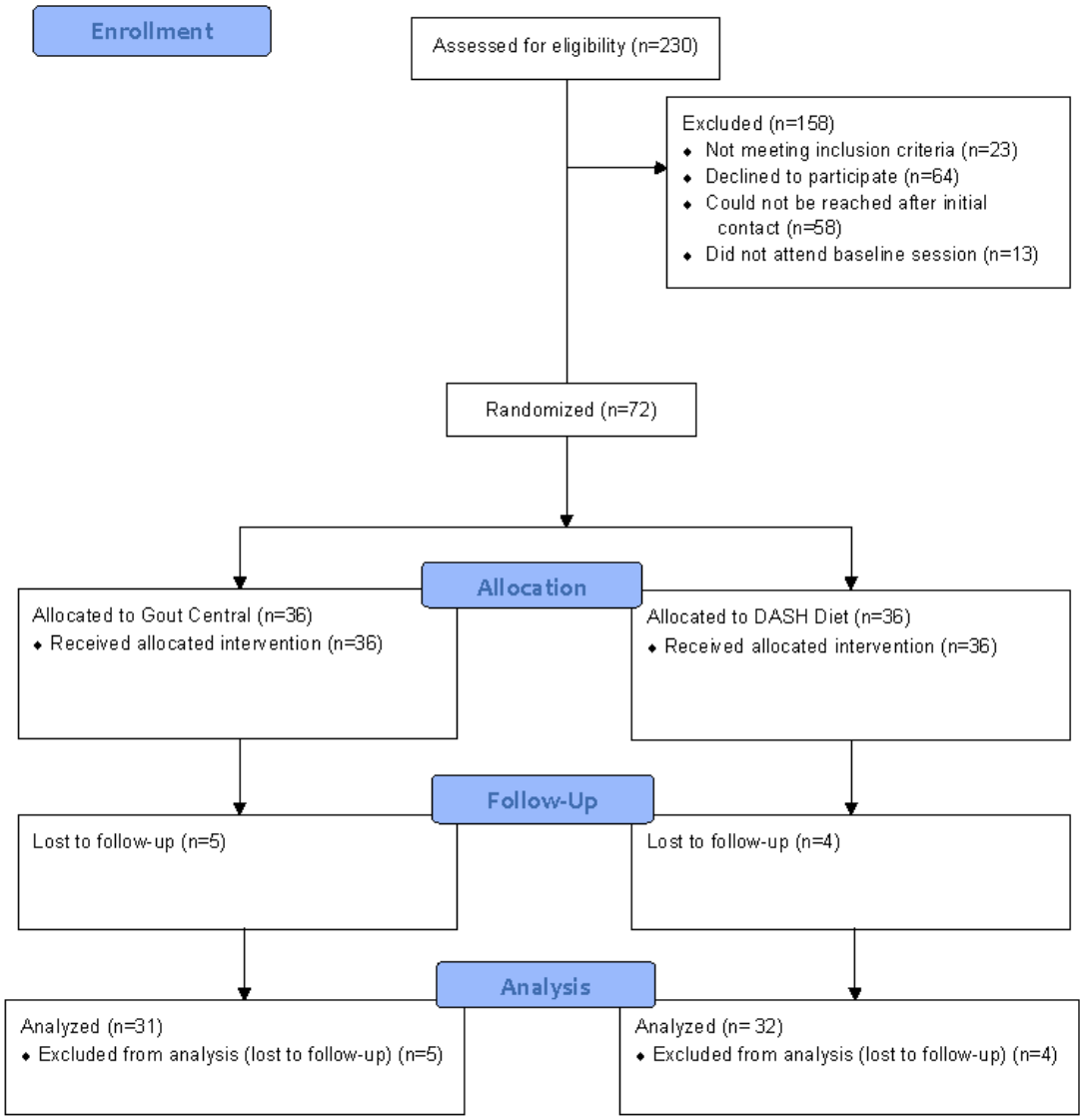
Illness perceptions were measured using a gout-specific Brief Illness Perceptions Questionnaire (B-IPQ) [28]. The 8-item B-IPQ examines the five key illness perception dimensions as well as items measuring the patient's concern, understanding, and emotional response to illness. The eight items are measured on a 0-10 Likert scale, with higher scores indicative of stronger endorsement: consequences (how much gout affects the patient's life), timeline (how long the patient thinks gout will continue), personal control (how much control the patient has over his or her gout), treatment control (how much the patient's medication can control gout), identity (severity of gout symptoms), concern (how concerned the patient is about his or her gout), understanding (how well the patient feels he/she understand their gout), and emotional response (how much gout affects the patient emotionally.) The B-IPQ has satisfactory reliability and validity across a range of chronic illnesses [28].

Power Calculation, Sample Size, and Statistical Analyses

Due to the lack of intervention studies that have examined differences in self-reported user engagement of health apps from which to estimate an effect size, we determined the sample size required to detect a medium effect size (0.6) between the two groups in user engagement (based on the uMARS). For the power calculation, we used an independent samples *t* test with 80% power and a 5% significance level, which indicated that 72 participants were required (36 in each group). To our knowledge, the only other study comparing user engagement scores (using the uMARS) across health apps used 5-6 participants per app group based on recommendations from the usability testing literature [29]. Our power calculation was based on recommendations for studies in which a standardized effect size is unknown [30].

Of the 72 participants, 9 were lost to follow-up and did not complete any of the follow-up questionnaires (Figure 1). The analysis therefore constituted a per protocol analysis. Independent samples *t* tests and analysis of covariance (ANCOVA) for variables that were unbalanced at baseline were used to examine differences between groups postintervention in user engagement, self-care behaviors, and illness perceptions. The results remained unchanged when controlling for baseline covariates; therefore, unadjusted means are reported. Means, SDs, and 95% CIs are presented with the analyses. Effect sizes were calculated using Cohen *d*, interpreted as <0.2 (small), 0.3-0.7 (medium), and 0.8 (large) [31].

Figure 1. Flow diagram of participant recruitment, randomization, and attrition. DASH: Dietary Approaches to Stop Hypertension.



Participant Feedback

Thirty-nine participants answered the open-ended question from the uMARS. The comments were analyzed independently by two researchers (AS and KS) using directed content analysis [32], appropriate where predetermined categories or concepts are being explored. After the initial stage of coding, the two researchers met to resolve any differences. The comments were initially grouped under three broad categories—positive experiences, negative experiences, and suggestions for improvement—with a frequency count for positive and negative comments. Participants' feedback was also mapped onto the four uMARS domains: Engagement, Functionality, Esthetics, and Information.

Results

During recruitment, 230 patients were contacted, of which 72 consented to participate and were randomized. The most common reasons for nonparticipation were not owning a smartphone or not wishing to participate in an mHealth intervention (Figure 1).

Baseline Measures

The mean (SD) age for the total sample at baseline (n=72) was 49 (15) years, and the majority of the participants were male (86%). Sixty percent were married, 17% were in a relationship, and 23% were single. Thirty-six percent identified as European New Zealanders; 25%, as Pasifika; 19%, as Maori; and 20%, as other ethnic groups. Regarding previous app use, most participants reported using apps (95.8%), with the number of apps used ranging from 0 to 125. In addition, 40% reported already using health apps. The majority of the participants reported drinking alcohol (73.6%), but not smoking (91.7%). A total of 69% reported being physically active. Baseline demographics were similar between the treatment and control groups (Table 1), except for age, whereby the participants in the DASH app group were older.

Table 1. Baseline demographic and clinical characteristics (N=72).

Characteristics	Gout Central (n=36)	DASH[a] Diet (n=36)
Age (years), mean (SD)	45 (14)	53 (15)
Sex (male), n (%)	30 (83)	32 (89)
Ethnicity, n (%)		
New Zealand European	14 (39)	12 (33)
Maori	5 (14)	9 (25)
Pacific	10 (28)	8 (22)
Other ethnic groups	7 (19)	7 (19)
Disease duration (years), mean (SD)	11 (13)	15 (10)
Number of flares in the past 3 months, mean (SD)	2.2 (5.1)	0.9 (1.7)
Serum urate (mmol/L) level, mean (SD)	0.41 (0.12)	0.35 (0.11)
Serum creatinine (μmol/L) level, mean (SD)	102 (46)	99 (52)
Current gout medications, n (%)		
Urate-lowering therapies[b]	20 (56)	29 (81)
Anti-inflammatory medications	16 (44)	11 (31)
Polypharmacy (≥5 long-term medications)	4 (11)	10 (28)
Comorbidities (five most prevalent), n (%)		
Hypertension	3 (8)	6 (17)
Type 2 diabetes	3 (8)	3 (8)
Kidney disease	4 (11)	1 (3)
Cardiovascular disease	2 (6)	5 (14)
Hypercholesterolemia	1 (3)	1 (3)

[a]DASH: Dietary Approaches to Stop Hypertension.

[b]Urate-lowering therapies: allopurinol and febuxostat; anti-inflammatory medications: colchicine, nonsteroidal anti-inflammatory drugs, and corticosteroids.

Regarding baseline clinical measures, the majority of the sample (68%) reported currently taking ULT, 38% reported taking anti-inflammatory medications, and an additional 19% reported being on more than five long-term medications. The mean serum urate level of the participants at baseline was 0.38 (0.11) mmol/L. In addition, 44% reported other comorbidities, with the most common reported as hypertension, type 2 diabetes, kidney disease, cardiovascular disease, and hypercholesterolemia. The clinical variables were similar between the treatment and control groups, except for the number of participants currently taking ULTs, which were higher in the control group (control: n=29 vs treatment: n=20; Table 1).

When examining illness perceptions at baseline in the two groups, the following beliefs were unbalanced: timeline, personal control, treatment control, and identity beliefs (Table 2). Self-care behaviors were similar at baseline between the two treatment arms.

Table 2. Mean differences in illness perceptions at baseline and follow-up.

Illness perceptions (score range 0-10)	Gout app, mean (SD)	DASH[a] app, mean (SD)	Mean difference (95% CI)	*P* value[b]
Consequences beliefs				.04
Baseline	4.50 (3.32)	4.11 (3.51)	–0.39 (–1.99 to 1.22)	
Follow-up	4.61 (2.63)	2.19 (2.36)	–2.43 (–3.68 to –1.18)	
Timeline beliefs				.70
Baseline	5.94 (3.54)	8.03 (2.90)	2.08 (0.56 to 3.61)	
Follow-up	6.81 (3.31)	7.19 (3.72)	0.38 (–1.40 to 2.16)	
Personal control beliefs				.28
Baseline	5.90 (3.02)	7.67 (2.43)	1.76 (0.48 to 3.05)	
Follow-up	6.90 (2.66)	7.56 (2.08)	0.66 (–0.54 to 1.86)	
Treatment control beliefs				.54
Baseline	7.44 (2.69)	8.92 (1.54)	1.47 (0.44 to 2.50)	
Follow-up	7.42 (3.00)	7.88 (2.81)	0.46 (–1.01 to 1.92)	
Identity beliefs				.001
Baseline	5.11 (2.94)	3.64 (3.09)	–1.47 (–2.89 to –0.06)	
Follow-up	4.06 (2.25)	2.09 (2.32)	–1.97 (–3.12 to –0.82)	
Concern beliefs				.26
Baseline	5.89 (3.07)	5.67 (3.43)	–0.22 (–1.75 to 1.31)	
Follow-up	5.26 (3.01)	4.34 (3.38)	–0.91 (–2.53 to 0.70)	
Understanding beliefs				.23
Baseline	7.44 (2.26)	8.03 (1.81)	0.58 (–0.38 to 1.55)	
Follow-up	8.13 (1.96)	7.34 (3.07)	–0.79 (–2.09 to 0.52)	
Emotional response beliefs				.002
Baseline	4.81 (3.46)	4.56 (3.83)	–0.25 (–1.97 to 1.47)	
Follow-up	5.00 (3.01)	2.63 (2.84)	–2.38 (–3.85 to –0.90)	

[a]DASH: Dietary Approaches to Stop Hypertension.

[b]*P* value refers to analysis of covariance for between-group comparisons postintervention.

Differences in User Engagement Between Groups Postintervention

With regard to the engagement and information subscale scores, participants rated the gout app as more engaging than the dietary app, with a mean difference of –0.58 (*P*=.003; effect size=0.77), and as more informative than the dietary app, with a mean difference of –0.34 (*P*=.04; effect size=0.53). Although the subjective app quality score (derived from four items that examine overall user experience) was also higher in the gout app group than in the dietary app, this difference was not statistically significant, with a mean difference of –0.36 (*P*=.11). Lastly, when evaluating the six individual items that examined perceived impact of the app, the only statistically significant differences were found in awareness of the importance of gout (mean difference of –0.64; *P*=.049; effect size=0.51) and knowledge/understanding of gout (mean difference of –0.70; *P*=.03; effect size=0.57), which were both higher in the gout app group than in the dietary app group. There was little evidence that either group used the app more during the 2-week trial (in terms of days used or minutes), with *P* values of .81 and .52, respectively (Table 3).

Table 3. Differences in user engagement between app groups postintervention measured by the user version of the Mobile Application Rating Scale (score range: 1-5).

Measures	Gout app (n=31)	DASH[a] app (n=32)	Mean difference (95% CI)	*P* value
Engagement subscale score	3.26 (0.73)	2.68 (0.77)	–0.58 (–0.96 to –0.21)	.003
Information subscale score	3.92 (0.51)	3.58 (0.76)	–0.34 (–0.67 to –0.01)	.04
Subjective app quality score	3.06 (0.82)	2.70 (0.94)	–0.36 (–0.81 to 0.08)	.11
Perceived impact: Awareness	3.42 (1.15)	2.78 (1.36)	–0.64 (–1.27 to –0.003)	.049
Perceived impact: Knowledge/understanding	3.32 (1.14)	2.63 (1.26)	–0.70 (–1.30 to –0.09)	.03
Perceived impact: Attitudes	3.19 (1.17)	2.59 (1.46)	–0.60 (–1.27 to 0.07)	.08
Perceived impact: Intention to change	3.06 (1.21)	2.56 (1.32)	–0.50 (–1.14 to 0.14)	.12
Perceived impact: Help seeking	2.84 (1.10)	2.59 (1.29)	–0.24 (–0.85 to 0.36)	.42
Perceived impact: Behavior change	3.10 (1.19)	2.59 (1.27)	–0.50 (–1.12 to 0.12)	.11
App use (days)	7.90 (3.95)	8.13 (3.27)	0.22 (–1.60 to 2.05)	.81
App use (minutes)	11.34 (13.00)	9.64 (6.91)	–1.70 (–6.92 to 3.52)	.52

[a]DASH: Dietary Approaches to Stop Hypertension.

Differences in Self-Care and Illness Perceptions Postintervention

Independent samples *t* tests and ANCOVA (adjusting for age, ULT, and baseline illness perceptions) were conducted to examine differences postintervention. There were no differences in any self-care behaviors between the two groups (*P*>.05) postintervention (results not tabulated). The gout app group demonstrated stronger negative beliefs regarding the impact of gout (mean difference 2.43; *P*=.04; effect size=0.97), stronger beliefs regarding the severity of symptoms (mean difference of –1.97; *P*=.001; effect size=0.86), and a stronger emotional response to gout (mean difference of –2.38; *P*=.002; effect size=0.81) at follow-up. None of the other illness beliefs demonstrated differences postintervention (Table 2).

Participant Feedback

The comments from the participants were grouped under three broad categories: positive experiences, negative experience, and suggestions for improvement. Textbox 1 presents a summary of the qualitative feedback, including how the feedback mapped onto the four domains of the uMARS (Engagement, Functionality, Esthetics, and Information).

Textbox 1. Summary of participants' positive/negative feedback and suggestions for improvement grouped according to the four domains of the user version of the Mobile Application Rating Scale (Engagement, Functionality, Esthetics, and Information).

Gout app

Engagement

Positive experiences:

- Tracking medications and urate levels
- Setting reminders
- Relevant for people who are newly diagnosed
- Useful during flare-ups

Negative experiences:

- Not useful for ongoing self-management of gout
- Lacking novelty
- Only useful if person has regular blood tests
- Not relevant to New Zealanders and different ethnic groups
- Not adapted for people with low health literacy

Suggestions for improvement:

- Provide more information for Pacific peoples
- Provide more videos and interactive features
- Add links to healthcare team

Functionality

Suggestions for improvement:

- Ability to enter more data
- Ability to track foods

Esthetics

Suggestions for improvement:

- Include more visual content

Information

Suggestions for improvement:

- Graphing urate levels
- Add links to other sources of information

Dietary Approaches to Stop Hypertension

Engagement

Negative experiences:

- Not motivating
- Not relevant to New Zealanders
- Not relevant to different cultural and ethnic groups
- Not relevant to gout

Suggestions for improvement:

- Customizing/tailoring app to the individual
- Provide social support
- More interactive content

Functionality

Negative experiences:

- Hard to use

Suggestions for improvement:

- Make it more simple

Esthetics

- Busy/confusing interface

Information

Positive experiences:

- Good recipes

Negative experiences:

- Poor dietary advice
- Not adapted for people with lower health literacy

Suggestions for improvement:

- Tracking/logging information

Positive Experiences

Positive feedback from the gout app group (n=10) reflected their satisfaction with the app's ability to track health information, set reminders, and monitor self-care behaviors.

> *I particularly liked being able to track medicines and set reminders; look at graphs showing urate levels; list all my meds in one place; and be reminded to drink more water.* [Participant #66, Gout app]

In contrast, only two respondents in the DASH app group (n=2) provided positive feedback, both based on the information and dietary content of the app.

> *Very good app, very informative. Great recipes.* [Participant #40, DASH app]

Negative Experiences

Feedback from the gout app group suggested that the app was better suited to people who were newly diagnosed or patients with frequent flare-ups, rather than for long-term self-management.

> *I have only had a gout attack 3 times in the last 10 years. I consider that not having gout attacks regularly negates the use off the app and is not much help in my case. Also it is USA of origin and some parts are not much use in NZ.* [Participant #10, Gout app]

Other negative feedback from both app groups (n=3 for the Gout app and n=8 for the DASH app) included the lack of customization to New Zealand or lack of tailoring to ethnic groups more prone to gout in New Zealand such as Maori or Pacific Peoples.

> *Should be more customisable...Have current or relevant stats applicable to Pacific People. Go to gout sites or locations for help.* [Participant #7, Gout app]

Suggestions for Improvement

Both app groups suggested the need for more interactive features such as video content and links to health professionals or other patients.

> *For millennial users, it would be better to put more interactive features such as a link to videos explaining gout in a visual way...it would be better to not just link the app to our personal healthcare team but to have an online chatting with other available healthcare personnel (possibly to create a series of interactive Q&A with some available healthcare personnel).* [Participant #9, Gout app]

> *Wouldn't it be nice to have an interactive function with other users and create a network that allows people to connect and support each other.* [Participant #19, DASH diet app]

Other suggestions for improvement included simplifying information, especially for people with low health literacy or non-English speakers.

> *This app also has very useful information that is only good for those who understand the jargon used. If I have no clue about the language used I'm not going to spend time trying to decipher the information. This part of the app needs to be simplified and made fun for all people and not just an educated few.* [Participant #55, DASH diet app]

> *The biggest challenge for the use of this app would be literacy both in Health and English. People with*

English as a second language would struggle. [Participant #58, Gout app]

Discussion

This is the first study to examine the impact of a commercially available gout app on health outcomes. This is also the first study to examine the utility of a commercially available gout app from the patient's perspective, by examining user engagement. Our primary hypothesis was largely confirmed by the findings, with the gout self-management app demonstrating higher user engagement scores than the dietary app. However, this did not translate to improvements in self-care behaviors. The findings regarding illness beliefs at follow-up were more mixed, with the gout self-management app associated with stronger consequence beliefs, illness identity beliefs, and emotional response beliefs than the dietary app.

Several possible reasons exist to explain why higher user engagement did not translate to improvements in self-care behaviors. First, despite higher uMARS scores, there was no difference in time that participants spent on or used to access the gout app compared with the dietary app during the study. The average number of days spent using the app (of 14 days) for both groups was only 8. Both the apps in this study provided no feedback or goal-setting functions, thus requiring the user to be intentional about their app usage and access it without personalized feedback or any specific behavior change strategies. Second, as suggested by the participant feedback, it is possible that participants determined that the gout app (Gout Central) would only be helpful during a gout flare and thus used it less. This explanation may provide insight into why there was no difference in usage time between the two apps and no changes in self-care behaviors. Even though the gout app was chosen because it was the best available app for gout [7,17], it may not provide all the appropriate tools necessary to manage gout when it is asymptomatic. Therefore, Gout Central may not meet the needs of users in terms of continuous self-management and care of the condition between flares.

A third possible explanation is that there was little integration with daily self-management behaviors in gout. Many existing health apps focus on providing educational content, basic health monitoring, or various reminders [33], but fail to fully utilize the unique capabilities of smartphone technology (eg, real-time data collection and data visualization technology) [34]. For gout, specifically, this may mean daily medication reminders, flare diaries, visualizing serum urate fluctuations, and food diary capabilities, which encourage a user to engage with their self-care regardless of whether they are having a flare. Real-time data tracking would be especially desirable for people who are testing their serum urate levels using commercially available test meters that provide an immediate result, which would then allow for real-time serum urate tracking. Even though the gout app provided a wealth of information about gout as well as some opportunities for data tracking, real-time reminders and real-time data tracking were both missing.

Our findings regarding illness perceptions were more mixed. The more engaging gout app resulted in stronger negative illness perceptions about gout at follow-up, including stronger beliefs regarding the negative impact of having gout, the severity of symptoms, and the emotional response to gout. It is possible that using the gout app reminded participants about the negative aspects of their illness or the seriousness of their illness, which, over the long-term, may positively impact self-care behaviors. For example, in a study of patients with kidney disease, having stronger illness identity beliefs (ie, beliefs regarding severity of symptoms) was associated with more proactive coping [35]. On the other hand, we have previously found that stronger consequence beliefs and emotional response beliefs are associated with increased disability and mortality in gout [15,16]. It is also possible that negative illness beliefs (ie, more pessimistic beliefs about gout) may impact the app use itself, as participants may not want to be reminded that they have gout. Due to the short follow-up in this study, we cannot confirm either of these possibilities or whether the change in illness perceptions will be sustained or influence self-care behaviors in the long-term.

Targeting illness perceptions presents another promising area for future research in mHealth interventions. Only a handful of mHealth interventions have attempted to modify illness perceptions [36-38], with promising effects on outcomes including medication adherence as well as objective health outcomes. Considering that many commercially available health apps lack theory or evidence-based behavioral strategies [39-41], this could be an important area for future research.

The open-ended participant feedback largely supported our quantitative results, with more positive comments recorded overall for the Gout Central app than the DASH diet app. The participant feedback also highlighted that gout patients want apps that allow them to track health-related information, that are customized to their needs, that provide interactive features, and that are simple to use.

This study had many strengths: It is the first study to examine the efficacy of a commercially available gout app on self-reported health outcomes and to consider the association between user engagement and health outcomes in gout. It also successfully recruited a diverse group of patients of whom 43% were Maori or Pasifika. However, there are several limitations that should be noted. As we did not develop the apps, we were also unable to objectively examine user engagement and had to rely on self-report. Furthermore, after data collection was completed for this study, the gout app Gout Central was updated; therefore, these results may not reflect the most up-to-date version of the app. Another limitation was the short follow-up duration. Despite this short time period, there was some loss to follow-up, which necessitated a per protocol analysis rather than an intention-to-treat analysis. Lastly, these findings may not be generalizable to other populations.

The findings from this short-term trial demonstrate the many challenges associated with not only developing engaging and effective apps for patients with chronic health conditions but also the challenges present in testing commercially available apps. Our findings suggest that, at least over a period of 2 weeks, using the gout app was more engaging than using a general dietary app for patients with gout, but that the gout app is unlikely to change self-care behaviors in the short-term. It is

becoming increasingly clear in the mHealth literature that the design process needs to be user-centered in order to produce the best outcomes for patients [42-44]. Work in this area is currently underway by Nguyen and colleagues [45], who have recently developed a gout self-management app. User-centered design has been tested in apps for type 1 diabetes, heart health, and asthma and has shown favorable outcomes in the preliminary stages [44]. Adopting a user-centered development approach would likely result in higher-quality apps, which are useful and accessible [46].

In conclusion, working with gout patients to develop an app that best suits their needs, targets daily self-management behaviors in between flares with specific behavior change strategies, and modifies illness beliefs could be a promising way forward in the use of mHealth to manage gout.

Acknowledgments

We thank all the participants who took part in our study. AS was supported by the University of Auckland's Faculty Research Development Fund grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2MB - mhealth_v7i8e15021_app1.pdf]

References

1. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
2. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
3. Research2Guidance. Internet. mHealth App Developer Economics 2014: The State of the Art of mHealth App Publishing URL: http://mhealtheconomics.com/mhealth-developer-economics-report/ [accessed 2019-10-20]
4. FDA. 2016. Mobile Medical Applications URL: http://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/default.htm [accessed 2019-10-20]
5. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
6. 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]
7. Nguyen AD, Baysari MT, Kannangara DRW, Tariq A, Lau AYS, Westbrook JI, et al. Mobile applications to enhance self-management of gout. Int J Med Inform 2016 Oct;94:67-74. [doi: 10.1016/j.ijmedinf.2016.06.021] [Medline: 27573313]
8. Zhu Y, Pandya BJ, Choi HK. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. Arthritis Rheum 2011 Oct;63(10):3136-3141. [doi: 10.1002/art.30520] [Medline: 21800283]
9. Kuo CF, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nat Rev Rheumatol 2015;11(11):649-662. [doi: 10.1038/nrrheum.2015.91] [Medline: 26150127]
10. Dalbeth N, Dowell T, Gerard C, Gow P, Jackson G, Shuker C, et al. Gout in Aotearoa New Zealand: the equity crisis continues in plain sight. N Z Med J 2018 Dec 09;131(1485):8-12. [Medline: 30408813]
11. Singh JA. Challenges faced by patients in gout treatment: a qualitative study. J Clin Rheumatol 2014 Apr;20(3):172-174 [FREE Full text] [doi: 10.1097/RHU.0000000000000091] [Medline: 24662562]
12. Scheepers LEJM, van Onna M, Stehouwer CDA, Singh JA, Arts ICW, Boonen A. Medication adherence among patients with gout: A systematic review and meta-analysis. Semin Arthritis Rheum 2018 Dec;47(5):689-702. [doi: 10.1016/j.semarthrit.2017.09.007] [Medline: 29198878]
13. De Vera MA, Marcotte G, Rai S, Galo JS, Bhole V. Medication adherence in gout: a systematic review. Arthritis Care Res (Hoboken) 2014 Oct;66(10):1551-1559 [FREE Full text] [doi: 10.1002/acr.22336] [Medline: 24692321]
14. Derksen C, Serlachius A, Petrie KJ, Dalbeth N. "What say ye gout experts?" a content analysis of questions about gout posted on the social news website Reddit. BMC Musculoskelet Disord 2017 Nov 23;18(1):488 [FREE Full text] [doi: 10.1186/s12891-017-1856-y] [Medline: 29169350]

15. Serlachius A, Gamble G, House M, Vincent ZL, Knight J, Horne A, et al. Illness Perceptions and Mortality in Patients With Gout: A Prospective Observational Study. Arthritis Care Res (Hoboken) 2017 Dec;69(9):1444-1448 [FREE Full text] [doi: 10.1002/acr.23147] [Medline: 27813363]
16. Dalbeth N, Petrie KJ, House M, Chong J, Leung W, Chegudi R, et al. Illness perceptions in patients with gout and the relationship with progression of musculoskeletal disability. Arthritis Care Res (Hoboken) 2011 Nov;63(11):1605-1612. [doi: 10.1002/acr.20570] [Medline: 22034122]
17. Kieser A, Dalbeth N, Serlachius A. Keeping Up with the Applications: Lessons Learned Evaluating Gout Apps. Telemed J E Health 2019 Apr;25(4):272-273. [doi: 10.1089/tmj.2018.0091] [Medline: 29969385]
18. O'Brien H, Toms E. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr 01;59(6):938-955 [FREE Full text] [doi: 10.1002/asi.20801]
19. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
20. Localytics - Blog. App User Retention Improves in the US URL: http://info.localytics.com/blog/app-user-retention-improves-in-the-us [accessed 2019-10-20]
21. Mulvaney SA, Ritterband LM, Bosslet L. Mobile intervention design in diabetes: review and recommendations. Curr Diab Rep 2011 Dec;11(6):486-493. [doi: 10.1007/s11892-011-0230-y] [Medline: 21960031]
22. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
23. Rai SK, Fung TT, Lu N, Keller SF, Curhan GC, Choi HK. The Dietary Approaches to Stop Hypertension (DASH) diet, Western diet, and risk of gout in men: prospective cohort study. BMJ 2017 May 09;357:j1794. [doi: 10.1136/bmj.j1794] [Medline: 28487277]
24. Neogi T, Jansen TLTA, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2015 Oct;74(10):1789-1798 [FREE Full text] [doi: 10.1136/annrheumdis-2015-208237] [Medline: 26359487]
25. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
26. Talbot F, Nouwen A, Gingras J, Gosselin M, Audet J. The assessment of diabetes-related cognitive and social factors: the Multidimensional Diabetes Questionnaire. J Behav Med 1997 Jun;20(3):291-312. [Medline: 9212382]
27. Dalbeth N, House ME, Horne A, Petrie KJ, McQueen FM, Taylor WJ. Prescription and dosing of urate-lowering therapy, rather than patient behaviours, are the key modifiable factors associated with targeting serum urate in gout. BMC Musculoskelet Disord 2012 Sep 16;13:174 [FREE Full text] [doi: 10.1186/1471-2474-13-174] [Medline: 22978848]
28. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006 Jun;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
29. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the Quality of Mobile Phone Apps for Weight Management: User-Centered Study With Employees From a Lebanese University. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836 [FREE Full text] [doi: 10.2196/mhealth.9836] [Medline: 30672742]
30. 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]
31. Cohen J. Statistical power analysis for the behavioral sciences. New York: Routledge; 1988:1134742703.
32. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
34. Liu C, Zhu Q, Holroyd K, Seng E. Status and trends of mobile-health applications for iOS devices: A developer's perspective. Journal of Systems and Software 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049] [Medline: 30872122]
35. Vélez-Vélez E, Bosch RJ. Illness perception, coping and adherence to treatment among patients with chronic kidney disease. J Adv Nurs 2016 Apr;72(4):849-863. [doi: 10.1111/jan.12873] [Medline: 26689295]
36. Petrie KJ, Perry K, Broadbent E, Weinman J. A text message programme designed to modify patients' illness and treatment beliefs improves self-reported adherence to asthma preventer medication. Br J Health Psychol 2012 Feb;17(1):74-84. [doi: 10.1111/j.2044-8287.2011.02033.x] [Medline: 22107110]
37. Perera AI, Thomas MG, Moore JO, Faasse K, Petrie KJ. Effect of a smartphone application incorporating personalized health-related imagery on adherence to antiretroviral therapy: a randomized clinical trial. AIDS Patient Care STDS 2014 Nov;28(11):579-586 [FREE Full text] [doi: 10.1089/apc.2014.0156] [Medline: 25290556]

38. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. BMJ 2018 Dec 17;361:k1959 [FREE Full text] [doi: 10.1136/bmj.k1959] [Medline: 29773539]
39. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
40. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
41. 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]
42. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
43. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth Interventions With Children and Young People. Front Psychiatry 2018;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
44. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
45. Nguyen AD, Frensham LJ, Wong MX, Meslin SM, Martin P, Lau AY, et al. mHealth App Patient Testing and Review of Educational Materials Designed for Self-Management of Gout Patients: Descriptive Qualitative Studies. JMIR Mhealth Uhealth 2018 Oct 15;6(10):e182 [FREE Full text] [doi: 10.2196/mhealth.9811] [Medline: 30322835]
46. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]

Abbreviations

ANCOVA: analysis of covariance
B-IPQ: Brief Illness Perceptions Questionnaire
DASH: Dietary Approaches to Stop Hypertension
HDEC: Health and Disability Ethics Committee
mHealth: mobile health
uMARS: user version of the Mobile Application Rating Scale
ULT: urate-lowering therapies

Edited by G Eysenbach; submitted 13.06.19; peer-reviewed by A Nguyen, R Day; comments to author 02.07.19; revised version received 04.07.19; accepted 04.07.19; published 13.08.19.

Please cite as:
Serlachius A, Schache K, Kieser A, Arroll B, Petrie K, Dalbeth N
Association Between User Engagement of a Mobile Health App for Gout and Improvements in Self-Care Behaviors: Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(8):e15021
URL: http://mhealth.jmir.org/2019/8/e15021/
doi:10.2196/15021
PMID:31411147

Original Paper

Relationship Between Adherence to Remote Monitoring and Patient Characteristics: Observational Study in Women With Pregnancy-Induced Hypertension

Thijs Vandenberk[1,2,3,4,5*], MSc, PhD; Dorien Lanssens[1,2,3,4,5*], MSc, PhD; Valerie Storms[1,2,3,4], MSc, PhD; Inge M Thijs[1,2,3,4,5], MSc, PhD; Lotte Bamelis[6], MSc, PhD; Lars Grieten[1,2,3,4,5], MSc, PhD; Wilfried Gyselaers[1,2,3,4,5], MD; Eileen Tang[7], MSc, PhD; Patrick Luyten[7,8], MSc, PhD

[1]Faculty of Medicine and Life Sciences, Hasselt University, Diepenbeek, Belgium

[2]Department of Obstetrics & Gynecology, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Future Health, Ziekenhuis Oost-Limburg, Genk, Belgium

[4]Limburg Clinical Research Center, Hasselt University, Diepenbeek, Belgium

[5]Mobile Health Unit, Hasselt University, Diepenbeek, Belgium

[6]Centre for Translational Psychological Research TRACE, Ziekenhuis Oost-Limburg, Genk, Belgium

[7]Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Leuven, Belgium

[8]Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Thijs Vandenberk, MSc, PhD
Faculty of Medicine and Life Sciences
Hasselt University
Agoralaan
Diepenbeek, 3590
Belgium
Phone: 32 11268111
Email: thijsvandenberk@hotmail.com

Abstract

Background: Pregnancy-induced hypertension (PIH) is associated with high levels of morbidity and mortality in mothers, fetuses, and newborns. New technologies, such as remote monitoring (RM), were introduced in 2015 into the care of patients at risk of PIH in Ziekenhuis Oost-Limburg (Genk, Belgium) to improve both maternal and neonatal outcomes. In developing new strategies for obstetric care in pregnant women, including RM, it is important to understand the psychosocial characteristics associated with adherence to RM to optimize care.

Objective: The aim of this study was to explore the role of patients' psychosocial characteristics (severity of depression or anxiety, cognitive factors, attachment styles, and personality traits) in their adherence to RM.

Methods: Questionnaires were sent by email to 108 mothers the day after they entered an RM program for pregnant women at risk of PIH. The Generalized Anxiety Disorder Assessment-7 and Patient Health Questionnaire-9 (PHQ-9) were used to assess anxiety and the severity of depression, respectively; an adaptation of the Pain Catastrophizing Scale was used to assess cognitive factors; and attachment and personality were measured with the Experiences in Close Relationships-Revised Scale (ECR-R), the Depressive Experiences Questionnaire, and the Multidimensional Perfectionism Scale, respectively.

Results: The moderate adherence group showed significantly higher levels of anxiety and depression, negative cognitions, and insecure attachment styles, especially compared with the over adherence group. The low adherence group scored significantly higher than the other groups on other-oriented perfectionism. There were no significant differences between the good and over adherence groups. Single linear regression showed that the answers on the PHQ-9 and ECR-R questionnaires were significantly related to the adherence rate.

Conclusions: This study demonstrates the relationships between adherence to RM and patient characteristics in women at risk of PIH. Alertness toward the group of women who show less than optimal adherence is essential. These findings call for further research on the management of PIH and the importance of individual tailoring of RM in this patient group.

Trial Registration: ClinicalTrials.gov NCT03509272; https://clinicaltrials.gov/ct2/show/NCT03509272

(JMIR Mhealth Uhealth 2019;7(8):e12574) doi:10.2196/12574

KEYWORDS

remote monitoring; gestational hypertensive diseases; monitoring, ambulatory; hypertension, pregnancy-induced; surveys and questionnaires; treatment adherence and compliance

Introduction

Pregnancy-induced hypertension (PIH), which is a complication in 6% to 10% of pregnancies, is defined as a systolic blood pressure (BP)>140 mmHg and diastolic BP>90 mmHg. PIH refers to 1 of 4 conditions: (1) pre-existing hypertension, (2) gestational hypertension, (3) pre-eclampsia, and (4) unclassifiable hypertension [1]. It is a major cause of maternal, fetal, and neonatal morbidity and mortality [1,2]. The assessment of women with pregnancies complicated by PIH includes clinical follow-up, serological investigations, and fetal ultrasound. The type and frequency of follow-up depends on the kind and severity of the hypertensive disorder [1]. The goal of treatment is to prevent significant cerebrovascular and cardiovascular events in the mother, without compromising fetal well-being [3].

New techniques to support these strategies have recently been developed, including remote monitoring (RM), which can be broadly defined as the use of telecommunication technologies to facilitate the transmission of medical information and services between health care providers and patients [4]. RM is a relatively new approach (dating back to the early 1990s) that facilitates patient management at home [5]. As part of the Hasselt University and Limburg Clinical Research Program, Ziekenhuis Oost-Limburg (ZOL, Genk, Belgium), a large hospital in the east of Belgium, added RM to the prenatal care of women with PIH. All women diagnosed with PIH who delivered at the outpatient prenatal clinic of ZOL were included. Women received RM on demand of the responsible obstetrician before admission or after discharge from the prenatal ward. The criteria to initiate RM were PIH at gestational age more than 12 weeks where an intensive follow-up until delivery was desirable. Women without a mobile phone, a gestational age less than 12 weeks, a fetus with congenital malformations, and women who refused informed consent were excluded and received conventional care. Women consenting for RM were asked to perform 2 BP measurements a day with the iHealth Blood Pressure Monitor (iHealth Feel), fill in once a week their weight on the app, and to wear continuously an iHealth activity tracker (iHealth Wave; iHealth, Paris, France). The data from the monitor devices were transmitted to a Web-based dashboard developed by the Mobile Health Unit of the University of Hasselt and the Future Health Department of ZOL. Predetermed alarm signals were set; 1 midwife performed remote follow-up of all transformed data at the dashboard. Alarm events were communicated with the obstetrician in charge to discuss management options before contacting and instructing patients at home. Therapeutic interventions were according to local management. Our first results were promising, suggesting that the addition of RM to the prenatal care protocol for women at risk of gestational hypertensive disease reduces prenatal hospitalization (until the moment of delivery), inductions, and pre-eclampsia compared with the levels in women who receive conventional care. Furthermore, it is likely that women monitored with RM will enter labor spontaneously and will be more likely to be diagnosed with gestational hypertension rather than pre-eclampsia than women treated with conventional care [6]. RM has also been effective in the follow-up of pregnant women with issues such as problematic BP and bodyweight [7,8]. However, adherence to RM is an important concern. Several studies have reported low rates of adherence to RM [9-11]. Adherence refers to the extent to which a patient follows a prespecified treatment regimen or protocol [12]. The methods used to measure treatment adherence are either direct or indirect. Direct methods include observation and the assessment of metabolites or biological markers in the blood. Indirect methods include self-report questionnaires, pill counts, rate of prescription refilling, and the clinical assessment of patients' physiological markers [13,14].

In developing new strategies (including RM) to optimize the obstetric care of pregnant women, it is important to investigate the patients' characteristics, as these will potentially affect their adherence to RM. The peripartum period has long been known to be associated with increased levels of stress and anxiety, related to the transition to parenthood and parental tasks and concerns associated with this transition [15]. However, when PIH is present, it potentially increases the already elevated levels of stress and anxiety associated with this normative and normally adaptive heightened "maternal preoccupation" in the perinatal period [16,17]. In this context, cognitive factors, such as catastrophizing or rumination, and insecure attachment styles and personality factors, such as perfectionism and dependency, have been associated with problems in negotiating the challenges of parenthood, which are expressed as increased levels of anxiety and depression [18-20]. Mothers with a tendency to catastrophize, for instance, might be excessively worried about PIH and might, therefore, show either poor adherence to RM because they wish to avoid potentially threatening information or alternatively, they might engage excessively in RM, and RM might become an obsession for them. Individuals with avoidant attachment styles and high levels of self-critical perfectionism might show a similar pattern of avoidance or over adherence, whereas those with anxious attachment or dependent personality traits might become overly compliant with RM. Poor medical outcomes and higher mortality rates in individuals with attachment avoidance and self-critical perfectionism have been associated with a tendency to deny health problems and compulsive autonomy (the belief that one must be able to manage one's problems on one's own) [21]. In contrast, individuals with high levels of attachment anxiety and dependency traits typically seek help readily, typically leading

to better health outcomes (eg, earlier detection of cancer), but also with excessive use of medical care [22-24]. Regarding perfectionism, different dimensions have been discerned [25]: self-oriented perfectionism refers to having high personal standards and the need to constantly live up to these high standards, whereas other-oriented perfectionism refers to expecting perfection and high performance from others; finally, socially prescribed perfectionism refers to a constant striving to live up to others' high standards and expectations. Individuals with high levels of self-oriented or socially prescribed perfectionism might show excessive adherence to RM, whereas individuals with increased levels of other-oriented perfectionism might show low adherence because of a skeptical attitude toward others and the RM program in particular.

To the best of our knowledge, no research to date has examined the relationships between adherence to RM and patient characteristics. Therefore, the primary endpoint of this study was to explore the roles of depression and anxiety, cognitive factors, and attachment and personality traits in relation to adherence to RM. On the basis of the findings discussed above, we expected that anxiety and depression, and cognitive factors, such as rumination and catastrophizing, would be increased in low and excessively adherent mothers. Similarly, we expected high levels of attachment avoidance and self-critical perfectionism to be associated with low or over adherence. Furthermore, we hypothesized high levels of attachment anxiety and dependent personality features to be related to over adherence. Finally, we expected high levels of self-oriented or socially prescribed perfectionism to be associated with over adherence, whereas high levels of other-oriented perfectionism were hypothesized to be related to low adherence. The secondary endpoint of the study was the relation between the individual questionnaire and the adherence rate.

Methods

Study Protocol

This study is part of the Pregnancy Remote Monitoring (PREMOM) study, an observational study involving 8 hospitals in Limburg (Belgium), undertaken to optimize gestational outcomes in pregnancies complicated with PIH. The PREMOM protocol and main results have been reported elsewhere [6,26,27]. Briefly, women consenting to RM underwent obstetric surveillance with a wireless BP monitor and an activity tracker. They were asked to make 1 BP measurement in the morning and 1 in the evening, to enter 1 weight measurement weekly and to wear the activity tracker 24 hours a day until delivery or hospital admission. When alarm signals were detected (systolic BP>140 mmHg, diastolic BP>90 mmHg, or weight gain>1 kg/day) by the responsible midwife, the obstetrician-in-charge was contacted to discuss the management options before the patient was contacted at home. The types of interventions were (1) expectant management, (2) ambulatory blood sampling and 24-hour urine collection at home, (3) adjustment of antihypertensive therapy and/or physical activity, (4) admission to the prenatal ward, or (5) induction of labor. The therapeutic interventions were based on local management strategies.

Pregnant women were given information about the study at the start of their RM program. All the women provided written informed consent to participate in the study. The Ziekenhuis Oost-Limburg Medical Ethics Committee approved the study.

The characteristics of the participants were collected at inclusion in the PREMOM program. Demographic and obstetric information was collected at recruitment and after delivery from the hospital administration and/or billing records.

All participants received an email containing a SurveyMonkey link. After logging in, the participants were asked to complete 6 questionnaires (see subsection Questionnaires).

Participants

A total of 124 mothers from the PREMOM study were invited to participate in this study. A total of 7 (5.65%) of them declined participation because of lack of interest. Of the remaining 117 pregnant women, 7 (5.98%) were hospitalized in the prenatal ward with complications before they could complete the questionnaires. In total, 110 pregnant women (88.71%) completed the questionnaires, 2 of whom were excluded from the final analysis because their data were invalid because of failure to fill out the questionnaires correctly.

Questionnaires

The Generalized Anxiety Disorder-7 (GAD-7) assessment scale was used to assess anxiety. It consists of 7 items to be rated on a 4-point Likert scale ranging from 0 to 3. The Patient Health Questionnaire-9 (PHQ-9) was used to measure the severity of depression. This questionnaire consists of 9 items to be rated on a 4-point Likert scale ranging from 0 to 3. The Pain Catastrophizing Scale assesses painful experiences and indicators of negative thoughts. It consists of 13 items to be scored on a balanced 5-point Likert scale ranging from 1 to 5. This questionnaire was adapted by the research team to include pregnancy-related questions. Anxious and avoidant attachment styles were measured by the 36 items from the Experiences in Close Relationships-Revised Scale (ECR-R), to be rated on a 7-point Likert scale ranging from 1 to 7. The Depressive Experiences Questionnaire for Adolescents was used to assess self-criticism and dependency, consisting of 20 items to be scored on a 7-point Likert scale ranging from 1 to 7. Finally, the Multidimensional Perfectionism Scale measures self-oriented, other-oriented, and socially prescribed dimensions of perfectionism, using 45 items to be rated on a balanced 7-point Likert scale ranging from 1 to 7. For all 6 questionnaires, higher scores indicate higher levels of anxiety, depression, cognitive, attachment, or personality traits of interest.

Adherence

Patients' adherence to their scheduled daily measurements was determined by tracking the total number of scheduled events and then counting the actual number of measurements made. This from the moment of inclusion, until 90 days later. A total of 180 measurements was expected: 90 days x 2 measurements a day. The adherence rate was calculated as follows: number of measurements made/180 potential measurements × 100%. This ratio provides a robust measure of adherence. With this formula, adherence ranges between 0% (in case the patient did

not make any measurement during her pregnancy) to over 100% (in case the pregnant woman performed more than 2 BP measurements a day). The study population of pregnant women was, in discussion with midwives and gynecologists, divided into 4 study groups: (1) Those with an adherence rate<30% (low adherence). An adherence rate below 30% is really insufficient in the follow-up of the pregnant women. When the last blood measurement of a women remains far from the clinical threshold (90 mmHg Diastolic or 140 mmHg Systolic BP), it is not critical to receive only 1 measurement in 2 days (corresponds with 30%). Whereas 30% adherence rate is way too low when the currently (and last) BPs where elevated; (2) Those with an adherence rate of 30% to 80% (moderate adherence); (3) Those with an adherence rate of 80% to 100% (good adherence); and (4) Those with an adherence rate>100% (overadherence). A large group of women seemed to be really motived to adherence to the monitoring program, and the bulk of these women seem to fall within the 30 to 80% adherence rate. In discussions with midwifes and gynecologists, women with adherence rates between 80 to 100% seemed to be a highly motivated group, whereas those with adherence rates>100% were considered to be perhaps overly anxious and concerned. An additional analysis was performed based on equal sample size groups. This analysis can be found in Multimedia Appendix 1.

Statistical Analysis

Data were analyzed with the RStudio version 3.2.2 (RStudio Inc) statistical software. The Shapiro-Wilk test was used to assess whether the data were normally distributed. Nonparametric tests were used when the normality assumption was violated. Non-normally distributed data are expressed as medians and interquartile ranges (IQRs). Analysis of variance was used to test within-group comparisons. An independent *t* test (parametric) and/or the Mann-Whitney *U* test (nonparametric) was used for between-group comparisons. Single linear regression was performed to determine which personal characteristics had a significant relation with the adherence rate. *P* values <.05 were considered statistically significant.

Results

Participant Demographic and Obstetric Characteristics

In total, 108 participants completed the questionnaires. The patient demographic and obstetric characteristics are presented in Table 1. In the total sample, the median adherence was 89.4% (IQR: 54.7-103.3), the median age was 30 years (IQR: 28-33), the median prepregnancy weight was 76 kg (IQR: 66-91), the mean height was 167 cm (SD 7), the median body mass index was 27 kg/m^2 (IQR: 24-32), and 37.9% (41/108) of the women were primiparous. There were no significant differences in any of these demographic or obstetric characteristics among the 4 adherence groups (see Table 1).

Relationships Between Patient Characteristics, Questionnaire, and Adherence to Remote Monitoring

As expected, the results showed that several patient characteristics were associated with adherence, particularly in the groups with lower levels of adherence, although unexpected findings also emerged (see Multimedia Appendix 2). Specifically, participants in the moderate adherence group were characterized by the highest levels of anxiety and depression, particularly compared with the overadherence group, although these differences were quite modest. However, the moderate adherence group showed significantly elevated levels of rumination, magnification, and helplessness (cognitive factors) and elevated levels of both attachment anxiety and avoidance compared with the good and overadherence groups. There were no significant differences between the good and overadherence groups. Contrary to expectation, self-criticism and dependency were not associated with adherence. Other-oriented perfectionism was the only patient personality trait that distinguished the low adherence group from the other 3 groups, suggesting that this group of patients was characterized by high levels of criticism toward others, and particularly toward others who failed to meet their expectations.

Single linear regression showed that the PHQ-9 (*P*=.01) and ECR-R (*P*=.01) questionnaires were significantly related to the adherence rate. Multimedia Appendix 2 describes for each questionnaire and adherence group the median (IQR) or mean (SD) and *P* values.

Table 1. Characteristics of the study participants.

Variable	Low adherence, range: (0.0-27.8)	Moderate adherence, range: (36.1-78.3)	Good adherence, range: (81.7-100.0)	Overadherence, range: (100.6-156.1)	*P* value
Number of participants, n	12	32	31	33	—[a]
Adherence (%), median (IQR[b])	9.2 (0.0-18.6)	56.4 (47.2-71.4)	90.4 (88.3-97.5)	107.2 (103.9-116.1)	—
Age (years), median (IQR)	31 (29-34)	30 (28-36)	30 (28-32)	30 (28-33)	.77
Prepregnancy weight (kg), median (IQR)	76 (64-88)	78 (68-90)	72 (67-87)	82 (65-95)	.37
Height (cm), mean (SD)	170 (4)	166 (6)	168 (8)	166 (7)	.89
BMI[c] (kg/m^2), median (IQR)	28 (22-32)	28 (24-31)	26 (24-30)	29 (24-34)	.36
Primigravida, n (%)	7 (58)	8 (25)	7 (23)	19 (58)	.15

[a]No significance test was performed for this descriptive demographic variable.

[b]IQR: interquartile range.

[c]BMI: body mass index.

Discussion

Principal Findings

We investigated the relationships between patient characteristics questionnaires of pregnant women at risk of PIH and their adherence rates in an RM program. To our knowledge, this is the first study to investigate the potential role of patient characteristics questionnaires in adherence to an RM program for pregnant women at risk of PIH. A total of 3 interesting sets of findings emerged.

First, as expected, women exhibiting moderate adherence showed higher scores for negative psychosocial traits. Specifically, the moderately adherent group was notably characterized by high levels of both attachment avoidance and anxiety as well as tendencies to ruminate, feel helpless, and magnify problems. It seems these women may have shown suboptimal adherence to the program because they worried highly and ruminated upon potential negative outcomes. As a result, they may want to avoid any potential confrontation with threatening information. If this assumption is correct, this has important implications for the future implementation of RM because the early identification of these women may increase their adherence and thus, prevent negative outcomes.

Second, although women in the low adherence group (<30% adherence) seemed to have similar psychosocial characteristics to the women who showed good and overadherence, they were distinguished by markedly elevated levels of other-oriented perfectionism. This suggests that these women display poor adherence because they are very critical of others, perhaps including the health care staff proposing and initiating RM. As other-oriented per "difficult to reach," particularly with an intervention that involves very little personal contact between the patient and the health care provider. Therefore, when RM is implemented, it may be crucial to screen for these traits and to develop a preferably brief and cost-effective intervention to address these issues in such women.

Third, in marked contrast to our expectations, there were no significant differences between the good and overadherence groups. We expected traits such as high levels of self-criticism and rumination to be associated with very high levels of adherence, reflecting a maladaptive preoccupation with RM, leading to excessive health care behaviors. However, none of these traits distinguished this group of mothers from those with good adherence. Therefore, what we initially thought would reflect "high" adherence (ie, measuring BP more than the requested 2 times a day), might reflect the normal "maternal preoccupation" with the health of their baby that characterizes mothers in the peripartum period [28]. This maternal preoccupation is thought to reflect a biological and psychological preparedness to give birth, which from an evolutionary perspective, is adaptive. Therefore, these findings warn against interpreting the seemingly overadherence of mothers to RM as problematic. Of course, there may be a group of mothers for whom this normative preoccupation becomes a maladaptive preoccupation, but further research is required to investigate this.

Finally, the findings of this study showed a significant relation between the PHQ-9 and ECR-R questionnaire and the adherence rate.

Strengths and Limitations

Our study is the first to demonstrate relationships between patient psychosocial characteristics and adherence rates in an RM program. Our results should contribute to an increased use of RM in obstetric care, as encouraged by the European Communities in the eHealth Action Plan. An awareness of the influence of patient characteristics on adherence rates can be useful in selecting particular pregnant women for an RM program. Although the results of this study are encouraging, a number of limitations must be taken into account in future research. First, the generalizability of the results may be affected by the single-center design of the study. Second, the results of the study relied on self-reported data. To include diagnoses in questionnaires, for example GAD-7, clinical diagnostic interviews would be required. Furthermore, the difficulty in assessing depression prenatally is that several symptoms of depression, such as fatigue, appetite change, and sleep problems, are also associated with pregnancy. During clinical diagnostic interviews, a study-specific guide can be used to determine whether the study participants perceive their symptoms to be

pregnancy-related [29,30]. Third, the questionnaires were completed on a single moment basis. It is possible that exceptional events influenced the women's responses to the questionnaires.

The overall adherence rate in this study (mean 79.56%; median 89.44%) corresponds to reported rates of 70.40% and 90.00% adherence to BP measurements [31,32]. As reported by many studies, the adherence rate usually decreases steadily over time. This reduction was more evident in the first weeks or months after the start of RM [32,33]. Participants' nonadherence to the manual entry of daily information, especially in long-term monitoring programs, is also a problem [34].

Recommendations for Further Research

Multiple trajectories and predictors of health-related quality of life (HRQOL) have been determined in women during pregnancy. For instance, young maternal age, low education, financial dissatisfaction, unplanned pregnancy, pregnancy-related symptoms, depression, and domestic violence may be associated with low HRQOL [35]. Future studies should investigate the influence of these variables on adherence rates to RM. The results of the low adherence group may indicate that these mothers underreport distress because of a critical and hostile attitude toward RM. Future research is required to investigate this issue. The study of Biaggi et al [36] demonstrated that depression rates tend to increase with each trimester and that anxiety and pain interference also increases significantly over time during the third trimester [37,38]. Future research should confirm the results of this study in other longitudinal periods of pregnancy. Maternal anxiety during pregnancy is associated with several adverse outcomes, including spontaneous abortion, increased cesarean section, pre-eclampsia, placental abruption, preterm labor, low birth weight, smaller head circumference, and lower mental development scores in infants [39,40]. Future research should investigate the relationships between several adverse outcomes of pregnancy and adherence rates.

Conclusions

The peripartum period has long been known to be associated with increased levels of stress and anxiety, which can be exacerbated by PIH and negatively influence adherence rates. This study shows that anxiety, depression, and negative cognitive and attachment styles, but also other-oriented perfectionism, are characteristic of women with less than optimal adherence. As the results of the low adherence group threaten both the well-being and the follow-up of the patient, further research is required to determine possible strategies to improve the management of PIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adherence analysis based on equal sample size groups.
[PDF File (Adobe PDF File), 158KB - mhealth_v7i8e12574_app1.pdf]

Multimedia Appendix 2
Answers questionnaires related to adherence groups.
[PDF File (Adobe PDF File), 69KB - mhealth_v7i8e12574_app2.pdf]

References

1. Kintiraki E, Papakatsika S, Kotronis G, Goulis DG, Kotsis V. Pregnancy-induced hypertension. Hormones (Athens) 2015;14(2):211-223 [FREE Full text] [doi: 10.14310/horm.2002.1582] [Medline: 26158653]
2. Ayala DE, Hermida RC. Ambulatory blood pressure monitoring for the early identification of hypertension in pregnancy. Chronobiol Int 2013 Mar;30(1-2):233-259. [doi: 10.3109/07420528.2012.714687] [Medline: 23006127]
3. Kattah AG, Garovic VD. The management of hypertension in pregnancy. Adv Chronic Kidney Dis 2013 May;20(3):229-239 [FREE Full text] [doi: 10.1053/j.ackd.2013.01.014] [Medline: 23928387]
4. Odibo IN, Wendel PJ, Magann EF. Telemedicine in obstetrics. Clin Obstet Gynecol 2013 Sep;56(3):422-433. [doi: 10.1097/GRF.0b013e318290fef0] [Medline: 23903374]
5. Cruz JB, Brooks D, Marques A. Home telemonitoring in COPD: a systematic review of methodologies and patients' adherence. Int J Med Inform 2014 Apr;83(4):249-263. [doi: 10.1016/j.ijmedinf.2014.01.008] [Medline: 24529402]
6. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
7. Ganapathy RG, Grewal A, Castleman JS. Remote monitoring of blood pressure to reduce the risk of preeclampsia related complications with an innovative use of mobile technology. Pregnancy Hypertens 2016 Oct;6(4):263-265. [doi: 10.1016/j.preghy.2016.04.005] [Medline: 27939464]
8. Rhoads S, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring implementation of m-Health monitoring in postpartum women with hypertension. Telemed J E Health 2017 Dec;23(10):833-841 [FREE Full text] [doi: 10.1089/tmj.2016.0272] [Medline: 28475431]

9. Toogood PA, Abdel MP, Spear JA, Cook SM, Cook DJ, Taunton MJ. The monitoring of activity at home after total hip arthroplasty. Bone Joint J 2016 Nov;98-B(11):1450-1454. [doi: 10.1302/0301-620X.98B11.BJJ-2016-0194.R1] [Medline: 27803219]
10. Law LE, Edirisinghe N, Wason JM. Use of an embedded, micro-randomised trial to investigate non-compliance in telehealth interventions. Clin Trials 2016 Dec;13(4):417-424. [doi: 10.1177/1740774516637075] [Medline: 26968939]
11. Lyngå P, Fridlund B, Langius-Eklöf A, Bohm K. Perceptions of transmission of body weight and telemonitoring in patients with heart failure? Int J Qual Stud Health Well-being 2013 Dec 16;8:21524 [FREE Full text] [doi: 10.3402/qhw.v8i0.21524] [Medline: 24345687]
12. Souza W, Jardim PC, Brito LP, Araújo FA, Sousa AL. Self measurement of blood pressure for control of blood pressure levels and adherence to treatment. Arq Bras Cardiol 2012 Feb;98(2):167-174 [FREE Full text] [Medline: 22249412]
13. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
14. Erdine S, Arslan E. Monitoring treatment adherence in hypertension. Curr Hypertens Rep 2013 Aug;15(4):269-272. [doi: 10.1007/s11906-013-0369-9] [Medline: 23828148]
15. Slade AC, Cohen LJ, Sadler LS, Miller M. The psychology and psychopathology of pregnancy: reorganization and transformation. In: Zeanah CH, editor. Handbook of Infant Mental Health, Third Edition. New York: The Guilford Press; 2009:22-39.
16. Black KD. Stress, symptoms, self-monitoring confidence, well-being, and social support in the progression of preeclampsia/gestational hypertension. J Obstet Gynecol Neonatal Nurs 2007;36(5):10. [Medline: 17880312]
17. Leeners B, Stiller R, Neumaier-Wagner P, Kuse S, Schmitt A, Rath W. Psychosocial distress associated with treatment of hypertensive diseases in pregnancy. Psychosomatics 2008;49(5):413-419. [doi: 10.1176/appi.psy.49.5.413] [Medline: 18794510]
18. Karsnitz D, Ward S. Spectrum of anxiety disorders: diagnosis and pharmacologic treatment. J Midwifery Womens Health 2011;56(3):266-281. [doi: 10.1111/j.1542-2011.2011.00045.x] [Medline: 21535374]
19. Weisberg R, Paquette JA. Screening and treatment of anxiety disorders in pregnant and lactating women. Womens Health Issues 2002;12(1):32-36. [Medline: 11786290]
20. Sutter-Dallay A, Giaconne-Marcesche V, Glatigny-Dallay E, Verdoux H. Women with anxiety disorders during pregnancy are at increased risk of intense postnatal depressive symptoms: a prospective survey of the MATQUID cohort. Eur Psychiatry 2004 Dec;19(8):459-463. [doi: 10.1016/j.eurpsy.2004.09.025] [Medline: 15589703]
21. Mikulincer M, Shaver PR. Attachment, group-related processes, and psychotherapy. Int J Group Psychother 2007 Apr;57(2):233-245. [doi: 10.1521/ijgp.2007.57.2.233] [Medline: 17419673]
22. Bornstein R, Languirand FM. Healthy Dependency: Leaning on Others Without Losing Yourself. New York: William Morrow; 2003.
23. Ciechanowski P, Sullivan M, Jensen M, Romano J, Summers H. The relationship of attachment style to depression, catastrophizing and health care utilization in patients with chronic pain. Pain 2003 Aug;104(3):627-637. [Medline: 12927635]
24. Ciechanowski PS, Walker EA, Katon WJ, Russo JE. Attachment theory: a model for health care utilization and somatization. Psychosom Med 2002;64(4):660-667. [Medline: 12140356]
25. Hewitt P, Flett GL, Turnbull-Donovan W, Mikail S. The multidimensional perfectionism scale: reliability, validity, and psychometric properties in psychiatric samples. Psychol Assess 1991 Sep;3(3):464-468. [doi: 10.1037/1040-3590.3.3.464]
26. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
27. Lanssens DV, Vandenberk T, Smeets CJ, De Cannière H, Vonck S, Claessens J, et al. Prenatal remote monitoring of women with gestational hypertensive diseases: cost analysis. J Med Internet Res 2018 Dec 26;20(3):e102 [FREE Full text] [doi: 10.2196/jmir.9552] [Medline: 29581094]
28. Mayes L, Leckman JF. Parental representations and subclinical changes in postpartum mood. Infant Ment Health J 2007 May;28(3):281-295. [doi: 10.1002/imhj.20136] [Medline: 28640466]
29. Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Arch Womens Ment Health 2012 Oct;15(5):367-374. [doi: 10.1007/s00737-012-0295-x] [Medline: 22983357]
30. Yonkers K, Smith MV, Gotman N, Belanger K. Typical somatic symptoms of pregnancy and their impact on a diagnosis of major depressive disorder. Gen Hosp Psychiatry 2009;31(4):327-333 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.03.005] [Medline: 19555792]
31. Port K, Palm K, Viigimaa M. Daily usage and efficiency of remote home monitoring in hypertensive patients over a one-year period. J Telemed Telecare 2005;11(Suppl 1):34-36. [doi: 10.1258/1357633054461705] [Medline: 16035987]
32. Celler B, Argha A, Varnfield M, Jayasena R. Patient adherence to scheduled vital sign measurements during home telemonitoring: analysis of the intervention arm in a before and after trial. JMIR Med Inform 2018 Apr 09;6(2):e15 [FREE Full text] [doi: 10.2196/medinform.9200] [Medline: 29631991]

33. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient compliance in home-based self-care telehealth projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]
34. Jonassaint C, Shah N, Jonassaint J, De Castro L. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: the sickle cell disease mobile application to record symptoms via technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
35. Bai G, Raat H, Jaddoe VW, Mautner E, Korfage IJ. Trajectories and predictors of women's health-related quality of life during pregnancy: a large longitudinal cohort study. PLoS One 2018;13(4):e0194999 [FREE Full text] [doi: 10.1371/journal.pone.0194999] [Medline: 29614087]
36. Biaggi A, Conroy S, Pawlby S, Pariante CM. Identifying the women at risk of antenatal anxiety and depression: a systematic review. J Affect Disord 2016 Feb;191:62-77 [FREE Full text] [doi: 10.1016/j.jad.2015.11.014] [Medline: 26650969]
37. Ballard C, Davis R, Cullen PC, Mohan RN, Dean C. Prevalence of postnatal psychiatric morbidity in mothers and fathers. Br J Psychiatry 1994 Jun;164(6):782-788. [Medline: 7952984]
38. Simpson W, Glazer M, Michalski N, Steiner M, Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Can J Psychiatry 2014 Aug;59(8):434-440 [FREE Full text] [doi: 10.1177/070674371405900806] [Medline: 25161068]
39. Austin MP. Antenatal screening and early intervention for "perinatal" distress, depression and anxiety: where to from here? Arch Womens Ment Health 2004 Feb;7(1):1-6. [doi: 10.1007/s00737-003-0034-4] [Medline: 14963727]
40. Zhong Q, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sánchez SE, et al. Diagnostic validity of the Generalized Anxiety Disorder-7 (GAD-7) among pregnant women. PLoS One 2015;10(4):e0125096 [FREE Full text] [doi: 10.1371/journal.pone.0125096] [Medline: 25915929]

Abbreviations

BP: blood pressure
ECR-R: Experiences in Close Relationships-Revised Scale
GAD-7: Generalized Anxiety Disorder-7
HRQOL: health-related quality of life
IQR: interquartile range
PHQ-9: Patient Health Questionnaire-9
PIH: pregnancy-induced hypertension
PREMOM: Pregnancy Remote Monitoring
RM: remote monitoring

Edited by C Dias; submitted 22.10.18; peer-reviewed by S Vencatachellum, M Habibovic, T Dilles; comments to author 04.12.18; revised version received 19.01.19; accepted 09.04.19; published 28.08.19.

Please cite as:
Vandenberk T, Lanssens D, Storms V, Thijs IM, Bamelis L, Grieten L, Gyselaers W, Tang E, Luyten P
Relationship Between Adherence to Remote Monitoring and Patient Characteristics: Observational Study in Women With Pregnancy-Induced Hypertension
JMIR Mhealth Uhealth 2019;7(8):e12574
URL: http://mhealth.jmir.org/2019/8/e12574/
doi:10.2196/12574
PMID:31464190

Original Paper

How Resource Scarcity and Accessibility Affect Patients' Usage of Mobile Health in China: Resource Competition Perspective

Qing Ye[1,2], MA; Zhaohua Deng[2], PhD; Yanyan Chen[1], PhD; Jiazhi Liao[3], PhD; Gang Li[3], PhD; Yaobin Lu[4], PhD

[1]Department of Information Management, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3]Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[4]School of Management, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Zhaohua Deng, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
No 13 Hangkong Road, Qiaokou District
Wuhan, 430030
China
Phone: 86 19945036079
Email: zh-deng@hust.edu.cn

Abstract

Background: The last decade has witnessed many achievements in China's health care industry, but the industry still faces major challenges among which the uneven distribution of medical resources and the imbalance between supply and demand are the most pressing problems. Although mobile health (mHealth) services play a significant role in mitigating problems associated with health care delivery, their adoption rates have been low.

Objective: The objective of this study was to explore the impact of resource scarcity and resource accessibility on the adoption of mHealth from the perspective of resource competition, to examine the concerning factors, and to provide a theoretical basis for promoting mHealth in China.

Methods: We used 229,516 original registration records of outpatients to conduct an empirical analysis to examine the adoption of mHealth services from the perspective of resource competition.

Results: The adoption rate of mobile services for outpatients was low, accounting for only 31.5% (N=71,707). The empirical results indicated that resource scarcity (beta=.435, *P*=.01) and accessibility (beta=−.134, *P*=.02) have a significant impact on the adoption of mHealth. In addition, gender (beta=.073, *P*=.01) and age (beta=−.009, *P*<.001) are significantly related to adoption of mHealth. Experience with mHealth has a moderating role in the relationship between resource scarcity (beta=−.129, *P*=.02), accessibility (beta=.138, *P*=.04), and adoption of mHealth.

Conclusions: In this study we demonstrate that the external environment (resource scarcity and resource accessibility) has a significant impact on the adoption of mHealth. This study also demonstrates that experience with mHealth has a moderating role in the relationship between the elements of the external environment. Finally, we confirm that mHealth is a key factor in the delivery and allocation of medical resources and provide a theoretical basis for government agencies to develop policies on mHealth.

(JMIR Mhealth Uhealth 2019;7(8):e13491) doi:10.2196/13491

KEYWORDS

mobile health; technology adoption; moderating effect; resource scarcity; resource accessibility; resource competition

Introduction

Background

In recent years, China's health care industry has made a series of significant improvements, but it still faces many challenges, including uneven distribution of medical resources, imbalance between supply and demand, etc [1,2]. One particularly enormous challenge lies in ensuring the availability and equal distribution of high-quality resources. Such resources are mainly concentrated in large hospitals located in large or medium-sized cities, which are difficult for patients in rural and remote areas to access [2,3]. The number of physicians and nurses per 1000 persons in China is 2.06 and 2.13, respectively, which is much lower than in developed countries [1].

In addition, in recent years, the use and penetration of mobile phones have grown in China because of the rapid development of technology and the obvious reduction in cost [1,4]. China now ranks among the countries with the highest number of smartphones per capita [5]. With this proliferation of mobile technology, mobile health (mHealth) has become the main driving force in the health care industry. mHealth can be defined as the delivery of health care services through mobile technologies, which mainly comprise online consultations, appointment registration, and health recommendations [6]. Among these apps, appointment registration services, a fundamental and vital channel through which patients seek medical resources, have gained particularly widespread acceptance in China. With the rapid development of mobile technologies and the continuing penetration of mobile devices, mHealth has become a valuable tool and popular option to improve health care and service delivery in underserved regions [7-9]. Chinese government agencies have made policy decisions to support and promote the growth of this important emerging industry [10]. Health care companies and hospitals have begun to pay attention to this issue and have been attempting to provide health services directly to patients through mobile and wireless technologies. These services have the potential to be highly beneficial, especially for patients in rural and remote areas.

Although mHealth presents many opportunities for improving health care delivery, its adoption and use have been low [2,7,11,12]. A robust body of work in information systems has studied factors associated with mHealth adoption and use [11,13-15]. Although extant studies have explored and discussed reasons for these low adoption rates [14,16], empirical research examining this question from the perspective of resource competition is still rare. Previous studies have revealed that a growing number of patients who ordinarily seek health care services offline will move to online channels when medical resources are limited, which will inevitably affect patients' adoption and use of mHealth [11,15]. However, extant studies of mHealth adoption have mainly focused on patient-related factors, ignoring the impact of the external environment on technology adoption [17,18]. In the context of health care in China, the impact of external factors such as resource scarcity and accessibility caused by lack of and uneven distribution of medical resources on patients' technology adoption behavior is still unclear.

This paper aims to explore the impact of resource scarcity and resource accessibility on the adoption of mHealth from the perspective of resource competition, to examine the factors involved, and to provide a theoretical basis for promoting mHealth in China. In this research, we pose the following questions:

1. How do resource scarcity and accessibility affect the adoption of mHealth by patients in China?
2. How does patient experience with mHealth moderate the relationship between resource scarcity, resource accessibility, and adoption of mHealth?
3. Does mHealth play a role in the delivery of medical resources?

To answer these questions, we use a dataset of 227,539 outpatient registration records from both online and offline channels to conduct an empirical study. In the next section, we review the related literature. After we construct a conceptual model and develop 4 hypotheses, we present our research method and report our results. Finally, we discuss our research implications and note some limitations.

Literature Review and Theoretical Background

Mobile Health Services

With the rapid development of mobile internet technology and growing prevalence of smartphones, mHealth has become an increasingly practical, innovative approach to health care delivery, especially in rural and remote areas [19-21]. Many applications for mHealth services have been reported in the extant literature, including chronic disease management [22,23], mental health services [24], health solutions for pregnant women and teenage youth [25-27], and health monitoring [28,29].

mHealth overcomes geographical boundaries, enhances the equity and accessibility of medical resources, and provides an effective channel for patients in rural and remote areas to access medical services [30,31]. At present, hospitals provide a variety of services through mobile platforms available in China, which mainly include the WeChat platform and independently developed applications. Services provided comprise online consultation [32,33], appointment registration [7,30,34], electronic medical prescription, and online payment. Among these services, appointment registration is the service most heavily utilized, with which our study is most concerned. Previous studies have shown that experience has a positive significant impact on patients' adoption of mHealth [13,16,35]. In the context of mobile registration, historical registration records provide us with patients' experience-related data, affording us an empirical context to demonstrate the moderating effect between resource competition and patients' adoption of mHealth.

Medical Resource Scarcity and Accessibility

Health care issues, which include aging populations, chronic illnesses, rising costs, and access disparities, represent a major challenge in China [3,4,12]. As previously mentioned, the numbers of physicians and nurses per 1000 residents (2.06 and 2.13, respectively in 2014) are relatively inadequate compared with those of developed countries [1]. The shortage and uneven

distribution of medical resources is in fact the underlying reason for a number of current health care problems [3].

The finiteness of medical resources results in their scarcity, leading to a supply and demand problem in which the supply of medical resources cannot meet the growing demand of the people, who want access to these knowledge-intensive services when they are sick [36,37]. This results in competition among patients. Obviously, hospitals can only provide limited services to patients every day; thus, only patients who have registered can access medical services. With the development of internet technology and smart devices, patients can now access medical resources via both online (eg, online registration for medical appointments) and offline channels (eg, registered via full service or self-service in a hospital) [30]. Competition for medical resources will lead to behavioral changes. In this context, the online channel is an important supplement to the offline channel, and patients will compete for medical resources via these 2 channels. This change in behavior will affect patients' adoption of technology. On the basis of these arguments, the impact of medical resource competition on patients' adoption of technology is very worthy of examination.

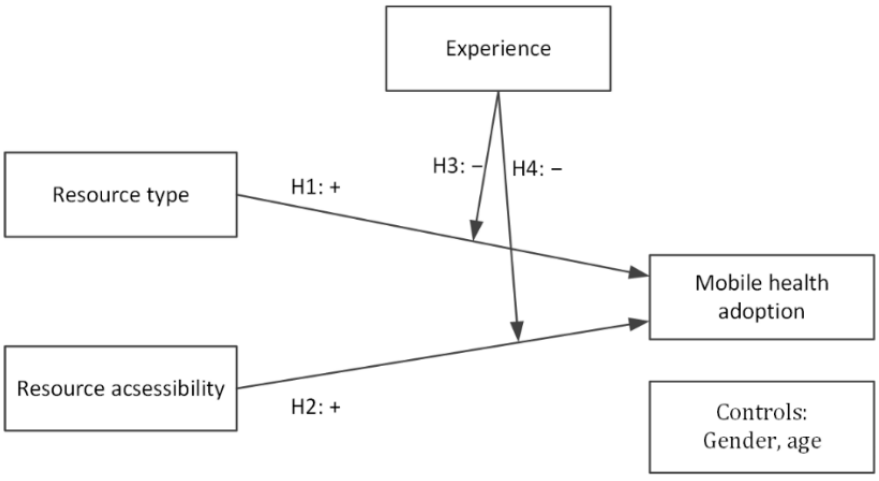
Extending Channel Complementarity Theory to mHealth Adoption

Technology adoption in health domains has attracted wide attention of scholars and has been studied from many theoretical perspectives such as technology acceptance model, social exchange theory, and identity theories [38-40]. On the basis of channel complementarity theory, this study focuses on the impact of external environment (resource scarcity and accessibility) on technology adoption. Dutta-Bergman constructed channel complementarity theory and suggested the idea of media complementarity [41,42]. Channel complementarity theory argues that individuals will use any available channel to meet their needs [43]. We conceptualize the process of appointment channels selection as an individual information acquisition process by drawing on the theory of channel complementarity. With the development of internet technologies and mHealth, a variety of services channels are emerging. Patients will choose any available channels to obtain needed medical resources, including mobile service channel (online) and traditional channels (offline). The complementary or displacement effects of the channels are manifested under the influence of the external environment. In our research context, the scarcity of medical resources will lead to the competition of patients for medical resources, thus affecting the choice of patients' service channels and ultimately affecting patients' technology adoption. In addition, the accessibility of medical resources for patients is varying, which will inevitably affect the choice of service channels and the adoption of mHealth.

Research Model and Hypotheses

Extant studies mainly focus on the impact of individual characteristics or technological factors on adoption of mHealth [44-46]. To our knowledge, few studies have considered the impact of the external environment, which may also exert a great influence on patients and their behavior [11,15]. Therefore, we believe that it is necessary to address this gap by studying the impact of the external environment on the adoption of mHealth. In this study, the external environment is characterized in reference to the scarcity and accessibility of medical resources. The scarcity of medical resources is measured according to the patient's choice of 1 of 2 different physician types, chief physician (higher value, more scarce) or associate chief physician (lower value, less scarce). Accessibility is measured by the distance from the patient's location to the hospital. In addition, we examine the moderating role of experience with mHealth in the relationship between medical resource scarcity and accessibility and mHealth adoption. The conceptual model of our research is illustrated in Figure 1. The research hypotheses in detail are presented as follows.

Figure 1. Conceptual model. H: hypothesis.



Medical Resource Scarcity

We define the scarcity of medical resources as limited medical resources relative to people's diverse health needs. In this study, we use the value and abundance of medical resources to measure resource scarcity. When patients compete for medical resources, they are mainly competing for medical resources of different values. Medical resources of high value are relatively scarce, presenting the problem of unbalanced supply and demand. Patients' competition for high-value medical resources is more

intense, compared with the competition for lower-value medical resources. Mobile registration service provides patients with a very important online channel through which all patients with a smart device can access medical resources. As overall competition intensifies, traditional offline competition will certainly shift to online channels so that patients will compete for medical resources via mHealth platforms. Therefore, we can use medical resource scarcity as a measurement dimension of resource competition to examine the impact of medical resource competition on the adoption of mHealth.

Mobile registration services in China mainly provide 2 types of medical resources, patients may choose to make an appointment with a chief physician or with an associate chief physician. The difference between these 2 practitioner types lies in their level of expertise, so chief physicians are of higher value for patients and competition for appointments with them is more intense. According to channel complementarity theory, patients will meet their medical needs through all channels that can obtain the required medical resources, and mHealth just provides such opportunities and channels. Patients will compete for these 2 types of medical resources through both offline and online channels, including mobile registration. Therefore, the scarcity of resources will affect the channel choice of patients. On the basis of these arguments, we hypothesize the following:

H1: Medical resource scarcity has a significant positive impact on the adoption of mHealth. The adoption rate of mHealth is high when medical resources have a higher value.

Medical Resource Accessibility

Accessibility of medical resources refers to the degree of difficulty for patients to access medical resources, and distance is an important indicator to measure the accessibility of medical resources. Medical resources are unevenly distributed because of the regional imbalances of economic development in China, which result in lower resource accessibility for many patients. High-quality medical resources are mainly concentrated in large hospitals in big cities, making it difficult for patients in rural and remote areas to access them. As the distance between the patient and the hospital increases, the accessibility of medical resources, especially those of higher quality, will decrease concurrently. In addition to the cost of medical treatment, more remote patients face additional costs such as transportation and accommodations. We hypothesize that to reduce these associated costs, patients who are far away from the hospital will generally register via mHealth service and confirm that the registration is successful before they go to the hospital.

According to the channel complementarity theory, patients with lower accessibility of medical resources will show channel displacement effect, that is, they will reduce medical costs through the mobile channel. Therefore, we can safely assume that the accessibility of medical resources may affect patients' adoption of mHealth. Distance to the hospital can be used as a measure of medical resource accessibility and can be used to examine the impact of accessibility of medical resources on the adoption of mHealth. In this context, we believe that the accessibility of medical resources is low for patients who are far away from the hospital. To compensate for this disadvantage, patients must use mHealth services as a tool to obtain medical resources. Thus, we formulate the following hypothesis:

H2: The accessibility of medical resources associated with mHealth is negatively related to patient adoption of mHealth. Patients who must travel longer distances to a hospital have a higher mHealth adoption rate.

The Moderating Effect of Experience

Existing studies have shown that experience with mHealth has a positive impact on its adoption [7,13,47,48]. In this study, we agree with this argument and will further confirm it in the following empirical section. In doing so, we draw on the attribution theory, which argues that people tend to attribute their actions to the conditions of the external environment and as a result, they may limit their actions [49,50]. In the context of our research, the scarcity and accessibility of medical resources are the external environmental factors that may limit patients' adoption of mobile services, and the experience of success or failure with mobile registration will affect patients' attribution. Attribution of previous successes or failures in a specific action (in our case, mobile registration) will affect patients' expectations, emotions, and efforts the next time they have the opportunity to use mobile registration. Patients' experiences with mobile registration are often unsuccessful or may even fail because of the scarcity of and competition for medical resources. This unsuccessful or failed experience has a certain impact on the relationship between resource scarcity, accessibility, and adoption of mHealth. Therefore, it is necessary to study the moderating role of patient experience in the relationship between resource scarcity, accessibility, and the adoption of mHealth. In our research context, we argue that this moderating effect is negative. On the basis of the arguments above, we propose the following hypotheses:

H3: Experience with mHealth has a negative moderating role in the relationship between medical resource scarcity and patients' adoption of mHealth.

H4: Experience with mHealth has a negative moderating role in the relationship between accessibility of medical resources and patients' adoption of mHealth.

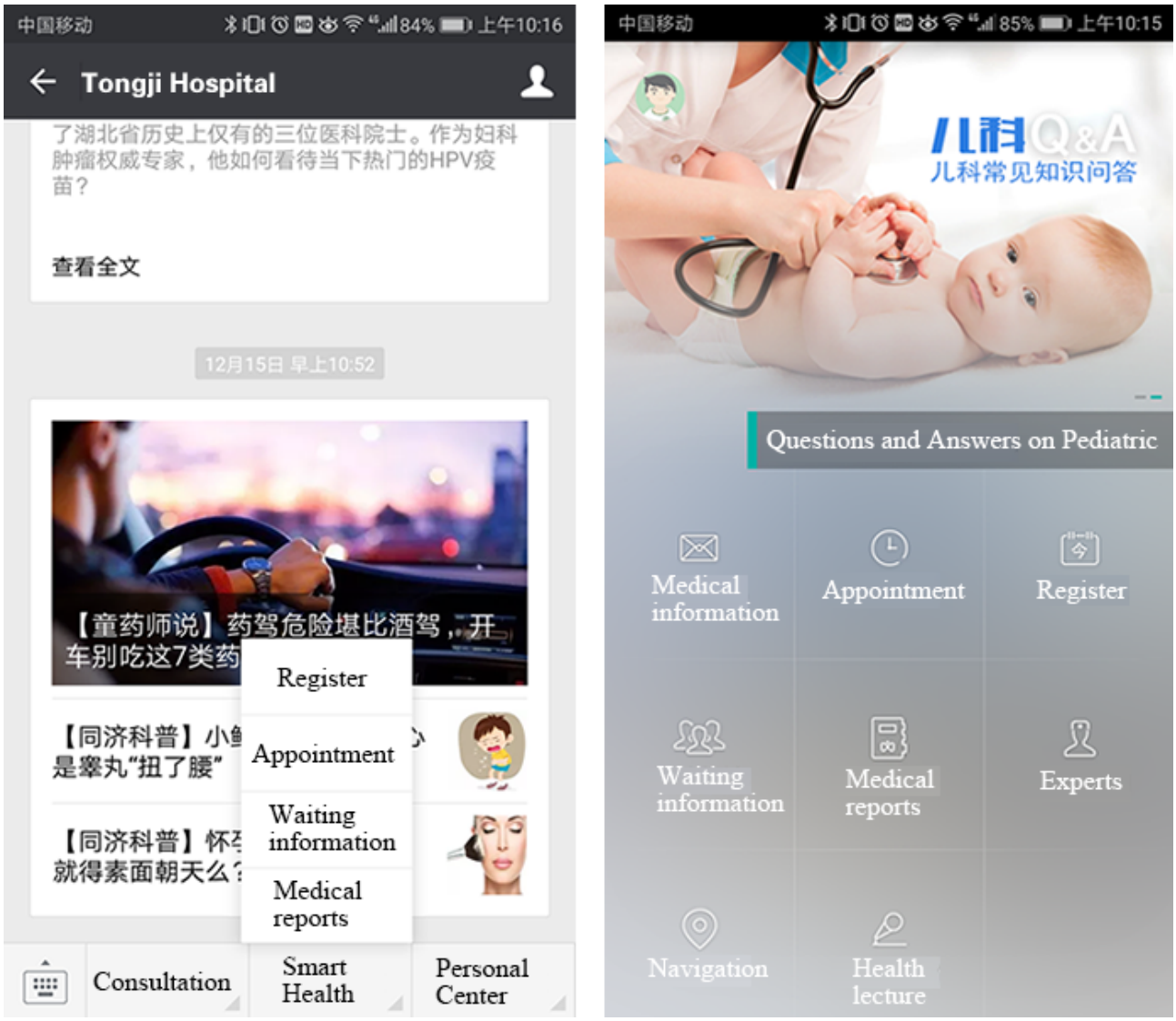
Methods

Research Context

In this study, our research environment is Tongji Hospital, which is a large, tertiary, multispecialty referral hospital in China. Founded in 1900 by Dr Erich Paulun, a German physician, Tongji Hospital is an innovative modern hospital integrating medical care, teaching, and research [51]. In addition, Wuhan Tongji Hospital is one of China's top 10 most well-known hospitals, according to Fudan rankings. Tongji Hospital has offered a mobile online appointment registration service since 2014, utilizing platforms such as WeChat, the E-Tongji app, and some third-party registration platforms such as Guahao website and the China Mobile 12580 platform (see Figure 2 for snapshots of the WeChat platform and E-Tongji app). Tongji Hospital offers traditional offline registration channels as well. The hospital provides 2 types of medical resources, chief physician and associate chief physician, which, as mentioned

above, are accorded different values, as patients in China tend to value chief physicians over associate chief physicians. Both resource types are in short supply, providing a suitable environment for us to study the impact of resource scarcity on the adoption of mHealth. Patients in Tongji Hospital come from many provinces in China, providing us with a wide range of distance to hospital measurements, which we use to gauge the accessibility of medical resources for each patient.

Figure 2. Tongji hospital registration snapshots from the WeChat platform and E-Tongji app.



Sample and Data Collection

In our research, we used 4 specific datasets. The first comprised 229,516 original registration records of outpatients from both online and offline channels. The registration records were collected from the appointment and registration platform of the Optical Valley branch of Wuhan Tongji Hospital between June 2016 and May 2017. All registration records were anonymized in accordance with patient privacy regulations. Each registration record contains patient's age, patient's gender, type of medical resource (chief physician or associate chief physician) selected, a variable for prior experience with mobile registration, and the first 7 digits of the patient's mobile phone number. The second dataset contains the travel distance between the city in which the patient is located and the Optical Valley branch of Wuhan Tongji Hospital, as calculated by Baidu Map. The third dataset was used to obtain patients' location by linking patients' mobile phone numbers to publicly available city information. A patient's city of residence was deduced based on the first 7 digits of their mobile phone number (similar to what one can do using area codes in the United States). Any incomplete or erroneous records were removed to ensure the reliability of the results. Of the 229,516 original registration records, a total of 227,539 (99.14%) records were analyzed after screening.

Variables and Models

Our research variables are presented in Table 1. Gender and age of patient are control variables. The independent variables are experience, resource scarcity, and resource accessibility. Experience with mHealth is identified according to whether the patient has previously used a mobile registration service. Resource scarcity is measured by medical resource type (chief physician or associate chief physician). In a microenvironment (a specific hospital), there are generally fewer physicians with higher title than those with lower title, which indicates that they are more scarce. Resource accessibility is measured by distance

from patient location to the hospital. Adoption of mHealth, which we define as patient registration via any of the several mHealth service channels, serves as the dependent variable. The available mHealth service channels include the WeChat platform, E-Tongji app, and some third-party registration platforms.

Table 1. Definition of variables.

Variable	Definition and measurement	Symbols
Dependent variable		
Adoption	A binary variable was used to measure whether the patient made an appointment via mobile health service; 0=never, 1=yes	—[a]
Independent variables		
Resource scarcity (RS)	RS is measured by the type of medical resource; 0=associate chief physician, 1=chief physician	RS
Resource accessibility (RA)	RA is measured by distance to the hospital; 1=less than 300 km, 0=more than 300 km. 300 km is the distance from the farthest city in the province to the city where the hospital is located	RA
Experience (EXP)	Whether patients have experience with mobile registration service; 0=no, 1=yes	EXP
Gender	Value 0=a male patient and value 1=a female	—
Age	Patient age in years	—

[a]Not applicable.

We conducted logistic regression using SPSS version 23.0 to test our research hypotheses. For this study, our analysis comprises 3 steps. In the first step, we estimate our model using only control variables. Second, we add the main effect, including experience, resource scarcity, and resource accessibility. Finally, we incorporate the interaction effect into our model. The full empirical model can be expressed as follows:

$$\text{logit}(Y_i) = \beta_0 + \beta_1 \text{Gender}_i + \beta_2 \text{Age}_i + \beta_3 \text{Experience}_i + \beta_4 \text{ResourceScarcity}_i + \beta_5 \text{ResourceAccessibility}_i + \beta_6 \text{ResourceScarcity}_i \times \text{Experience}_i + \beta_7 \text{ResourceAccessibility}_i \times \text{Experience}_i + \varepsilon_i$$

where β_0 is the constant term. β_1 and β_2 are the coefficients associated with control variables. β_3, β_4, and β_5 are the coefficients associated with examined variables. β_6 and β_7 are the coefficients of interaction terms. Gender_i, Age_i, and Experience_i represent patient characteristics. $\text{ResourceScarcity}_i$ and $\text{ResourceAccessibility}_i$ account for the external environment of medical resource competition. ε_i is the error term.

Results

Descriptive Study

The descriptive statistics and correlations for variables in our study are shown in Table 2. The adoption rate of mobile services for outpatients was low, accounting for only 31.5% (N=71,707) and the mean age of outpatients was 37.109 years (SD 18.987). The results indicate that the dependent variables are correlated with the independent variables, and the correlation between study variables and control variables is low. These results show that multiple collinearity is not a significant problem in our research, which guarantees the accuracy of the model estimation.

Table 2. Descriptive statistics and correlation matrix (N=227,539).

Variable number	Descriptive statistics construct	Mean (SD)	Variable number				
			1	2	3	4	5
1	Age of patient	37.109 (18.987)	1.000[a]	—[b]	—	—	—
2	Gender of patient	0.549 (0.498)	0.058[c]	1.000	—	—	—
3	Experience	0.193 (0.395)	−0.027[c]	0.031[c]	1.000	—	—
4	Resource scarcity	0.510 (0.499)	0.009[c]	−0.009[c]	0.033[c]	1.000	—
5	Resource accessibility	0.107 (0.309)	0.009[c]	0.020[c]	0.012[c]	0.006[d]	1.000
6	Adoption	0.315 (0.465)	−0.077[c]	0.019[c]	0.093[c]	0.010[c]	0.294[c]

[a]Correlations between 2 variables are calculated using Pearson correlation analysis.

[b]Not applicable.

[c]$P<.001$.

[d]$P<.005$.

Empirical Results

The regression results are presented in Table 3. Model 1 includes only control variables— *gender* and *age* —which are both significantly related to the adoption of mHealth. Specifically, patient age has a significant negative impact on the adoption of mHealth and patient gender has a significant positive relationship to the adoption of mHealth. Model 2 includes the main effects, in addition to the control variables. We find that all the main effects' coefficients are significant. Experience with mHealth and medical resource scarcity are significantly positively related to the adoption of mHealth, whereas medical resource accessibility is significantly negatively related. Model 3 includes the interaction terms, in addition to the main effects and control variables. In model 3, all the main effects' coefficients are significant, which is consistent with model 2. The interaction effects have significant negative coefficients.

Table 3. Regression results.

Variable	Model 1	Model 2	Model 3
Constant	−0.504 (0.011)[a]	−1.045 (0.013)	−1.067 (0.013)
Age of patient	−0.009[b] (0.000)	−0.009[b] (0.000)	−0.009[b] (0.000)
Gender of patient	0.102[b] (0.009)	0.073[b] (0.010)	0.073[b] (0.010)
Experience with mobile health (EXP)	—[c]	1.484[d] (0.011)	1.568[d] (0.017)
Resource scarcity (RS)	—	0.404[b] (0.010)	0.435[d] (0.011)
Resource accessibility (RA)	—	−0.104[d] (0.015)	−0.134[d] (0.017)
RS×EXP	—	—	−0.129[d] (0.022)
RA×EXP	—	—	0.138[d] (0.037)
Log-likelihood	282,093.080	262,017.222	261,970.485
Cox and Snell R-square	0.007	0.090	0.091
Nagelkerke R-square	0.009	0.127	0.127

[a]The values in parentheses are standard deviation.

[b]$P<.01$.

[c]Not applicable.

[d]$P<.05$.

In H1, we posited that medical resource scarcity has a significant positive impact on the adoption of mHealth, which occurs at a high rate when medical resources have a higher value. As shown in Table 3, the coefficients of medical resource scarcity are positive and significant in model 2 ($beta_4$=.404, $P<.05$) and model 3 ($beta_4$=.435, $P<.05$). Therefore, H1 is supported. In H2, we hypothesized that the accessibility of a medical resource associated with mHealth is negatively related to the adoption of mHealth, so that patients who must travel a greater distance to the hospital would have a higher adoption rate of mHealth. As shown in Table 3, the coefficients of medical resource accessibility are negative and significant in model 2 ($beta_5$=−.104, $P<.05$) and model 3 ($beta_5$=−.134, $P<.05$). Thus, hypothesis H2 is supported. In H3 and H4, we posited that experience with mHealth has a negative moderating role in the relationship between medical resource scarcity and patients' adoption of mHealth, whereas medical resource accessibility has a positive moderating effect. The interaction effects' coefficients are significant in model 3 ($beta_6$=−.129, $P<.05$; $beta_7$=.138, $P<.05$), thus supporting hypotheses H3 and H4. In the next section, we will test the robustness of our models.

Robustness Check

To obtain a more fine-grained understanding of our sample, we conducted visualization analysis of the geographical distribution and registration channels of outpatients. In our sample, 206,327 of 227,539 outpatients (90.68%) who registered through both online and offline channels were from Hubei Province. To demonstrate the geographical distribution characteristics of the patients who used online channels, we identified the number of patients by distinct colors at the national and provincial level. The visualization results are shown in Figure 3. At the national level (see Figure 3) and among those patients who registered online, most patients (N=65,760, 91.71%) came from Hubei Province, followed by Henan and other neighboring provinces. Within Hubei Province (see Figure 3) and among those patients who registered online, a total of 53,047 (80.67%) patients came from Wuhan, with Ezhou (N=2328, 3.5%) ranking second and Huanggang (N=2181, 3.3%) ranking third. This visualization shows that at both the national and provincial levels, the number of patients who used online channels gradually decreased as the distance to the hospital increased.

We also examined the online channels used for registration, with Figure 4 displaying the trends in proportional use of each channel. This graph shows that the WeChat platform displays an upward trend in proportional use each month between June 2016 and May 2017, whereas the use of the E-Tongji app decreases. Furthermore, the proportional use of the WeChat platform is higher than the E-Tongji app in all months. WeChat and the E-Tongji app constitute the official registration channels

provided by the hospital directly, but in addition, unofficial channels, such as Guahao website and the China Mobile 12580 platform, constitute an intermediary platform. During our study period, the proportional use of official channels increased from 75.02% to 85.89%, whereas the proportional use of unofficial channels decreased from 24.97% to 14.11%.

On the basis of Figures 3 and 4, we find that most of the hospital's patients are concentrated in Hubei and that channel selection is relatively stable. The likely reason is that each province has its own medical resources. Therefore, we narrow our sample so as to only consider patients from Hubei and then test the stability of the models. In the robustness check, we incorporate medical resource accessibility as a continuous variable into the models, whereas other variable definitions remain unchanged. The results, presented in Tables 4 and 5, indicate that our models are stable when we control for the impact of geographical factors.

Figure 3. Geographical distribution of patients who register online.

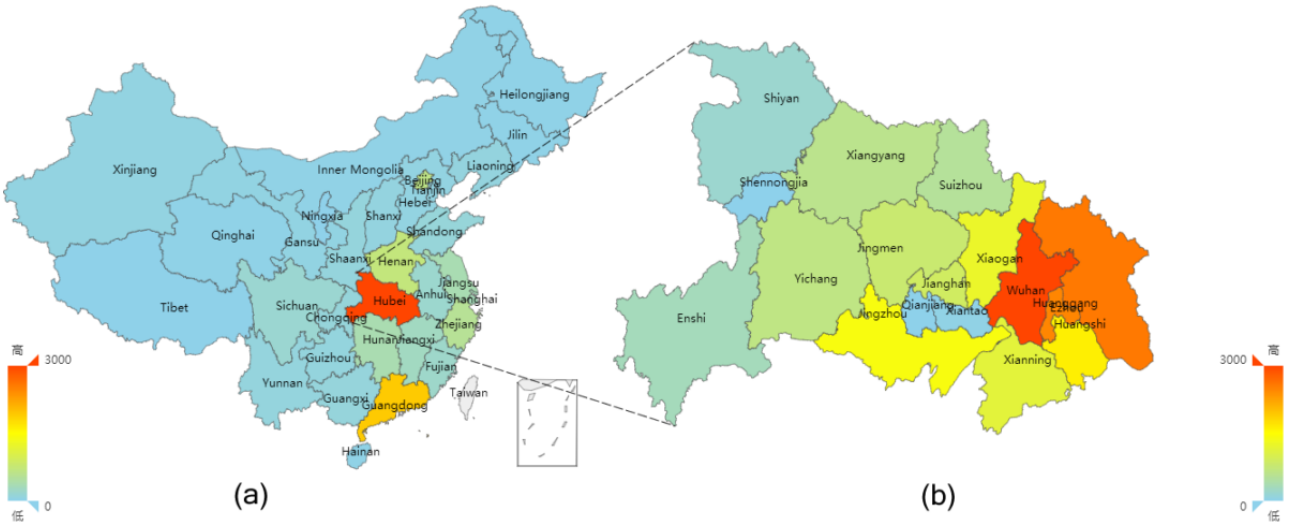


Figure 4. Trend in the proportional use of registration channels.

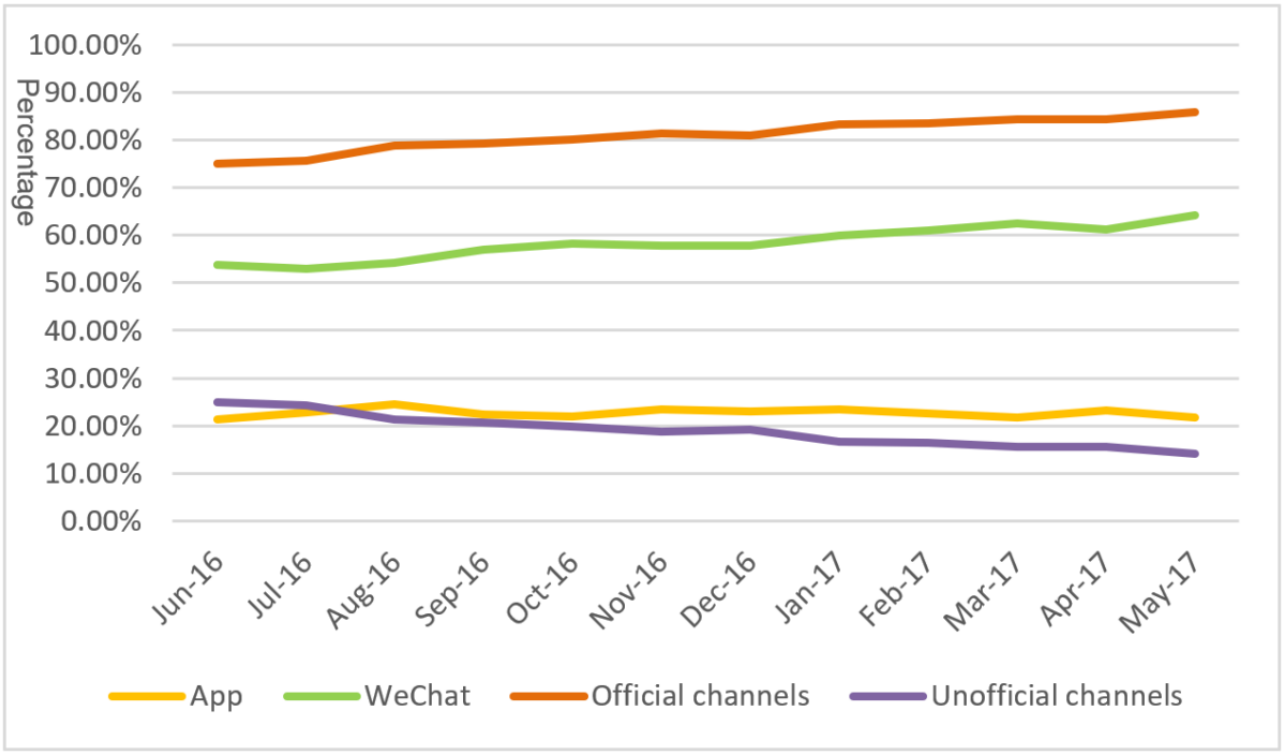


Table 4. Descriptive statistics and correlation matrix (N=206,327).

Variable number	Descriptive statistics construct	Mean (SD)	Variable number				
			1	2	3	4	5
1	Age of patient	37.170 (19.148)	1.000[a]	—[b]	—	—	—
2	Gender of patient	0.553 (0.497)	0.055[c]	1.000	—	—	—
3	Experience	0.194 (0.396)	−0.029[c]	0.028[c]	1.000	—	—
4	Resource scarcity	0.511 (0.499)	0.009[c]	−0.011[c]	0.034[c]	1.000	—
5	Resource accessibility	70.149 (60.985)	0.061[c]	0.001[e]	−0.022[c]	0.015[d]	1.000
6	Adoption	0.314 (0.464)	−0.081[c]	0.017[c]	0.297[c]	0.094[c]	0.018[c]

[a]Correlation coefficients between 2 variables are Pearson correlation coefficients.

[b]Not applicable.

[c]$P<.001$.

[d]$P<.005$.

[e]Not significant.

Table 5. Regression results.

Variable	Model 1	Model 2	Model 3
Constant	−0.493 (0.011)[a]	−1.095 (0.014)	−1.124 (0.015)
Age of patient (Age)	−0.009[b] (0.000)	−0.009[b] (0.000)	−0.009[b] (0.000)
Gender of patient (Gender)	0.093[b] (0.010)	0.066[b] (0.010)	0.066[b] (0.010)
Experience with mobile health (EXP)	—[c]	1.498[d] (0.012)	1.621[d] (0.022)
Medical resource scarcity (MRS)	—	0.403[b] (0.010)	0.430[d] (0.012)
Medical resource accessibility (MRA)	—	−0.001[b] (0.000)	−0.001[b] (0.000)
MRS×EXP	—	—	−0.112[d] (0.024)
MRA×EXP	—	—	0.001[b] (0.000)
Log-likelihood	255,344.365	236,735.682	236,692.326
Cox and Snell R-square	0.007	0.093	0.093
Nagelkerke R-square	0.010	0.130	0.130

[a]The values in parentheses are standard deviation.

[b]$P<.01$.

[c]Not applicable.

[d]$P<.05$.

Discussion

Principal Findings

The goal of this research was to explore the impact of external environment (resource scarcity and resource accessibility) on the mHealth adoption and to identify the moderating role of experience. Drawing from channel complementarity theory and attribution theory, we proposed our research hypotheses and conducted empirical research.

The empirical results of our study support all our hypotheses and confirm some previous arguments. In line with previous work, our results indicate that age is negatively and significantly related to the adoption of mHealth services, with young people showing higher acceptance than the elderly [2,13]. Extant studies have also demonstrated that gender is strongly associated with mHealth adoption [7,52]. In our research, we also identify a statistically significant difference in mHealth adoption between women and men. Overall, women exhibit more positive attitudes toward mHealth services, which has been rarely reported before. One possible explanation is that as the popularity of mobile devices has increased, the technological knowledge gaps between men and women have narrowed, and women are more likely to handle the task of making doctor's appointments for other family members. Existing studies have shown that experience with mHealth is a crucial factor determining the use of mHealth [48], which is also confirmed by our study. Our

results reveal that experience has a significant and positive impact on the adoption of mHealth.

Based on the main effect of our model, empirical results demonstrate that the external environment, characterized here as an environment of medical resource scarcity, has a positive and significant impact on the adoption of mHealth, whereas medical resource accessibility is negative and significant. In our study, medical resource scarcity is assessed based on which medical resource type (chief physician or associate chief physician) a patient selected. These different physician types are of different values to patients, and the fact that patients in China place a higher value on chief physicians and that their numbers are relatively limited both prompt patients to move from offline to online registration to compete for this limited medical resource. Therefore, we observe that resource scarcity is positively and significantly related to the adoption of mHealth. In terms of accessibility, which we measure using the distance from the patient's location to the subject hospital, we find that resource accessibility has a negative and significant impact on the adoption of mHealth. Patients have a higher mHealth adoption rate when medical resource accessibility is low. This result shows that mHealth services play a significant role in the delivery of medical resources and improve the fairness and accessibility of health care in rural and remote areas. Finally, considering the interaction effect in our model, the results indicate that experience with mHealth has a negative moderating role in the relationship between medical resource scarcity and patients' adoption of mHealth, whereas medical resource accessibility has a positive moderating effect.

Contributions

Our results contribute to the existing literature in several ways. First, we addressed the research gap previously discussed, noting that existing studies about technology adoption have focused mainly on patient-related factors rather than external environments [15,16]. On the basis of channel complementarity theory and attribution theory, we conducted an empirical study to analyze the adoption of mHealth from the perspective of resource competition and found that medical resource scarcity and accessibility have significant impacts on the adoption of mHealth. Second, we found that patients have a higher mHealth adoption rate when they must travel greater distances to the hospital. We therefore argue that mHealth has a significant role in health care delivery and medical resource allocation [1]. Finally, we confirmed that patients' prior experience with mHealth is positively and significantly related to the adoption of mHealth, and we found that experience has a moderating effect in the relationship between resource scarcity, accessibility, and adoption of mHealth.

We also made several useful contributions to the practice and policy surrounding mHealth. As it establishes the important role of mHealth in the delivery and allocation of medical resources, our work provides a theoretical basis for government agencies to develop policies on mHealth. For health care providers and stakeholders, we propose several feasible measures to increase the mHealth adoption rate, which will maintain their advantage and competitiveness in the health care market. Our results also suggest that health care providers and policymakers can take measures to encourage patients to use mHealth services and therefore increase the overall adoption rate, such as cultivating habitual patient use of mHealth services, reducing online registration fees, and taking mobile channel priority.

Limitations

There are some limitations to our study. First, as the data were provided by one specific hospital, our results may be influenced by this hospital's particular characteristics, and the universal adaptability of our results may be limited. Second, we only used age and gender as our control variables, so we may be lacking some patient-related variables, such as education, income level, and technological competence, which could affect our results. Third, in this study, we used 2 different physician types to measure medical resource scarcity, both of which are expert resources; however, outpatient medical resources in China also include attending physicians, which would be considered more of a general medical resource. In this study, we did not examine the differences between expert and general resources, which could potentially be significant. We plan to incorporate this distinction and further enrich our results in future research.

Conclusions

In summary, our study confirms that gender, age, and experience are significantly related to the adoption of mHealth, findings which are in line with existing studies. Our empirical results reveal that medical resource scarcity and accessibility have positive and negative impacts, respectively, on the adoption of mHealth. Experience with mHealth plays a moderating role in the relationship between resource scarcity, accessibility, and adoption of mHealth. We hope that further research will enrich our understanding of mHealth adoption and the impact of external environmental factors.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (award number 71671073 and 71810107003) and the Natural Science Foundation of Hubei Province (award number 2018CFB739). The authors thank Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology for providing data for the analysis. The authors appreciate the editors and 4 anonymous reviewers for their constructive comments during the review process. The authors would like to sincerely appreciate Prof Guodong (Gordon) Gao for his assistance on earlier drafts of this paper.

Conflicts of Interest

None declared.

References

1. Zhang X, Lai K, Guo X. Promoting China's mHealth market: a policy perspective. Health Policy Technol 2017 Dec;6(4):383-388. [doi: 10.1016/j.hlpt.2017.11.002]
2. Sun J, Guo Y, Wang X, Zeng Q. mHealth for aging China: opportunities and challenges. Aging Dis 2016 Jan;7(1):53-67 [FREE Full text] [doi: 10.14336/AD.2015.1011] [Medline: 26816664]
3. Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. Lancet 2014 Aug;384(9945):805-818. [doi: 10.1016/S0140-6736(14)61120-X]
4. Stefanacci RG, Riddle A. China - A half billion older adults getting and giving assistance. Geriatr Nurs 2017 Nov;38(6):591-595. [doi: 10.1016/j.gerinurse.2017.10.005] [Medline: 29103611]
5. Newzoo. Top Countries/Markets by Smartphone Penetration & Users URL: https://newzoo.com/insights/rankings/top-50-countries-by-smartphone-penetration-and-users/ [accessed 2019-05-31] [WebCite Cache ID 78mAX0tYB]
6. Zhang X, Guo X, Wu Y, Lai K, Vogel D. Exploring the inhibitors of online health service use intention: a status quo bias perspective. Inf Manage 2017 Dec;54(8):987-997. [doi: 10.1016/j.im.2017.02.001]
7. Hoque MR. An empirical study of mHealth adoption in a developing country: the moderating effect of gender concern. BMC Med Inform Decis Mak 2016 Dec 3;16:51 [FREE Full text] [doi: 10.1186/s12911-016-0289-0] [Medline: 27142844]
8. Hamilton EC, Saiyed F, Miller CC, Eguia A, Fonseca AC, Baum GP, et al. The digital divide in adoption and use of mobile health technology among caregivers of pediatric surgery patients. J Pediatr Surg 2018 Aug;53(8):1478-1493. [doi: 10.1016/j.jpedsurg.2017.08.023] [Medline: 28927983]
9. WHO Global Observatory for eHealth. Mhealth: New Horizons For Health Through Mobile Technologies (Global Observatory For Ehealth). Geneva: World Health Organization; 2019.
10. Wu H, Deng Z. Knowledge collaboration among physicians in online health communities: a transactive memory perspective. Int J Inf Manage 2019 Dec;49:13-33. [doi: 10.1016/j.ijinfomgt.2019.01.003] [Medline: 10]
11. 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]
12. Henriquez-Camacho C, Losa J, Miranda JJ, Cheyne NE. Addressing healthy aging populations in developing countries: unlocking the opportunity of eHealth and mHealth. Emerg Themes Epidemiol 2014 Dec;11(1):136 [FREE Full text] [doi: 10.1186/s12982-014-0021-4] [Medline: 25642276]
13. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 Dec;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
14. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 2;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
15. Maiga G, Namagembe F. Predicting adoption of mHealth technology in resource constrained environments. 2014 May 7 Presented at: IST-Africa Conference Proceedings; May 6-9, 2014; Le Meridien Ile Maurice, Mauritius p. 1. [doi: 10.1109/ISTAFRICA.2014.6880628]
16. Khatun F, Heywood AE, Ray PK, Hanifi SM, Bhuiya A, Liaw S. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
17. Kaphle S, Chaturvedi S, Chaudhuri I, Krishnan R, Lesh N. Adoption and usage of mHealth technology on quality and experience of care provided by frontline workers: observations from rural India. JMIR Mhealth Uhealth 2015 May 28;3(2):e61 [FREE Full text] [doi: 10.2196/mhealth.4047] [Medline: 26023001]
18. Huang H, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
19. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza J, et al. Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed Eng Online 2015;14(Suppl 2):S6 [FREE Full text] [doi: 10.1186/1475-925X-14-S2-S6] [Medline: 26329639]
20. Silva B, Rodrigues J, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
21. Singh K, Drouin K, Newmark L, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]
22. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 6;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
23. Yao Q, Liu F, Hou X. Research on blood pressure health management system on mobile terminal. J Adv Oxid Technol 2018;21(2) http://www.jaots.net/journal/volume21/number2.html.
24. Thornton L, Kay-Lambkin F, Tebbutt B, Hanstock T, Baker A. A mobile phone–based healthy lifestyle monitoring tool for people with mental health problems (MyHealthPA): development and pilot testing. JMIR Cardio 2018 Oct 4;2(2):e10228. [doi: 10.2196/10228]

25. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: Evaluation of a mobile health app. Brain Inj 2017;31(10):1279-1286 [FREE Full text] [doi: 10.1080/02699052.2017.1332388] [Medline: 28665690]
26. Rivera-Romero O, Olmo A, Muñoz R, Stiefel P, Miranda M, Beltrán LM. Mobile health solutions for hypertensive disorders in pregnancy: scoping literature review. JMIR Mhealth Uhealth 2018 May 30;6(5):e130 [FREE Full text] [doi: 10.2196/mhealth.9671] [Medline: 29848473]
27. Arbour MW, Stec MA. Mobile applications for women's health and midwifery care: a pocket reference for the 21st century. J Midwifery Womens Health 2018 Dec 15;63(3):330-334. [doi: 10.1111/jmwh.12755] [Medline: 29763990]
28. Villanueva J, Suarez M, Garmendia O, Lugo V, Ruiz C, Montserrat J. The role of telemedicine and mobile health in the monitoring of sleep-breathing disorders: improving patient outcomes. Smart Homecare Technol Telehealth 2017 Feb;Volume 4:1-11. [doi: 10.2147/SHTT.S108048]
29. Quercia K, Tran PL, Jinoro J, Herniainasolo JL, Viviano M, Vassilakos P, et al. A mobile health data collection system for remote areas to monitor women participating in a cervical cancer screening campaign. Telemed J E Health 2018 Dec;24(4):277-282. [doi: 10.1089/tmj.2017.0146] [Medline: 28846504]
30. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Dec 4;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
31. Ahmed T, Lucas H, Khan AS, Islam R, Bhuiya A, Iqbal M. eHealth and mHealth initiatives in Bangladesh: a scoping study. BMC Health Serv Res 2014 Jun 16;14:260 [FREE Full text] [doi: 10.1186/1472-6963-14-260] [Medline: 24934164]
32. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Dec;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
33. Kamsu-Foguem B, Foguem C. Telemedicine and mobile health with integrative medicine in developing countries. Health Policy Technol 2014 Dec;3(4):264-271. [doi: 10.1016/j.hlpt.2014.08.008]
34. Colaci D, Chaudhri S, Vasan A. mHealth interventions in low-income countries to address maternal health: a systematic review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
35. Khatun F, Heywood AE, Ray PK, Bhuiya A, Liaw S. Community readiness for adopting mHealth in rural Bangladesh: a qualitative exploration. Int J Med Inform 2016 Dec;93:49-56. [doi: 10.1016/j.ijmedinf.2016.05.010] [Medline: 27435947]
36. Hou J, Michaud C, Li Z, Dong Z, Sun B, Zhang J, et al. Transformation of the education of health professionals in China: progress and challenges. Lancet 2014 Aug;384(9945):819-827. [doi: 10.1016/S0140-6736(14)61307-6]
37. Wu Q, Zhao L, Ye X. Shortage of healthcare professionals in China. Br Med J 2016 Sep 22;354:i4860. [doi: 10.1136/bmj.i4860] [Medline: 27659864]
38. Hoque M. An empirical study of mHealth adoption in a developing country: the moderating effect of gender concern. BMC Med Inform Decis Mak 2016 Dec 3;16(1):51 [FREE Full text] [doi: 10.1186/s12911-016-0289-0] [Medline: 27142844]
39. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manage Inf Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
40. Mishra A, Anderson C, Angst C, Agarwal R. Electronic health records assimilation and physician identity evolution: an identity theory perspective. Inf Syst Res 2012 Sep;23(3-part-1):738-760. [doi: 10.1287/isre.1110.0407]
41. Dutta-Bergman M. Complementarity in consumption of news types across traditional and new media. J Broadcast Electron Media 2004 Mar;48(1):41-60. [doi: 10.1207/s15506878jobem4801_3]
42. Dutta-Bergman M. Interpersonal communication after 9/11 via telephone and internet: a theory of channel complementarity. New Media Soc 2016 Jun 30;6(5):659-673. [doi: 10.1177/1461448040470866]
43. Rains S, Ruppel E. Channel complementarity theory and the health information-seeking process: further investigating the implications of source characteristic complementarity. Commun Res 2013 Nov 11;43(2):232-252. [doi: 10.1177/0093650213510939]
44. Lee E, Han S. Determinants of adoption of mobile health services. Online Inf Rev 2015 Aug 10;39(4):556-573. [doi: 10.1108/OIR-01-2015-0007]
45. Mackert M, Mabry-Flynn A, Champlin S, Donovan E, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Dec 4;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
46. Zhang X, Han X, Dang Y, Meng F, Guo X, Lin J. User acceptance of mobile health services from users' perspectives: the role of self-efficacy and response-efficacy in technology acceptance. Inform Health Soc Care 2017 Mar;42(2):194-206 [FREE Full text] [doi: 10.1080/17538157.2016.1200053] [Medline: 27564428]
47. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 2016 Mar;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
48. Hossain MN, Okajima H, Kitaoka H, Ahmed A. Consumer acceptance of eHealth among rural inhabitants in developing countries (a study on portable health clinic in Bangladesh). Procedia Comput Sci 2017;111:471-478. [doi: 10.1016/j.procs.2017.06.049]
49. Carroll JS, Payne JW. The psychology of the parole decision process: a joint application of attribution theory and information-processing psychology. In: Cognition and Social Behavior. Oxford, England: Lawrence Erlbaum; 1976.

50. Sjovall AM, Talk AC. From actions to impressions: cognitive attribution theory and the formation of corporate reputation. Corp Reputation Rev 2004 Oct 1;7(3):269-281. [doi: 10.1057/palgrave.crr.1540225]
51. Tongji Hospital. About Tongji Hospital URL: http://english.tjh.com.cn/About-TJH/Index.shtml [accessed 2019-05-31] [WebCite Cache ID 78m9otxCZ]
52. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating m-health acceptance from a protection motivation theory perspective: gender and age differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 26.01.19; peer-reviewed by D Vogel, F Khatun, H Wu, H Yang; comments to author 28.04.19; revised version received 15.05.19; accepted 16.05.19; published 09.08.19.

Please cite as:
Ye Q, Deng Z, Chen Y, Liao J, Li G, Lu Y
How Resource Scarcity and Accessibility Affect Patients' Usage of Mobile Health in China: Resource Competition Perspective
JMIR Mhealth Uhealth 2019;7(8):e13491
URL: https://mhealth.jmir.org/2019/8/e13491/
doi:10.2196/13491
PMID:31400104

Original Paper

The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study

Isaac Vaghefi[1], PhD; Bengisu Tulu[2], PhD

[1]Seidenberg School of Computer Science and Information Systems, Pace University, New York, NY, United States

[2]Foisie Business School, Worcester Polytechnic Institute, Worcester, MA, United States

Corresponding Author:
Isaac Vaghefi, PhD
Seidenberg School of Computer Science and Information Systems
Pace University
1 Pace Plaza
New York, NY, 10038
United States
Phone: 1 (212) 346 1687
Email: sashrafvaghefi@pace.edu

Abstract

Background: Mobile health (mHealth) apps that support individuals pursuing health and wellness goals, such as weight management, stress management, smoking cessation, and self-management of chronic conditions have been on the rise. Despite their potential benefits, the use of these tools has been limited, as most users stop using them just after a few times of use. Under this circumstance, achieving the positive outcomes of mHealth apps is less likely.

Objective: The objective of this study was to understand continued use of mHealth apps and individuals' decisions related to this behavior.

Methods: We conducted a qualitative longitudinal study on continued use of mHealth apps. We collected data through 34 pre- and postuse interviews and 193 diaries from 17 participants over two weeks.

Results: We identified 2 dimensions that help explain continued use decisions of users of mHealth apps: users' assessment of mHealth app and its capabilities (user experience) and their persistence at their health goals (intent). We present the key factors that influence users' assessment of an mHealth app (interface design, navigation, notifications, data collection methods and tools, goal management, depth of knowledge, system rules, actionable recommendations, and user system fit) and relate these factors to previous literature on behavior change technology design. Using these 2 dimensions, we developed a framework that illustrated 4 decisions users might make after initial interaction with mHealth apps (to abandon use, limit use, switch app, and continue use). We put forth propositions to be explored in future research on mHealth app use.

Conclusions: This study provides insight into the factors that shape users' decisions to continue using mHealth apps, as well as other likely decision scenarios after the initial use experience. The findings contribute to extant knowledge of mHealth use and provide important implications for design of mHealth apps to increase long-term engagement of the users.

(JMIR Mhealth Uhealth 2019;7(8):e12983) doi:10.2196/12983

KEYWORDS

mobile health; mHealth; digital health; attrition; law of attrition; continued use; use decisions; goal persistence; IT assessment; smartphone; mobile app

Introduction

The use of smartphones to deliver health care services has been consistently on the rise for over a decade [1]. Accordingly, this topic has been attracting the attention of researchers and practitioners [2,3]. Given the pervasive and ubiquitous nature of smartphones, their powerful communication and interactive features, and access to internet, which brings an unbounded amount of health information, it is not surprising that mobile health (mHealth) apps have been a topic of investigation in health care. Using smartphones and various apps for transmitting electronic medical records [2], diagnosing and monitoring patients remotely [4], and implementing interventions that help people succeed in areas such as weight management, stress

management, smoking cessation, and dealing with chronic conditions [5-8] are examples of mHealth use. Recent surveys show that the market for these apps is rapidly evolving, bringing thousands of apps, aimed at various health purposes, to individuals at a minimal cost [9]. Despite the promise of mHealth apps, the use of mHealth tools has been limited [10], with reports suggesting that most individuals stop using them just before the fifth interaction, and a quarter of mHealth apps are used only once after installation [11,12]. Although promising in its value, it is less likely that the intended benefits of mHealth app use, such as improved access and quality of care, are going to be realized through such short-lived uses of the apps.

Studies that looked at individuals' adoption and use of mHealth have shown the importance of factors, such as users' motivation, existing health conditions, individual differences [7,13,14], and individuals' perceptions about usefulness and ease of use of mHealth [15,16]. Other studies highlighted the importance of design and persuasive nature of these apps and provided frameworks to guide theory-based design and development approaches [17,18]. Although these findings provide valuable insights on users' decisions to start using a technology, the question remains as to what guarantees the continued use of mHealth apps. This is a critical question, as research on the use of mHealth beyond the adoption phase is sparse [9] and anecdotal evidence has yielded inconsistent findings showing very minimal or no improvements in individuals' health because of adoption [6,19-21]. It is also a timely question, given the findings in information system (IS) research indicating that use of a system by itself will not be sufficient to provide the expected benefits [22,23]; an information technology (IT) tool should be used more than just a few times [24,25], or it should be used habitually [26], to deliver the expected positive outcomes, such as successful health behavior change. Current understanding of continued use of mHealth apps is limited and there is no study that directly focuses on this important issue.

As such, we conducted a qualitative, longitudinal, and exploratory study on continued use of health and wellness apps, a set of apps that are not disease-specific and aim to promote general wellness. The analysis revealed 2 important dimensions related to users' assessment of an mHealth app and its capabilities (user experience) and the users' persistence at achieving their health goals (intent). On the basis of these 2 dimensions, we proposed a 2×2 matrix to depict 4 type of users' decisions after adopting mHealth, which are to *abandon*, *limit*, *switch app*, and *continue use*. The results contribute to health informatics literature by providing a new perspective about how mHealth app use can be continued and the underlying key factors that could facilitate users' long-term engagement. For health care providers and mHealth app developers, these findings can be used to shape guidelines for better app design, whereas for users, they can ensure continued use decisions versus other possible choices.

The issue of adoption and use of technology has long been pursued by IS scholars. Earlier studies provided an overview of the basic predictors of adoption (see Venkatesh et al [27] for a review), mainly by adopting a cognitive perception-intention-use view to identify factors and antecedents that lead to adoption [28]. For instance, 2 prominent models, Technology Acceptance Model [29] and Unified Theory of Acceptance and Use of Technology [27], have illustrated the key role of several factors, including perceptions about a system's usefulness and ease of use, attitude [30], motivation [31], performance expectancy, effort expectancy, social influence, and facilitating conditions [27,32] on successful adoption of new technologies. Still, this literature argues that for a technology implementation to be considered successful, and for users to gain significant advantage, it is important that users continue to use a technology beyond the initial adoption stage [33]. Although the same models that were used to explain adoption can be helpful in describing users' postadoption behaviors [34], the accumulation of research knowledge shows that continued use of technology can be significantly different from initial adoption [35]. Hence, novel approaches were proposed to study the continued use of technology. For example, Bhattacherjee [35] proposed the expectation-confirmation model of continued use and explained that when expectations are positively confirmed via use experience, they can influence perceived usefulness and satisfaction, which increase continued use intentions. Although these studies provide valuable insights, the findings are intentionally abstract and general, so that they are applicable to a wide range of technologies.

mHealth technologies are designed to motivate and persuade behavior change to help users achieve their health and wellness goals. A number of studies have reported on the importance of design in persuasive technologies [36-38] and provided guidelines to improve the effectiveness of such solutions and their adoption [17]. Yet, to our knowledge, there is no study that focuses on the continued use of mHealth apps and the factors underlying this behavior. Our goal is to provide insights on this issue via an exploratory study, which we describe next.

Methods

Recruitment

Participants were recruited through an open call in 2 universities (1 public and 1 private) located in the Northeastern United States. We used purposive sampling [32] to select participants that satisfied the inclusion criteria: (1) own a smartphone, (2) be willing to use an mHealth app for at least 14 days, (3) not have had a past history of chronic diseases to ensure that the motivation of participants is not significantly different from typical, healthy mHealth users, and (4) have a specific health goal that can be reached using mHealth apps. We decided to focus on health and wellness mHealth apps because they represent the largest and most-commonly-used category of mHealth apps [3]. All interested individuals were screened over the phone or face-to-face to ensure their compliance with the inclusion criteria. We received approval from our Institutional Review Boards for the study approach described.

Data Collection

We used semistructured, face to face interviews and daily use diaries to collect data. The interview questions were developed based on a review of existing literature and further refined via discussions with 3 academic experts in health informatics.

During the first round of interviews (preuse interviews), participants answered questions about their approach to health and wellness, motivation to follow a healthy lifestyle, and level of confidence in making lifestyle changes to improve health. At the end of the preuse interviews, we asked participants to (1) identify a health and/or wellness goal toward which they wanted to work during the upcoming 14 days, (2) select a free mHealth app to use and download on their phones, and (3) describe how they are planning to use the new mHealth app.

Given that most users tend to withdraw from mHealth apps before the end of first week [39], we framed the longitudinal study for 14 consecutive days. Longitudinal studies vary in size and complexity, but the continuous monitoring of factors is common among all such studies [40]. This timeframe allowed participants to become familiar with the features of the app and decide whether they intended to continue using the app or abandon it. During the use period, we allowed participants to pick a new app to try if their original choice was not effective in helping them achieve their goals. Every evening, participants received an email with a link to the daily diary survey. The daily use diary included a single question, please describe your interactions with the app today? (eg, How many times and how long you used it? What features did you utilize? Any likes, dislikes?), to capture continuous data about app use. The diary data (see Multimedia Appendix 1 for a response example) allowed us to gather in-depth firsthand accounts of app use from the participants and reduce the likely effect of recall bias.

After the 14-day use period was over, participants were invited for the second round of interviews (postuse interviews). The participants were asked to describe their experience with the app, reflect back on their goals and motivation, assess if the app helped them achieve their goals, and discuss the reasons behind their positive or negative decisions to continue or withdraw use. Finally, the participants were asked to provide design suggestions for app developers that would improve their experience with mHealth apps and result in continued use. The pre- and postuse interviews took, on average, 21 min (13-47 min). The final dataset, collected from May to August 2017, included a total of 34 interviews and 193 daily usage diaries (some participants did not complete between 1 and 3 days of diary keeping). The participants received gift cards at the end of the closing interview.

Sample Characteristics

Participants were aged between 18 and 51 years, 70% (12/17) females, and 70% (12/17) iPhone users. Among 17 participants, 10 continued using the apps they picked during the initial interviews. However, 7 participants decided to try a new mHealth app because their original choices did not satisfy their needs. This is illustrated in Table 1 along with the participants' demographic information and areas they decided to focus on during the study.

Table 1. Study participant characteristics and goals.

Identification number	Age (years)	Sex	Phone	Selected apps (second choice, if changed)	Area of focus
1	23	F[a]	Android	MyFitnessPal	Diet
2	24	F	iPhone	ARise	Diet/physical activity
3	18	F	Android	Nike+ Training	Physical activity
4	50	M[b]	iPhone	Calorie Counter (Food Diary)	Diet
5	26	M	iPhone	Strong	Physical activity
6	35	F	iPhone	MapMyWalk	Physical activity
7	28	F	iPhone	Sleep Better (TracknShare LITE)	Sleep
8	51	F	iPhone	Weight Watchers	Diet
9	47	F	iPhone	Relax Lite	Mindfulness
10	33	F	iPhone	Aura and Headspace	Mindfulness
11	28	F	iPhone	5 Minute Home Workouts	Physical activity
12	29	F	iPhone	Female Fitness (Fitbit)	Physical activity
13	29	F	iPhone	Plant Nanny and Garmin	Diet (water)/physical activity
14	40	F	Android	Headspace (Aura)	Mindfulness
15	32	F	iPhone	Map my run (HabitBull)	Physical activity/habit building
16	24	M	Android	Samsung Health (Aura)	Physical activity/mindfulness
17	29	M	Android	Charity Miles (ASICS)	Physical activity

[a]Female.

[b]Male.

Data Analysis

All interviews and individual diaries (over 300 pages) were transcribed and added to QSR's NVivo application, which was used to code the data and conduct content analysis. To ensure anonymity, each participant was assigned a study identification number (ID). The analysis was performed using the grounded theory approach suggested by Strauss and Corbin (1998). During the open coding phase, we identified a total of 48 codes (eg, app customization, effort needed, reminder/alerts, motivation, activeness, content quality, context access, and continuance/discontinuance) that related to how participants described their use, assessed mHealth apps, and evaluated their willingness to keep using the app. During the axial coding phase, we discussed in multiple rounds how the 48 themes were related or distinct to determine overarching themes; the result was identification of 9 key dimensions that determined continued use of the app. Finally, focusing on the nature of decisions regarding continued use, we proposed a framework (2×2 matrix) based on users' *assessment of the app* in use and *persistence at health goals.* We provided qualitative appraisal of these dimensions based on information provided and usage patterns described by each respondent, and then mapped the results to the proposed framework.

Data Exclusion

Among the 19 individuals who were screened, 18 were eligible to participate in this study. A participant was excluded because of existing chronic conditions. Of the 18 eligible participants, 1 dropped out of the study during the 14-day use period and was not included in the analysis (final sample=17).

Results

Overview

The literature reports that when testing the influence of behavior change technologies on users' behavior, characteristics of users should be considered to make sense of the study results [41]. In other words, the user base for an intervention should meet the basic assumptions of the provided technology solution, such as participants' commitment to making behavioral changes in a specific domain (eg, diet or exercise). In this study, we selected people interested in using health and wellness mHealth apps and making behavioral changes based on a goal they chose, rather than a goal imposed on them. As we studied their interactions with the systems, we observed differences across 2 dimensions. One was their evaluation of the app based on the characteristics outlined earlier, whether enabling or not, and the other was their level of commitment to, and persistence in, achieving the goals they set for themselves. Previous studies focused on these 2 dimensions in isolation. The connection between these 2 dimensions emerged from the exploratory dataset revealing that continued mHealth app use can be influenced by (1) users' assessment of mHealth apps and (2) users' persistence at health goals.

Users' Assessment of Mobile Health Apps

The first dimension represents the initial user experience with mHealth apps and whether users have a positive or negative assessment of the capabilities of these apps. This dimension resembles previous study findings that highlight the role of satisfaction with IT use as an important step for users to extend the use of technology [35]. Yet, compared with the concept of satisfaction with technology use, which is all-inclusive and abstract and may apply to a wide range of issues related to user and technology interaction, we identified 9 factors and their subcategories that form users' assessment of mHealth apps through qualitative analysis and multiple iterations of coding. We present these factors, and some exemplary evidence (see Table 2) from the collected data, next.

Interface Design

Interface design related comments reflected participants' preference for *clean and simple* screens and a distaste for cluttered display and overwhelming *advertisements* on the screen. Although users' understanding of clean and simple may vary, the data revealed that when an interface is crowded with too much text or information, users have difficulty interacting with the app. However, a clean interface helps users navigate the app despite the complex nature of the app. This is relevant to the reduction principle, which suggests reducing complex behavior into simple tasks through the elimination of choices provided, and the liking principle, which suggests designing an interface that is appealing to users in persuasive system design (PSD) framework [17].

Navigation

Navigation (how users move through the menus and different features to accomplish their tasks) is another important factor that shapes users' opinions. The participants articulated their preference for an easy-to-understand *navigation menu and smooth flow* between screens of the app. As PSD framework suggests, reducing complexity is a critical principle to follow while designing successful, persuasive technologies.

The participants also expressed the need for *training* regarding the app's features and navigation menus (known as wizard) at the beginning of their interaction with the app. When participants reported working with wizards before use, they also reported that it helped them better engage with the app to achieve their goals (eg, participant 10 [P10]; or otherwise, withdraw quickly and look for another app, as they assessed that the app was not in line with their goals; eg, P3).

Table 2. Factors influencing decision to continue use.

Factors	Evidence from data
Interface: clean and simple design; appearance of advertisements	"I liked the way the dashboard looked. It was just so clean and so I said 'Alright well I'll download that and give it a try!' and I've been really happy with [using] it." [P12]; "I think the app had a lot of ads, and I know they have to make their money...when I was trying to add something, an ad keeps trying to pop up, it was frustrating." [P13]
Navigation: navigation menu and flow of pages; training and wizards	"I don't have to find all these different buttons and how to navigate through it. It's very simple to use. So, whereas the other app when there's so many different features...I don't have time to go through all of them." [P13]; "Having some sort of quick tutorial orientation...you have to have that...I find it helpful for most apps, so I understand what it does." [P7]
Notifications: alerts and reminders; control over alerts	"I would expect it to give me text updates or notifications, so I don't have to go into the app. Because if I have to go into the app [to check my progress] then I would be less likely to check...if it alerts me that would be wonderful." [P2]; "I mean it actually had a feature that you could set reminders. But...I don't like any notifications turned on...To me, it is always a distraction...it may work for others." [P5]; "I think it made me more active, especially because this [referring to a wearable device] has little red lights that pop up. So that kind of forced me hey I haven't walked for a while or maybe I'll do a walk around the building!" [P13]
Data collection: data entry convenience; need for extra device	"I mean part of the reason why the step app worked so well was that you literally turn it on it does everything. There isn't really a lot I need to do to interact with it further." [P6]; "Yes, that's the only thing I don't like right now is that I don't generally have pockets to carry my phone with me. So, I don't think it's accurately reflecting my step count. But if you carry it around it definitely would." [P12]
Goal management: setting up goals; notifications about progress	"I thought that was one of their big positives. For this app, I think the customizable side of it and being able to track exactly what I wanted is probably its biggest feature and something that I've been missing in other apps." [P7]; "You can click this and then you can go look at your trends over the past several days where here it's giving you the hourly trend or weekly ones." [P16]; "You can click this and then you can go look at your trends over the past several days where here it's giving you the hourly trend or weekly ones." [P16]; "The app, for instance, sent me emails saying that 'You have recorded your nutrition for seven days!' which I found pretty motivating. Kept me going!" [P8]
Depth of knowledge: available content; accuracy of data and content; completeness	"If an [nutrition] app had links to websites that explains how to ferment vegetables, or...links to helpful resources or articles recipes would help me more to get there." [P4]; "They have a lot of information and you can see kind of like during the night if it spikes when you woke up and it was pretty accurate that way and you could feel like a dream journal and put in you know if you had caffeine late and things like that to kind of track if that affects your sleep." [P7]; "So, the one app I had initially downloaded I thought had too much locked content, and I felt like I didn't have enough options. So, I deleted that, and I did download the other app." [P12]; "But there seem to be no consistent rules. It was overly complicated. I'm like I don't know how kids would play this. And there was no help document for me to read...and it was vastly inconsistent in terms of the content." [P6]; "I don't know if it's a bug or if it's supposed to be that way but if you have to pause it doesn't work and it's like you didn't even do it." [P14]
System rules: process of the app; clarity of rules and functions	"But there seem to be no consistent rules. It was overly complicated. I'm like I don't know how kids would play this. And there was no help document for me to read...and it was vastly inconsistent in terms of the content." [P6]; "I don't know if it's a bug or if it's supposed to be that way but if you have to pause it doesn't work and it's like you didn't even do it." [P14]
Actionable recommendations: personalized progress analysis; amount of usage time needed	"It did help me become more conscientious about getting some food in me three times a day at least. Becoming more aware of how many calories I was taking in. So, I could meet my goals." [P4]; "The only thing that [was needed] is to send me related notifications like '2000 steps from your goal for the day'." [P12]
Fit between user and system: match between features and user needs	"I think this app would work for a lot of people. For me what they provide value in, like in their add-ons, does not work. I get [the value] in other places already. So, if I didn't do podcasts, that would be a really nice way to introduce you to walking and running." [P6]; "From the notification, I knew how I was doing. So, it was nice because I wasn't doing anything extra to get this information. I didn't have to go in and really use the app where, with Aura [another app], I had to actively go in and open it up and make the three four minutes for each meditation." [P17]

Notifications

Notifications (*alerts and reminders* triggered by the app) encourage participants to use the app and make their interactions with the app efficient while increasing the likelihood of achieving their goals [17]. Most participants admit that they prefer simple notifications in the forms of smartphone alerts, text messages, or emails that provide a quick overview of their progress toward their health goals.

Although notifications could help motivate more use, the participants prefer some *control* over the frequency and type of notifications they receive. Depending on their preference or motivation to reach their personal goals, some users prefer more frequent notifications that could motivate continued use, whereas others prefer fewer notifications and consider them a distraction. For the latter group of users, notifications could become a deterrent to app use leading to use abandonment.

Data Collection

Data collection methods and tools utilized by the mHealth apps are expected to be *convenient* and require low effort, according to the participants. Automatic data collection without the need for users' frequent input, typically through passive sensing and use of a tracker, can influence the extent of their use and willingness to continue and expand usage of the app.

The analysis showed that efficiency can be achieved when users can quickly interact with a system to perform the intended task. In this study's context, when an app allows for automatic or quick data entry, it facilitates efficiency in use. This is consistent with the reduction principle in PSD framework [17]. Examples in the data include apps that collect health-related data (eg, step counting) or use guides/templates to speed up the data entry process (eg, default or latest inputs inserted for the user). In addition, all issues related to interface, notification, and navigation directly or indirectly help facilitate and speed up users' interaction with the app, which all contribute to efficiency.

In cases where manual data entry is unavoidable (eg, tracking diet), users seek features that provide convenient data entry, such as nutrition apps with comprehensive databases that provide nutritional values for a variety of options. Moreover, wearing and *carrying an additional device* all the time acts as a barrier toward full utilization of the system, as some may consider it an additional burden.

Goal Management

Goal management is a necessary functionality that enables users to reach their goals [17,18]. The participants expected the apps to allow them to *set goals* and track their performance against their goals. Depending on the context of use, this could translate into setting up daily, weekly, monthly, or longer-term goals that can be used as a reference point for assessing one's performance. This is related to the self-monitoring principle in PSD framework, which suggests that system features allowing users to keep track of their own performance facilitate progress toward their goals [17]. Examples in the data were defining cups of daily water intake, number of steps to walk, or number of week days to exercise or meditate. Although setting up goals appears to be necessary, many pointed to the need for flexibility in setting up goals, that is, users prefer to have alternative measurement tools and scales to pursue goals. This is related to the tailoring principle of persuasive design, which suggests that tailoring an app to the needs of its users improves its persuasiveness [17]. As an example, a diet management app would need to be able to record weight in pounds or kilograms and use other measurement scales, such as body fat or body mass index (ie, customized goal setting).

In addition, participants expressed that to be able to follow up with their goals, they needed the app to send regular *notifications about their progress*. These notifications, such as text messages, push pop-ups, or emails, help users stay motivated and on track with their goals over time. The analysis revealed that to facilitate continued use, users preferred a progress report (if possible visualized) that hinted at what needed to be done to reach predefined user goals.

Depth of Knowledge

Depth of knowledge provided, which refers to the freely *available content* in the app, was a key theme brought up by many of the participants. They distinguished between how much content is freely accessible and the amount of valuable knowledge represented in the app. For instance, a participant expected the meditation app to provide substantial quality content that was freely accessible. The expectation of free quality content appeared to be important when a premium (paid) version of the app was available. The *accuracy* of the information and knowledge represented in the app emerged as another key issue. The analysis showed that when users were not confident about the accuracy of data collected by the app, they would be hesitant to act upon the provided information and move toward the intended goals. Finally, perceived *completeness* of the content provided through the app could determine the extent of users' interaction with the app. When users perceive the content in the app *sufficient* or *more* than what they believe is needed to reach their personal goals, they tend to use it. This is consistent with the expertise principle discussed in PSD framework [17]. However, if the app does not provide sufficient information regarding achieving goals, or requires a payment to unlock the content, especially if users believe that the amount is unjustified, it blocks the opportunity for continued use because of loss of credibility.

System Rules

The clarity of the system rules embedded in the technology, in other words the way the system is designed to work, emerged as a major theme in the analysis. Although the specifics may vary in each app and context, users expect to easily understand the process underlying the design of the app. When the *rules and functions* represented in the app are clear, the user can make an informed decision to commit to the app and continue using it toward achieving personal goals. If users do not understand the process of how a system works based on their initial interactions, they will not be motivated to continued use.

Actionable Recommendations

Although the participants used apps with different focuses and features, they stated that to realize the benefits, the apps needed to provide actionable recommendations for improving the current conditions and specify what needed to be done to reach the goals in the intended timeline. This refers to the system's ability to offer clear next steps that users can follow to reach their goals and is relevant to the tunneling principle in PSD framework, which suggests guiding users during the change process by providing means for action that helps them get closer to their goals [17]. When the app does not summarize user data and provide reports, individuals will have difficulty making sense of their actions, and hence, will not be motivated to continued use. Providing *personalized progress analysis* (a large range of relevant analyses based on the data collected from users) can assist users in tracking their goals and keep them motivated and engaged with the app, leading to continued use.

The *amount of usage time* needed to interact with an mHealth app, for instance, to input data, check progress, or get feedback, to receive accurate progress reports and actionable recommendations toward reaching health goals should be aligned with users' expectations. On the basis of the goals set by the participants, the time needed for interacting with their apps varied. Our analysis showed that to continue use, it was important that the users perceived the required amount of time needed to work with the app as adequate. When there is inconsistency between the time a user can allocate to the app versus the time required by the app, users tend to withdraw from it after a period of time. As personal mHealth app use is not

mandatory, it is important that users be able to spend sufficient time with this technology to reach the intended outcome.

Fit Between User and System

Finally, although every app embeds a different set of features and provides various functionalities, there should be a *fit between user and system*, defined in terms of the match between user attributes and app attributes. Our analysis shows that when there is a good match between user attributes (such as preferences, expectations, and personality traits) and app attributes (such as interface, features, content, navigation, and rules), the app's use will be continued. When the app offers users an opportunity to achieve their health goals, users may extend and continue their use of the app beyond the first few interactions.

Users' Persistence at Health Goals

The second dimension related to continued use that emerged from the analysis was the intent of the users. In addition to the factors related to the technology being used, motivation of users played a vital role in continuing to use an mHealth app. Having persistence at intended health goals and being able to pursue them despite the likely challenges appeared to play a key role toward continued use of mHealth apps. Behavior change is difficult to achieve, even with the use of persuasive technologies [17]. For instance, 1 of the participants reported that although she found that the app had powerful features, she did not have enough motivation to continue use beyond the scope of the study. In our sample, although all users committed to a health goal at the beginning of the study (see Table 1), the analysis of the exit interviews revealed that not all demonstrated persistence toward the goal throughout the course of the study (see Multimedia Appendix 2 for examples). The mHealth apps we studied are autogenous technologies that people choose to use to change their own behavior [17]; hence, the use is completely voluntary and self-motivated. Previous research has shown that when there is no mandate to enforce the use of a technology, people have less tendency or positive attitude toward using that technology [42], unless internally motivated to do so [43]. In such a situation—common to mHealth apps—those who have higher persistence toward reaching their goals (stronger intent) appear to have longer continued engagement with the technology.

Typology of Mobile Health App Use Decisions

The results revealed 2 dimensions that were related to the continued use of mHealth apps. The first dimension considers an overall assessment of user experience and how an mHealth app provides opportunities for reaching health goals through the factors identified in the results section. This subjective assessment made by users can vary from high to low, depending on the extent to which technology is enabling them to achieve the intended goals. For instance, a user may believe an app is a high enabler because it provides notifications, has a simple interface, and allows for automatic data entry techniques, whereas another user may believe the same app is a low enabler because of the insufficient health information the app provides. The second dimension considers the intent of a user by assessing their level of commitment to their health goals. This is usually demonstrated by assessing the extent to which the users exhibit undivided attention and persistent efforts toward achieving goals and could range from low to high.

Considering these 2 dimensions, we identified 4 possible scenarios as an outcome of users' initial experiences with mHealth apps, which are the decisions to (1) abandon use, (2) limit use, (3) switch app, and (4) continue use.

Abandon Use

The decision to abandon quadrant (low assessment and low persistence) represents a situation where users are skeptical about the capability of the mHealth apps they selected, but, at the same time, they do not show persistence at their health goals. In such conditions, we expect a user to abandon the mHealth app before having any meaningful interaction with it. An example is a user who shows willingness to *improve own's well-being* but withdraws from this goal when faced with an obstacle, for instance, going 1 day without exercising, while simultaneously assessing the app as insufficient for reaching that health goal. As a respondent put it:

> *I feel like having only one means of communication or accountability is not good for me. I think if I'm serious about it, then I need to go to the meetings and be more engaged. Even though the system holds me accountable for it, it was not enough.* [P8]

Limit Use

The decision to limit use quadrant (high assessment and low persistence) refers to a situation where an enabling mHealth app is available, yet users do not show persistence for pursuing their goals and stop use when they experience any difficulty. An example is a user who reports having a short and intermittent interaction with the selected app, although he or she found the app suitable for reaching the goals. In this situation, we expect users to have a limited use of the app, insufficient for significant improvement toward achieving goals.

Switch App

The decision to switch quadrant (low assessment and high persistence) refers to a situation where users show commitment toward their goals but find the mHealth app to be a low enabler because of limited features of the app. In such situations, we expect a user to continue use but substitute that app for a better choice. For instance, a respondent admitted that "it [my use] depends on whether I find the app useful or not, because meditation is something that seems really helpful for me and I'd like the idea, but the [app] implementation isn't working so well…so, I'll go try something else and see." (10). Although substituting can be distracting, it can still provide an opportunity for reaching goals if a better mHealth app is found and then used in the future.

Continued Use

The decision to continue use is indeed the ideal situation, where the app is enabling and users show persistence toward goals. Under this condition, we expect the users to continue engagement with the app for longer periods, which eventually helps them move toward their intended goals. Figure 1 illustrates

our assessment of the participants (the assigned ID), according to the qualitative review of their data.

On the basis of these findings, we developed 4 propositions that describe circumstances associated with the decisions mHealth users make.

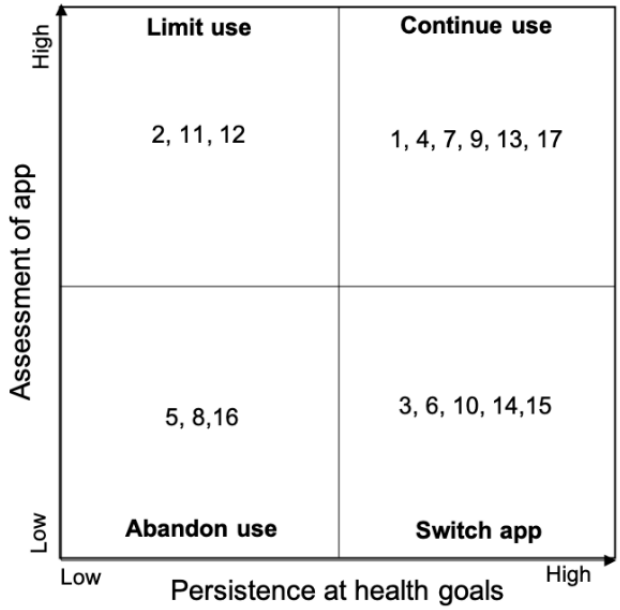
Proposition 1: When individuals have low assessment of the mHealth app and low persistence at their health goals, they will abandon using the mHealth app.

Proposition 2: When individuals have high assessment of the mHealth app and low persistence at their health goals, they will limit their use of the mHealth app.

Proposition 3: When individuals have low assessment of the mHealth app and high persistence at their health goals, they will switch to a different mHealth app of their choice.

Proposition 4: When individuals have high assessment of the mHealth app and high persistence at their health goals, they will continue to use the mHealth app.

Figure 1. Use decision scenarios regarding mobile health app use.



Discussion

Principal Findings

Despite the exponential rate at which new mHealth apps are introduced to market, most users stop usage soon after initial use. The aim of this study was to further our understanding of continued use of mHealth apps. Through the analyses of qualitative data collected via interviews and daily use diaries, we identified key factors that influenced users' decisions regarding continued use after the initial interaction with an app. Furthermore, based on the degree of users' assessment of the app and their persistence toward their goals, we highlighted 4 decisions: to abandon use, to limit use, to switch app, and to continue use. We put forth propositions that can guide future research that aims to understand behaviors regarding the use of mHealth apps.

Strengths

mHealth apps will continue to play a pivotal role in providing individual and customized health care services that can be reached anywhere, any time and at relatively low costs [7]; yet, the challenges regarding individuals' short-term use of these apps impede achieving the intended outcomes and making behavior changes. The exploratory study revealed that following PSD principles [17] can help improve the design of future technologies. In line with *less is more* recommendations in human-computer interaction literature [44], the results highlighted the importance of clean and simple interfaces as the gateway for users to have direct and straightforward interactions with mHealth apps. Clear rules, easy navigation through different parts of a system, and navigation wizards encourage continued use. Automatic data collection and simple data entry methods such as automatic food suggestions improve not only users' interactions with the app but also their satisfaction with the experience. Ultimately, when there is a good fit between users' needs and mHealth apps, continued use is likely to occur. Nonetheless, these practical implications should be considered with caution, as it is very challenging to consider all these factors at the same time when designing an mHealth app. Therefore, future research can investigate varying conditions and app-related characteristics that are relevant to each or a subset of factors promoting continued usage, which will provide a more granular view of the identified factors. Overall, these findings echo the calls for user-centered and goal-directed design approaches in the previous research [45]. The findings presented these design principles and concepts as perceived by the users during their decision-making process. Although the study provides important insights to better understand the underlying factors of continued mHealth app use, further research is needed, for instance, using surveys and larger sample, to test the effect of these relationships in various mHealth use contexts.

More importantly, the findings highlighted the importance of focusing on users' goals and their commitment to these goals.

Although motivation is shown to be sufficient for adoption of mHealth [13,46], we found that users who had persistence throughout the usage period were more likely to continue using the app and experience its positive outcomes [33]. As the use of mHealth apps is typically voluntary, their use should be proactively pursued. The study illustrates that persisting at goals while using the right system that fits users' needs could facilitate continued use, which could pave the way for achieving improved health outcomes. At the same time, we reveal other use scenarios that could result in the absence of goal persistence or lack of fit.

Limitations and Future Research

We acknowledge that this study has limitations. First, more than half of the participants were female, used an iPhone, and were highly motivated to take care of their health. These characteristics may have influenced the way they interacted with, and made decisions about, the mHealth app. Including larger dataset in a population (eg, balanced male/female; iPhone/Android; and motivated/unmotivated) will help improve the generalizability of the findings. Second, although we used 2 methods to collect longitudinal data, the provided information was self-reported and did not include objective measures. Using system log data in future studies may help provide additional insights on continued use of mHealth apps. Third, we focused on a limited range of mHealth apps (ie, health and wellness) as representative of mHealth apps. Future research is needed to replicate and extend the results to other contexts to have a more inclusive view of the continued use of mHealth. Fourth, participants were recruited in universities (although only 1 was a student) and received compensation for their participation.

In addition, we sent a daily reminder to participants to fill out their use diary, which could have influenced their interaction with the app and encouraged continued use. In the same vein, we acknowledge that other factors could influence continued use of mHealth app. For instance, previous research has shown the importance of privacy regarding adoption and use of mHealth app [47], yet privacy was mentioned only marginally in the dataset. Although such limitations are a common issue in exploratory research [32], further confirmatory studies are needed to validate and generalize the results in broader settings.

Finally, we note that the data focused on continued use as one way to successfully change the health behavior of individuals, yet we did not directly assess the behavior change of users per se. This presents a promising avenue for additional studies, for instance, using longer scope and more comprehensive data collection methods that pay specific attention to the relationship between continued mHealth use and health behavior change to assess how the former behavior instigates the latter.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A diary response example.
[PDF File (Adobe PDF File), 112KB - mhealth_v7i8e12983_app1.pdf]

Multimedia Appendix 2
Coding structure in qualitative analysis.
[PDF File (Adobe PDF File), 233KB - mhealth_v7i8e12983_app2.pdf]

References

1. Smith A. Pew Research Center. 2017. Record Shares of Americans Now Own Smartphones, Have Home Broadband URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2019-08-08]
2. 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]
3. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 1;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
4. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mhealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
5. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
6. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
7. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
8. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]

9. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
10. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
11. Perez S. TechCrunch. 2016. Nearly 1 in 4 People Abandon Mobile Apps After Only One Use URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/ [accessed 2019-01-20]
12. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and use of mhealth apps: a study from England. Pharmacy (Basel) 2017 Jun 14;5(2):pii: E33 [FREE Full text] [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
13. Cocosila M, Archer N. An Empirical Investigation of Mobile Health Adoption in Preventive Interventions. In: Proceedings of the 22nd BLED eConference eEnablement: Facilitating an Open, Effective and Representative eSociety. 2009 Presented at: BLED'09; June 14-17, 2009; Bled, Slovenia p. 189-202.
14. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 2;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
15. Wu JH, Wang SC, Lin LM. Mobile computing acceptance factors in the healthcare industry: a structural equation model. Int J Med Inform 2007 Jan;76(1):66-77. [doi: 10.1016/j.ijmedinf.2006.06.006] [Medline: 16901749]
16. Mallenius S, Rossi M, Tuunainen VK. Factors Affecting the Adoption and Use of Mobile Devices and Services by Elderly People—Results from a Pilot Study. In: Proceedings of 6th Annual Global Mobility Roundtable. 2007 Presented at: Mobility Roundtable'07; June 1-2, 2007; Los Angeles, California p. 12.
17. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1). [doi: 10.17705/1CAIS.02428]
18. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. London, UK: Ubiquity Publication; 2002.
19. Abebe NA, Capozza KL, des Jardins TR, Kulick DA, Rein AL, Schachter AA, et al. Considerations for community-based mhealth initiatives: insights from three Beacon communities. J Med Internet Res 2013 Oct 15;15(10):e221 [FREE Full text] [doi: 10.2196/jmir.2803] [Medline: 24128406]
20. 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]
21. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mhealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
22. Burton-Jones A, Volkoff O. How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. Inf Syst Res 2017 Sep;28(3):468-489. [doi: 10.1287/isre.2017.0702]
23. Burton-Jones A, Grange C. From use to effective use: a representation theory perspective. Inf Syst Res 2013 Sep;24(3):632-658. [doi: 10.1287/isre.1120.0444]
24. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
25. Bhattacherjee A, Premkumar G. Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. MIS Q 2004;28(2):229-254. [doi: 10.2307/25148634]
26. Limayem M, Hirt SG, Cheung CM. How habit limits the predictive power of intention: the case of information systems continuance. MIS Q 2007;31(4):705-737. [doi: 10.2307/25148817]
27. Venkatesh V, Morris M, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
28. Schwarz A, Chin W. Looking forward: toward an understanding of the nature and definition of IT acceptance. J Assoc Inf Syst 2007 Apr;8(4):230-243. [doi: 10.17705/1jais.00123]
29. Davis C, Claridge G. The eating disorders as addiction: a psychobiological perspective. Addict Behav 1998;23(4):463-475. [doi: 10.1016/S0306-4603(98)00009-4] [Medline: 9698975]
30. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
31. Davis FD, Bagozzi RP, Warshaw PR. Extrinsic and intrinsic motivation to use computers in the workplace. J Appl Soc Psychol 1992 Jul;22(14):1111-1132. [doi: 10.1111/j.1559-1816.1992.tb00945.x]
32. Patton MQ. Qualitative Research & Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
33. Jasperson J, Carter P, Zmud RW. A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. MIS Q 2005 Dec 19;29(3):525-557. [doi: 10.2307/25148694]
34. Kim SS, Malhotra NK, Narasimhan S. Two competing perspectives on automatic use: a theoretical and empirical comparison. Inf Syst Res 2005 Dec;16(4):418-432. [doi: 10.1287/isre.1050.0070]
35. Bhattacherjee A. An empirical analysis of the antecedents of electronic commerce service continuance. Decis Support Syst 2001 Dec;32(2):201-214. [doi: 10.1016/S0167-9236(01)00111-7]

36. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L, Abeele VV. A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. JMIR Mhealth Uhealth 2016 Oct 13;4(4):e118 [FREE Full text] [doi: 10.2196/mhealth.6286] [Medline: 27742604]
37. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
38. Pinzon OE, Iyengar MS. Persuasive technology and mobile health: a systematic review. In: Bang M, Ragnemalm EL, editors. Persuasive Technology: Design for Health and Safety: 7th International Conference on Persuasive Technology, Persuasive'12, Linköping, Sweden, June 6-8, 2012 Proceedings. New York City: Springer; 2012:45-48.
39. Fleit B. Lean Plum. 2016. Can Push Notifications Really Help App Retention? Find Out URL: https://www.leanplum.com/blog/can-push-notifications-really-help-app-retention-find-out/ [accessed 2019-01-12]
40. Coggon D, Barker D, Rose G. Epidemiology for the Uninitiated. Hoboken, NJ: John Wiley & Sons; 2009.
41. 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. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 3307-3316. [doi: 10.1145/2470654.2466452]
42. Karahanna E, Straub DW, Chervany NL. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. MIS Q 1999 Jun;23(2):183-213. [doi: 10.2307/249751]
43. van der Heijden H. User acceptance of hedonic information systems. MIS Q 2004;28(4):695-704. [doi: 10.2307/25148660]
44. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the User Interface: Strategies for Effective Human-Computer Interaction. India: Pearson; 2010.
45. Williams A. User-Centered Design, Activity-Centered Design, and Goal-Directed Design: A Review of Three Methods for Designing Web Applications. In: Proceedings of the 27th ACM International Conference on Design of Communication. 2009 Presented at: SIGDOC'09; October 5-7, 2009; Bloomington, Indiana, USA p. 1-8. [doi: 10.1145/1621995.1621997]
46. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V. A generalised adoption model for services: a cross-country comparison of mobile health (m-health). Gov Inform Q 2016 Jan;33(1):174-187. [doi: 10.1016/j.giq.2015.06.003]
47. Arora S, Yttri J, Nilse W. Privacy and security in mobile health (mhealth) research. Alcohol Res 2014;36(1):143-151 [FREE Full text] [Medline: 26259009]

Abbreviations

ID: identification number
IS: information system
IT: information technology
mHealth: mobile health
PSD: persuasive system design
PX: participant X

Edited by G Eysenbach; submitted 28.11.18; peer-reviewed by S Feldman, Y Wang, A Paglialonga, C Fernández, C Mamillapalli; comments to author 31.03.19; revised version received 25.06.19; accepted 21.07.19; published 29.08.19.

Please cite as:
Vaghefi I, Tulu B
The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study
JMIR Mhealth Uhealth 2019;7(8):e12983
URL: http://mhealth.jmir.org/2019/8/e12983/
doi:10.2196/12983
PMID:31469081

Original Paper

Is Wearable Technology Becoming Part of Us? Developing and Validating a Measurement Scale for Wearable Technology Embodiment

Elizabeth C Nelson[1], BA, MS; Tibert Verhagen[2], PhD; Miriam Vollenbroek-Hutten[1,3], PhD; Matthijs L Noordzij[4], PhD

[1]Biomedical Signals and Systems, University of Twente, Enschede, Netherlands

[2]Center for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[3]Ziekenhuis Groep Twente, Almelo, Netherlands

[4]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Elizabeth C Nelson, BA, MS
Biomedical Signals and Systems
University of Twente
PO Box 217
Enschede,
Netherlands
Phone: 31 642683161
Email: e.c.nelson@utwente.nl

Abstract

Background: To experience external objects in such a way that they are perceived as an integral part of one's own body is called embodiment. Wearable technology is a category of objects, which, due to its intrinsic properties (eg, close to the body, inviting frequent interaction, and access to personal information), is likely to be embodied. This phenomenon, which is referred to in this paper as *wearable technology embodiment*, has led to extensive conceptual considerations in various research fields. These considerations and further possibilities with regard to quantifying *wearable technology embodiment* are of particular value to the mobile health (mHealth) field. For example, the ability to predict the effectiveness of mHealth interventions and knowing the extent to which people embody the technology might be crucial for improving mHealth adherence. To facilitate examining *wearable technology embodiment*, we developed a measurement scale for this construct.

Objective: This study aimed to conceptualize wearable technology embodiment, create an instrument to measure it, and test the predictive validity of the scale using well-known constructs related to technology adoption. The introduced instrument has 3 dimensions and includes 9 measurement items. The items are distributed evenly between the 3 dimensions, which include body extension, cognitive extension, and self-extension.

Methods: Data were collected through a vignette-based survey (n=182). Each respondent was given 3 different vignettes, describing a hypothetical situation using a different type of wearable technology (a smart phone, a smart wristband, or a smart watch) with the purpose of tracking daily activities. Scale dimensions and item reliability were tested for their validity and Goodness of Fit Index (GFI).

Results: Convergent validity of the 3 dimensions and their reliability were established as confirmatory factor analysis factor loadings (>0.70), average variance extracted values (>0.50), and minimum item to total correlations (>0.40) exceeded established threshold values. The reliability of the dimensions was also confirmed as Cronbach alpha and composite reliability exceeded 0.70. GFI testing confirmed that the 3 dimensions function as intercorrelated first-order factors. Predictive validity testing showed that these dimensions significantly add to multiple constructs associated with predicting the adoption of new technologies (ie, trust, perceived usefulness, involvement, attitude, and continuous intention).

Conclusions: The wearable technology embodiment measurement instrument has shown promise as a tool to measure the extension of an individual's body, cognition, and self, as well as predict certain aspects of technology adoption. This 3-dimensional instrument can be applied to mixed method research and used by wearable technology developers to improve future versions through such things as fit, improved accuracy of biofeedback data, and customizable features or fashion to connect to the users'

personal identity. Further research is recommended to apply this measurement instrument to multiple scenarios and technologies, and more diverse user groups.

(JMIR Mhealth Uhealth 2019;7(8):e12771) doi:10.2196/12771

KEYWORDS

embodiment; wearable technology; measurement development; human technology interaction; eHealth; mHealth; wearable electronic devices; self-help devices; health information technology; medical informatics

Introduction

There has been an impressive increase in the usage of wearable technologies, digital devices that incorporate wireless connectivity and allow the user to seamlessly access, interact with, and exchange information anywhere and anytime [1], since their introduction into the marketplace [2,3]. Devices such as smart phones, activity trackers, and smart watches have been widely embraced [2,3] and seem to have become almost inseparable from the human body. Although various research fields address this phenomenon from their own perspectives, there seems to be consensus that users can interact with technology, accepting it as part of them and even experience it as part of their body [4]. Beyond body extension, users can extend their cognitive performance [5,6] through the constant access to information [7], self-identity [8-12], and through the highly personal experience [13,14] tailoring it to their personal preferences [15]. However, studies addressing this form of embodiment represent a relatively new area of focus and more empirical research is called for [4,16,17]. In addition, no study has yet combined the different embodiment experiences (body, cognition, and sense of self) to cover the full spectrum of the individual.

The embodiment experiences could be highly relevant for the study and use of wearable technology in health care. A substantial body of existing research addressed wearable technology in, among others, studies into health information recording [18,19], mood, sleep [20], personal sensing and biofeedback in mental health care [21-23], remote patient monitoring [24], medication adherence [25], and technology-assisted procedures [26]. Furthermore, there is an increasing use of wearable technologies that seamlessly fit into the user's everyday lifestyle, can be worn on the body or mated with human skin, and continuously and closely monitor the user's motion and vital signs (eg, pulse and blood pressure) [20,27], and as such provides the user with the information needed for self-assessment and change in health behaviors and health outcomes [18,28,29]. Generating insights into embodiment experiences and how to measure it may prove crucial for researchers and practitioners to further their understanding of what drives users to keep wearing the technology on the long term and be adherent to the health coaching associated with its measurements.

Past research on devices such as virtual reality [16] or cognitive prostheses [30] have addressed the embodiment of technology but were unable to measure it because of the lack of a measurement instrument [4,16,17]. In this paper, we aim to address this gap in the literature by proposing and validating the concept of *wearable technology embodiment* and operationalize it by developing a valid measurement instrument utilizing established conceptualization and measurement procedures [31]. This delineation implies that researchers and practitioners can use the instrument to measure the embodiment of some of the most widely adopted wearable technologies in the market today such as smartphones, activity trackers, and smart watches, as well technologies such as smart clothing/jewelry, head-mounted displays, and ear-worn technology [32].

Methods

Scale Development Procedure

To develop and test a measurement instrument for wearable technology embodiment, we followed an established scale development procedure [31,33-36] (Table 1). Scale development is a recognized process for developing and validating a definition and measurement scale for a construct that cannot be adapted from a similar scale or does not yet exist.

Table 1. Overview of scale development procedure.

Step and description	Actions undertaken in this study
1 Conceptualization: develop a conceptual definition of the construct	Conceptualization of target construct; scoping review; study selection; data extraction; define property; define entity; establish dimensionality of construct; construct definition
2 Development of measures: generate items to represent the construct and assess the content validity of the items	Item generation and sorting; expert interviews; item refinement
3 Method of validation: formally specify the measurement model	Formally specify the measurement model; include dependent variables for measurement
4 Scale evaluation and refinement: collect data, scale purification and refinement	Evaluate goodness of fit; assess validity at the construct level; assess reliability at the item level; eliminate problematic indicators
5 Validation: assess scale validity	Assess convergent validity; assess discriminant validity; test alternative models; test predictive validity

Step 1: Conceptualization

Conceptualization of Target Construct, Literature Review and Study Selection

To begin, we conducted a scoping review focused on the specific experience of embodiment of technologies worn or carried, for example, mobile phones or smart watches, searching for terms describing the experience of embodiment with a tool (Textbox 1). The scoping review included: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consultation [37,38]. Our review resulted in a total of 80 papers in disciplines that included electronic health, neuroscience, mobile computing, wearable computing, ubiquitous computing, psychology, sociology, and philosophy. Although some areas specifically describe the embodiment of technology, other research described the embodiment of tools such as Merleau Ponty's example of a blind person embodying their walking cane [39]. We believe these examples also apply to the embodiment of technologies. Therefore, studies which included the embodiment of a technical or nontechnical tool were included in the review. In total, 20 of the papers were discarded because they described embodiment without the use of technology or tools and therefore did not help in developing the concept. The 60 remaining papers were organized and analyzed for their description of the embodiment of technology/tools.

Textbox 1. Search terms used in scoping review.

Google Scholar, Science Direct, Science.gov, SpingerLink, WorldWideScience, JSTOR, and Web of Science

- Embodiment
  - AND (tool OR technology OR digital OR wearable OR mobile OR cognitive)
- Embodied interaction
  - AND (tool OR technology OR digital OR wearable OR mobile)
- Prosthesis
  - AND (cognitive OR embodiment OR technology OR digital OR wearable OR mobile)
- Phenomenology
  - AND **(**wearable OR mobile OR digital OR technology tool OR cognitive)

Data Extraction and Establishing the Entity and the Property

A structure was utilized to organize the study characteristics of the 60 papers, including: (1) authors and publication year, (2) main research findings, (3) research design details *(ie, experimental or nonexperimental design)*, (4) item embodied (ie, mobile and wearable technology or physical prosthetic), (5) description of the specific kind of embodiment, (6) measurement items, and (7) embodiment dimensionality. Using this literature review, we could establish the entity (the whom/what) and the property (the relevant process/aspect) of the construct, which created the foundation for our construct definition [31,35,36,40]. The entity here is obviously a person. We described the property as experiencing wearable technology and perceiving an extension of oneself [4,17,41,42].

Dimensionality of the Construct and Construct Definition

After having defined the property and the entity of wearable technology embodiment, we explored the dimensionality of the concept [43-45]. Although most of the examined literature suggested the existence of multiple dimensions, there seemed to be no overall consensus on either the number or naming of the dimensions. We therefore decided to uncover the most plausible categorization by systematically collecting, juxtaposing, and comparing possible dimensions as mentioned in previous studies. This process, which is also known as structured conceptualization [44], resulted in the emergence of 3 clear themes: *body extension*, *cognitive extension*, *and self-extension*. These themes were not only the most popular topics but also covered the full spectrum of the individual (ie, body, mind, and sense of self). Body extension refers to a physical addition or replacement of the body. For example, a robotic hand that communicates touch to the end of the human limb, improving dexterity [46]. Cognitive extension, sometimes called cognitive prostheses [17,47], refers to the experienced extension of one's cognitive capabilities such as navigation assistance or knowledge of the number of steps taken during a day. Finally, self-extension refers to an object being perceived as part of a person's identity or sense of self. For example, experiencing a mobile phone as a representation or extension of yourself and personalizing the technology to be congruent with your self-image [11,15,48,49].

We then combined the entity (a human individual), property (experiencing wearable technology and perceiving an extension), and dimensionality of the construct (body extension, cognitive extension, and self-extension) to concisely define our construct [31]. This led to the following definition of *wearable technology embodiment*: A person experiencing technology worn on or near the body perceiving a certain extension of the body, cognition, or (sense of) self.

Step 2: Development of Measures

Item Generation and Sorting

To come up with a first set of measurement items, we started the process of item generation, sorting, and selection (for a schematic overview of the item sorting and selection process, see Multimedia Appendix 1.) From our literature review, a first set of 24 preliminary items (12 body extensions, 7 cognitive extensions, 5 self-extensions; see Multimedia Appendix 2) was generated from the embodiment descriptions [43-45]. Some scales from the prosthesis literature were adapted to describe the extension of the body, which resulted in more items within this dimension. We then followed Hinkin and Tracey's [50] content validity assessment approach to adapt and refine the wording of the 24 items. In total, 2 members of the research team independently assessed the items within each dimension and then compared their assessments to determine whether they tied in with our definition, sat in the correct dimension, and read clearly and concisely [31,51,52]. During this process, 6 items were removed (items 8 to 12 and 19, see Multimedia Appendix 2), which led to an updated item pool of 18 items.

Expert Interviews and Item Refinement

To judge the content validity of the 18 measurement items, we made use of 8 expert interviews [31,43,52,53] to examine each item for comprehension, applicability, and fit into the construct's dimension [54]. In total, 2 of the experts worked as researchers at universities within human-computer interaction, one expert was an information systems researcher at a university and specialized in measurement scale development, 3 experts worked in wearable technology innovation and development, and 2 experts were individuals who were recently given wearables for the first time.

The experts were interviewed either in person or through a Skype conference call. After receiving a brief description and definition of wearable technology embodiment, they completed a Web-based survey, rating the applicability of each item from 1 (very inapplicable) to 5 (very applicable) and whether the item fits in the dimension. The experts chose whether they believed the item should be: (1) kept as is, (2) modified in a minor way, (3) modified in a major way, or (4) omitted [55]. Each item included an open field for suggestions or considerations [56]. The interviews ended with a brief discussion regarding the classification of 3 dimensions and possible suggestions for any new measurement items. The experts approved the dimensionality of our scale (ie, body extension, cognitive extension, and self-extension) and no new items were recommended. On the basis of their input, a few items were slightly reworded to improve their linguistic clarity. Using the feedback of the experts, 5 items were removed (items 5 to 7, 18, and 24; see Multimedia Appendix 2) because of repetition or lack of fit with the construct. This resulted in an updated measurement item pool of 13 items, which were subsequently used for empirical testing.

Step 3: Method of Validation

To test the preliminary measurement instrument, data were collected through a vignette-based survey (see Multimedia Appendix 3). The sample consisted of a group of 182 undergraduate business students attending an e-Business course at a university in The Netherlands. Participation was voluntary and the students were offered a small incentive of 5 extra credit points on an exam worth 100 points by including their student number in the survey. Each respondent was given 3 different vignettes, describing the hypothetical situation using a different type of wearable technology with the purpose of tracking daily activities: a smart phone, a smart wristband, and a smart watch. The 3 devices were chosen because: (1) they are typical examples of wearable technology used in everyday life, being the most widely adopted technologies worn on or near the body, (2) they match our conceptualization and definition of *digital devices that incorporate wireless connectivity and allow the user to seamlessly access, interact with, and exchange information anywhere and anytime*, and (3) they fit into our delineation of user-centered preventive care wearable technology that can continuously and closely monitor the user's motion and vital signs. Overall, the experience of participants with smart phones and relative unfamiliarity with activity tracking, smart wristbands, and smart watches supported our decision to make use of a vignette-based research design to confront them with the hypothetical use of wearable technology.

Each vignette asked the respondents to project themselves into the scenario of using the technology habitually to track their daily activities such as steps, hours of sleep, calorie burn, and achievement of personal health goals, regularly checking their progress throughout the day. The vignettes included pictures of the technology without brands to make the impression of the scenario as accessible as possible while avoiding branding bias. We randomized the order of the 3 vignettes for the respondents to avoid order bias. At the end of each vignette, the respondents were directed to a Web-based survey containing the 13 embodiment items (grouped into the 3 dimensions) and sociodemographic questions: age, gender, experience using the wearable technology (Table 2). To test the predictive validity of the construct [44], multi-item measurement instruments for trust, involvement, perceived usefulness, attitude toward use, and continuous intention also were included (see Multimedia Appendix 4 and full survey questions in attached documents). Before starting with a vignette, each respondent was told the length of time for the survey, who the investigator was, and the purpose of the study. The Web-based survey was pretested before sending it to the students, and included less than 8 questions per page, which were randomized to decrease order bias. Students were able to review or change their answers while filling in the survey. IP addresses were checked for multiple submissions and only completed questionnaires were analyzed. All surveys were checked for appropriate completion times.

Table 2. Sample characteristics (n=182).

Variables	Smartphone	Smart wristband	Smart watch
Technology use, n (%)			
Own and use quite often to track activity	37 (20.1)	5 (2.7)	5 (2.7)
Own but use seldom to track activity	72 (39.4)	7 (3.8)	6 (3.3)
Own but do not use to track activity	72 (39.4)	3 (1.7)	3 (1.7)
Do not own	2 (1.1)	168 (91.8)	169 (92.3)
Age, n (%)			
18-20	—[a]	49 (26.9)	—
21-23	—	105 (57.7)	—
24-26	—	24 (13.2)	—
27-30	—	4 (2.2)	—
Gender, n (%)			
Female	—	68 (37.4)	—
Male	—	114 (62.6)	—

[a]Same distribution.

Results

Step 4: Scale Evaluation and Refinement

Evaluate Goodness of Fit, Validity at Construct and Item Level, Eliminate Problematic Indicators

To test the dimensionality and further refine the scale items, we ran an exploratory factor analysis (EFA, principal component analysis with varimax rotation) on the set of 13 measurement items [57]. Aggregating the data of the 3 vignettes, each with 182 respondents, led to a sample that contained 546 responses. In total, 3 items were removed from the analysis (items 4, 16, and 17; see Multimedia Appendix 2) as they loaded substantially on 2 dimensions or more [58]. Rerunning the EFA with the 10 remaining items confirmed the 3 dimensions of wearable technology embodiment (Kaiser-Meyer-Olkin Measure of sampling adequacy ((0.80) Bartlett's test of sphericity 2390, P<.001) and accounted for 71.22% of the variance. All items loaded significantly on only 1 dimension and all factor loadings were above the recommended threshold value of 0.50 [58]; therefore, providing first evidence of the convergent and discriminant validity of the measurement instrument. As the largest factor within the EFA explained less than 50% of the variance (28.5%), evidence for common method bias was not found [31,59].

To further test the measurement instrument, we performed a confirmatory factor analysis (CFA) [60,61] using the software package IBM, SPSS Amos 23 [57] (maximum likelihood estimation). We tested the EFA solution of 3 dimensions as intercorrelated first-order factors [61]. See Table 3 for: Chi square degrees of freedom calculated probability, Minimum Discrepancy Degrees of Freedom (CMIN/ *df*), Goodness of Fit (GFI), Adjusted Goodness of Fit (AGFI)), Normed Fit Index, Incremental Fit Index, Tucker Lewis Index (TLI), Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), Akaike information criterion, Browne-Cudeck Criterion (BCC), Bayesian Information Criterion (BIC). After removing one item (item 23, see Multimedia Appendix 2) to improve the model fit, the found 9-item solution demonstrated a good fit with the data (CMIN/*df*) <5; GFI, AGFI, Normed Fit Index Incremental Fit Index (NFI), TLI, CFI >0.90; RMSEA <0.08; Table 3).

Table 3. Confirmatory factor analysis alternative model testing.

Model	Chi square (*df*)	*P* value	CMIN/df[a]	GFI[b]	AGFI[c]	NFI[d]	IFI[e]	TLI[f]	CFI[g]	RMSEA[h]	Akaike information criterion	BCC[i]	BIC[j]
3 first-order correlated	104.26 (24)	<.001	4.345	.96	.93	.94	.96	.94	.96	.078	146.26	147.05	236.62
3 first-order uncorrelated	273.79 (27)	<.001	10.141	.91	.85	.86	.87	.83	.87	.130	309.79	310.46	387.24
One first-order factor	812.73 (27)	<.001	30.101	.72	.53	.60	.61	.48	.61	.231	848.73	849.41	926.18

[a]CMIN/df: Minimum Discrepancy Degrees of Freedom.

[b]GFI: Goodness of Fit.

[c]AGFI: Adjusted Goodness of Fit Index.

[d]NFI: Normed Fit Index.

[e]IFI: Incremental Fit Index.

[f]Tucker Lewis Index.

[g]CFI: Comparative Fit Index.

[h]RMSEA: root mean square error of approximation.

[i]BCC: Browne-Cudeck Criterion.

[j]BIC: Bayesian Information Criterion.

Step 5: Assess Scale Validity

Test Alternative Models, Assess Convergent and Discriminant Validity

To further test the applicability of our dimensions, we tested 2 alternative models [52]: a model of 3 uncorrelated first-order factors and a model treating the 9 items as indicators of 1 first-order factor. The CFA results of the alternative models clearly showed that the alternative models did not have a good fit with the data (CMIN/*df*>5; GFI, AGFI, NFI, TLI, CFI<0.90; RMSEA<0.08; Akaike information criterion, BCC, BIC>scores of 3 first-order correlated model). These outcomes confirm that wearable technology embodiment is best modeled as a set of 3 correlated, first-order factors. In addition, the very poor fit of the 1 first-order factor reconfirmed the absence of common method bias.

The convergent validity of the 3 dimensions and their reliability (Table 4) was confirmed via the CFA factor loadings (>0.70) [58], average variance extracted (AVE) values (>0.50) [53], and minimum item to total correlations (>0.40) [62], which exceeded the established threshold value. The reliability of the dimensions was also confirmed as Cronbach alpha and composite reliability exceeded 0.70. The discriminant validity (Table 5).could be confirmed as the AVE of each construct exceeded the values of the crossconstruct squared correlations [63,64] .

Table 4. Convergent validity: Factor loadings, Cronbach alphas, composite reliabilities, (average variance extracted), and minimum item to total correlation.

Dimension and item	Factor loading (CFA)	Cronbach alpha	Composite reliability	Average variance extracted	Minimum item to total correlation
Body extension	—[a]	0.84	0.88	0.71	0.76
When using a <technology> it feels like it is part of my body	0.83	—	—	—	—
When using a <technology> it feels like it is an extension of my body	0.74	—	—	—	—
When using a <technology> it almost feels like it is incorporated into the body	0.86	—	—	—	—
Cognitive extension	—	0.72	0.84	0.64	0.80
Using <technology> heightens my knowledge about my activity	0.61	—	—	—	—
Using <technology> helps me learn about my activity	0.84	—	—	—	—
Using <technology> helps me gain understanding of my activity	0.62	—	—	—	—
Self-extension	—	0.86	0.88	0.71	0.76
When using a <technology> it feels like it is an extension of myself	0.76	—	—	—	—
When using a <technology> it feels like it is related to my sense of self	0.86	—	—	—	—
When using a <technology> it feels like it is a psychological extension of myself	0.81	—	—	—	—

[a]Not applicable.

Table 5. Discriminant validity testing: average variance extracted (italics) versus crossconstruct squared correlations between the constructs.

Constructs	Body extension	Cognitive extension	Self-extension	Trust	Involvement	Perceived usefulness	Attitude toward use	Continuous intention
Body extension	*0.87* [a]	—[b]	—	—	—	—	—	—
Cognitive extension	0.13	*0.80*	—	—	—	—	—	—
Self-extension	0.53	0.12	*0.88*	—	—	—	—	—
Trust	0.24	0.50	0.14	*0.86*	—	—	—	—
Involvement	0.13	0.11	0.33	0.01	*0.91*	—	—	—
Perceived usefulness	0.17	0.60	0.15	0.61	0.13	*0.86*	—	—
Attitude toward use	0.27	0.41	0.25	0.53	0.29	0.60	*0.84*	—
Continuous intention	0.18	0.24	0.25	0.35	0.49	0.40	0.59	*0.90*

[a]Italic scores (diagonal) are the average variance extracted of the individual constructs.

[b]Not applicable.

Test of Predictive Validity

To test the predictive validity of the wearable technology embodiment instrument, we utilized 5 dependent variables related to technology adoption: trust [55], involvement [65], perceived usefulness [66], attitude toward use [67], and continuous intention [68]. Structural equation modeling was used to test the extent to which the 3-dimensional wearable technology embodiment construct explained each of the 5 dependent variables by making use of the software package IBM SPSS Amos 23 [57] (maximum likelihood estimation).

The results of the analyses (see Multimedia Appendix 5) confirm a good fit with the data as the GFI, AGFI, NFI, TLI, and CFI exceed the recommended value of 0.90 and the RMSEA does not surpass the value of 0.08 [58]. The 3 dimensions of wearable technology embodiment explain considerable (trust and perceived usefulness) to acceptable amounts of the variance (involvement, attitude, and continuous intention) of the dependent variables [69]. Except for the influence of body extension on the attitude (beta=.17: *P*=.004), the standardized paths imply that the found influences account for a substantial proportion of the variance [70] (beta>.20). Overall, the results confirm the predictive validity of both the multidimensional construct and individual dimensions of the wearable technology embodiment instrument.

Discussion

Utilizing the development and validation process as prescribed by Mackenzie et al [31], we made use of literature study, expert interviews, and empirical data collected for 3 wearable technologies to conceptualize *wearable technology embodiment* and build a measurement instrument. We established wearable technology embodiment as a 3-dimensional concept consisting of the dimensions: body extension, cognitive extension, and self-extension.

Academic Implications

The findings of this study have 2 implications for academic research. First, the development of a measurement instrument serves researchers by quantifying the perception of wearable technology extending the user's body, cognitive capacity, and sense of self. Researchers could develop a mixed methods approach to extend qualitative findings or compare usage data in upcoming studies. Second, the results of the predictive validity testing suggest that wearable technology embodiment significantly adds to well-known constructs that have been applied previously to study the adoption and use of new technology (ie, trust, involvement, perceived usefulness, attitude toward use, and continuous intention). By adding to these constructs, wearable technology embodiment seems to function as a valuable extension of theoretical structures such as the technology acceptance model [71], theory of reasoned action [72], and expectation confirmation theory [73].

Practical Implications

Our results also have practical implications. Given that the dimensions of wearable technology embodiment seem to contribute positively to perceptions of trust, involvement, usefulness, and behavioral attitudes and intentions, wearable technology developers could benefit from this knowledge by developing devices in such as a way that they better fit the user's body shape (body extension), improve the acceptability of biofeedback data [74] (cognitive extension), and heighten customizable features and fashion [75-77], connecting to the user's personal identity (self-extension). Users are likely to evaluate the wearable technology more positively with customizations focused on body, cognition, and self-extension. Furthermore, wearable technology developers could make use of the outcomes of the predictive validity at the dimension level to further prioritize their efforts. For example, if the objective is to generate more trust in the technology, it seems advisable to focus on cognitive extension(s) as this was the strongest trust determinant in our model. When the aim is to generate more involvement, however, a focus on designing and developing self-extension(s) seems a better choice. Overall, integrating the 3 wearable technology embodiment dimensions into design and development priorities can aid practitioners in making more effective decisions.

Limitations and Recommendations

This study has been subject to a couple of limitations that could guide scholars in setting up future research. First, the sample consisted of a rather homogeneous group of students that can

be classified as millennials, most likely raised with technology [78], who are active users of emerging technologies [79]. The selection and use of this sample seems justified given that it reduces the likelihood that differences between the respondents such as age, educational background, and technological savviness may have biased our findings [80,81]. Furthermore, it is in line with the key objective of our work to setup and test a theoretically meaningful construct instead of generalizing found research effects to larger populations [82]. This is not to say, however, that research would not benefit from using the developed measurement instrument in future effect application studies with different, more heterogeneous samples [80]. We foresee this research as a next step in the field of technology embodiment studies.

Second, by making use of a vignette method, we were able to confront the respondents with the situation of using different forms of wearable technology. The use of the vignette method has several advantages. It simulates realism, can be tailored to the specific research problem [83], does not require respondents to have in-depth knowledge of the presented stimuli, and reduces the likelihood of confounding effects since participants respond to the same stimulus [84]. Still, it cannot reflect all facets of actual usage situations and we, therefore, suggest researchers to apply and crossvalidate the instrument in a real-word context.

Third, this study was framed within the context of using wearable technology to track daily activities such as movement, hours of sleep, calorie burn, and personal health goals. Past studies have highlighted that low adherence to mobile interventions is a common occurrence [85,86], yet the opportunity to measure and address health concerns is evident [64]. Our findings regarding embodiment positively influence constructs related to technology adoption (ie, trust, involvement, usefulness, attitude, and continuous intention) and suggest that this embodiment scale could give insight into a determination as to which individuals will adhere to the intervention. Predicting adherence levels and identifying individuals unlikely to adhere could help in understanding and possibly improving the low adherence rates during mobile and wearable health interventions.

Fourth, even though we do find that wearable technology embodiment functions as a determinant of constructs rooted in different theoretical frameworks, the focus of this study was on conceptualization and measurement and not on the extension of nomological networks. More theoretical rationale and validation are needed to substantiate our findings. We invite researchers to adopt the concept of wearable technology embodiment in their future studies.

Fifth, this study focused on some of the most popular wearables in consumer technology, which are used for preventative health care (ie, smart phone, activity tracker, and smart watch). Other devices that serve this purpose include smart textiles, tattoos, and jewelry. Next to these kind of wearables, the developed measurement instrument also could be applied to user-centered disease monitoring wearable devices such as wearable cameras that enhance chronic disease self-management [87], insulin monitors and pumps in the treatment of diabetes [19], smart gloves that assist rheumatoid arthritis patients in applying therapy [88], and medical-grade electrocardiogram wristwatches that assist cardiac patients to detect heart arrhythmia [89]. Furthermore, although our inquiry did not focus on care provider centered wearables, it seems logical that the measurement instrument could apply to wearable aids that are used during medical procedures such as Google glasses in surgery [28]. Still, more empirical exploration is needed to validate the applicability of our instrument for these kinds of wearable technology. We encourage researchers to do so in future studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flow diagram measurement item selection.
[PNG File, 83KB - mhealth_v7i8e12771_app1.png]

Multimedia Appendix 2
Measurement items in stages of development.
[PNG File, 161KB - mhealth_v7i8e12771_app2.png]

Multimedia Appendix 3
Survey vignettes-smartphone, smart wristband, and smart watch.
[PNG File, 676KB - mhealth_v7i8e12771_app3.png]

Multimedia Appendix 4
Constricts and items predictive validity testing.
[PNG File, 175KB - mhealth_v7i8e12771_app4.png]

Multimedia Appendix 5
Results predictive validity testing.

[PNG File, 103KB - mhealth_v7i8e12771_app5.png]

References

1. Bower M, Sturman D. What are the educational affordances of wearable technologies? Comput Educ 2015 Oct;88:343-353. [doi: 10.1016/j.compedu.2015.07.013]
2. Lamkin P. Forbes. 2018 Feb 22. Smartwatch Popularity Booms With Fitness Trackers on the Slide URL: https://www.forbes.com/sites/paullamkin/2018/02/22/smartwatch-popularity-booms-with-fitness-trackers-on-the-slide/ [accessed 2018-03-20]
3. Wilson HJ, Daugherty PR. Harvard Business Review. 2018 Mar 21. AI Will Change Health Care Jobs for the Better URL: https://hbr.org/2018/03/ai-will-change-health-care-jobs-for-the-better [accessed 2018-04-02]
4. Ugur S. Wearing Embodied Emotions: A Practice Based Design Research on Wearable Technology. New York: Springer; 2013.
5. Ishikawa T, Fujiwara H, Imai O, Okabe A. Wayfinding with a GPS-based mobile navigation system: a comparison with maps and direct experience. J Environ Psychol 2008 Mar;28(1):74-82. [doi: 10.1016/j.jenvp.2007.09.002]
6. Gallagher S. The socially extended mind. Cogn Syst Res 2013 Dec;25-26:4-12. [doi: 10.1016/j.cogsys.2013.03.008]
7. Okazaki S, Mendez F. Perceived ubiquity in mobile services. J Interact Mark 2013 May;27(2):98-111. [doi: 10.1016/j.intmar.2012.10.001]
8. Hogle LF. Enhancement technologies and the body. Annu Rev Anthropol 2005 Oct;34(1):695-716 [FREE Full text] [doi: 10.1146/annurev.anthro.33.070203.144020]
9. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Crit Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
10. Pitts V. Feminism, technology and body projects. Womens Stud 2005 Apr;34(3-4):229-247. [doi: 10.1080/00497870590964129]
11. Sultan F, Rohm A, Gao T. Factors influencing consumer acceptance of mobile marketing: a two-country study of youth markets. J Interact Mark 2009 Nov;23(4):308-320. [doi: 10.1016/j.intmar.2009.07.003]
12. Shilling C. The Body And Social Theory. California, USA: Sage Publications; 2012.
13. Lee D, Moon J, Kim YJ, Yi MY. Antecedents and consequences of mobile phone usability: linking simplicity and interactivity to satisfaction, trust, and brand loyalty. Inf Manag 2015 Apr;52(3):295-304. [doi: 10.1016/j.im.2014.12.001]
14. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Dec;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
15. Kang H, Shin W, Tam L. Differential responses of loyal versus habitual consumers towards mobile site personalization on privacy management. Comput Human Behav 2016 Mar;56:281-288. [doi: 10.1016/j.chb.2015.11.013]
16. Hoffmann L, Krämer NC. Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. Int J Hum Comput Stud 2013 Jul;71(7-8):763-774. [doi: 10.1016/j.ijhcs.2013.04.007]
17. de Preester H. Technology and the body: the (im)possibilities of re-embodiment. Found Sci 2011;16(2-3):119-137. [doi: 10.1007/s10699-010-9188-5]
18. Ananthanarayan S, Siek KA. Persuasive Wearable Technology Design for Health and Wellness. In: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops. 2012 Presented at: PervasiveHealth'12; May 21-24, 2012; San Diego, CA URL: https://ieeexplore.ieee.org/abstract/document/6240399 [doi: 10.4108/icst.pervasivehealth.2012.248694]
19. Seto E, Ware P, Logan AG, Cafazzo JA, Chapman KR, Segal P, et al. Self-management and clinical decision support for patients with complex chronic conditions through the use of smartphone-based telemonitoring: randomized controlled trial protocol. JMIR Res Protoc 2017 Nov 21;6(11):e229 [FREE Full text] [doi: 10.2196/resprot.8367] [Medline: 29162557]
20. Chu HC, Liu YM, Kuo FR. A mobile sleep-management learning system for improving students' sleeping habits by integrating a self-regulated learning strategy: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 29;6(10):e11557 [FREE Full text] [doi: 10.2196/11557] [Medline: 30373729]
21. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 Dec 8;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
22. Derks YP, Visser TD, Bohlmeijer ET, Noordzij ML. mHealth in mental health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. Int J Hum Factors Ergon 2017;5(1):61-92. [doi: 10.1504/IJHFE.2017.088418]
23. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing rest-activity profiles in schizophrenia using wearable and mobile technologies: development, implementation, feasibility, and acceptability of a remote monitoring platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
24. van Lier HG, Oberhagemann M, Stroes JD, Enewoldsen NM, Pieterse ME, Schraagen JM, et al. Design decisions for a real time, alcohol craving study using physio- and psychological measures. Int Conf Persuas Technol 2017:3-15. [doi: 10.1007/978-3-319-55134-0_1]

25. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 9;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
26. Wei NJ, Dougherty B, Myers A, Badawy SM. Using google glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018 Mar 6;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
27. 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]
28. Haghi M, Thurow K, Stoll R. Wearable devices in medical internet of things: scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
29. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016 Jan 21;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
30. Clark A. Natural-Born Cyborgs: Minds, Technologies, And The Future Of Human Intelligence. Oxford, UK: Oxford University Press; 2001.
31. MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Q 2011;35(2):293-334 [FREE Full text] [doi: 10.2307/23044045]
32. Costello K. Gartner. 2018. Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019 URL: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-
33. Alber JM, Bernhardt JM, Stellefson M, Weiler RM, Anderson-Lewis C, Miller MD, et al. Designing and testing an inventory for measuring social media competency of certified health education specialists. J Med Internet Res 2015 Sep 23;17(9):e221 [FREE Full text] [doi: 10.2196/jmir.4943] [Medline: 26399428]
34. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
35. Nunnally JC, Bernstein I. Psychometric Theory. New York: McGraw-Hill; 1994.
36. Spector PE. Summated Rating Scale Construction: An introduction. Newbury Park: Sage Publications; 1992.
37. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017;19(8):e277. [doi: 10.2196/jmir.7288] [Medline: 28765103]
38. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
39. Falk B, Merleau-Ponty M, Lefort C, Lingis A. The visible and the invisible. Philos Q 1970 [FREE Full text]
40. Schwab DP. Construct validity in organizational behavior. Res Organ Behav 1980;2:3-43 [FREE Full text]
41. Giummarra MJ, Gibson SJ, Georgiou-Karistianis N, Bradshaw JL. Mechanisms underlying embodiment, disembodiment and loss of embodiment. Neurosci Biobehav Rev 2008;32(1):143-160. [doi: 10.1016/j.neubiorev.2007.07.001] [Medline: 17707508]
42. Overton WF. Relationism and relational developmental systems: a paradigm for developmental science in the post-Cartesian era. Adv Child Dev Behav 2013;44:21-64. [doi: 10.1016/B978-0-12-397947-6.00002-7] [Medline: 23834001]
43. Garaus M, Wagner U. Retail shopper confusion: conceptualization, scale development, and consequences. J Bus Res 2016 Sep;69(9):3459-3467. [doi: 10.1016/j.jbusres.2016.01.040]
44. Wolfinbarger M, Gilly MC. eTailQ: dimensionalizing, measuring and predicting etail quality. J Retail 2003;79(3):183-198. [doi: 10.1016/S0022-4359(03)00034-4]
45. Zhang X, Ko M, Carpenter D. Development of a scale to measure skepticism toward electronic word-of-mouth. Comput Human Behav 2016 Mar;56:198-208. [doi: 10.1016/j.chb.2015.11.042]
46. Raspopovic S, Capogrosso M, Petrini FM, Bonizzato M, Rigosa J, Di Pino G, et al. Restoring natural sensory feedback in real-time bidirectional hand prostheses. Sci Transl Med 2014 Feb 5;6(222):222ra19 [FREE Full text] [doi: 10.1126/scitranslmed.3006820] [Medline: 24500407]
47. Wynn T. The cognitive life of things. In: Prefiguring Cyberculture: An Intellectual History. Cambridge: The MIT Press; 2014.
48. Ho SY. The attraction of internet personalization to web users. Electron Mark 2006;16(1):41-50. [doi: 10.1080/10196780500491162]
49. Vincent J. Emotional attachment and mobile phones. Know Techn Pol 2006 Mar;19(1):39-44. [doi: 10.1007/s12130-006-1013-7]
50. Hinkin TR, Tracey JB. An analysis of variance approach to content validation. Organ Res Methods 1999;2(2):175-186. [doi: 10.1177/109442819922004]
51. Lewis BT, Templeton GF, Byrd TA. A methodology for construct development in MIS research. Eur J Inf Syst 2005;14(4):388-400. [doi: 10.1057/palgrave.ejis.3000552]
52. Verhagen T, ven den Hooff B, Meents S. Toward a better use of the semantic differential in IS research: an integrative framework of suggested action. J Assoc Inf Syst 2015 Feb;16(2):108-143. [doi: 10.17705/1jais.00388]

53. Baldus BJ, Voorhees C, Calantone R. Online brand community engagement: scale development and validation. J Bus Res 2015 May;68(5):978-985. [doi: 10.1016/j.jbusres.2014.09.035]
54. Koc M, Barut E. Development and validation of new media literacy scale (NMLS) for university students. Comput Hum Behav 2016 Oct;63:834-843. [doi: 10.1016/j.chb.2016.06.035]
55. van der Heijden H, Verhagen T, Creemers M. Understanding online purchase intentions: contributions from technology and trust perspectives. Eur J Inf Syst 2003;12(1):41-48. [doi: 10.1057/palgrave.ejis.3000445]
56. Smith AK, Bolton RN, Wagner J. A model of customer satisfaction with service encounters involving failure and recovery. J Mark Res 1999 Aug 5;36(3):356-372. [doi: 10.2307/3152082]
57. Arbuckle JL. Amos Development. New York: IBM; 2014. AmosTM 23 Users Guide URL: http://amosdevelopment.com/download/User_Guide_23.pdf
58. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis, 7th Edition. London: Pearson; 2010.
59. Siemsen E, Roth A, Oliveira P. Common method bias in regression models with linear, quadratic, and interaction effects. Organ Res Methods 2009 Nov 30;13(3):456-476. [doi: 10.1177/1094428109351241]
60. Netemeyer RG, Bearden WO, Sharma S. Scaling Procedures: Issues and Applications. Thousand Oaks, CA: Sage Publications; 2003.
61. Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Q 1994;18(4):453-461. [doi: 10.2307/249524]
62. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. J Acad Mark Sci 1998;26(1):6-15. [doi: 10.1177/0092070398261002]
63. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 1981;18(3):382-388. [doi: 10.1177/002224378101800313]
64. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
65. Koufaris M. Applying the technology acceptance model and flow theory to online consumer behavior. Inf Syst Res 2002 Jun;13(2):205-223. [doi: 10.1287/isre.13.2.205.83]
66. Fang X, Chan S, Brzezinski J, Xu S. Moderating effects of task type on wireless technology acceptance. J Manag Inf Syst 2006;22(3):123-157 [FREE Full text] [doi: 10.2753/MIS0742-1222220305]
67. Verhagen T, Feldberg F, van den Hooff B, Meents S, Merikivi J. Understanding users' motivations to engage in virtual worlds: a multipurpose model and empirical testing. Comput Hum Behav 2012 Mar;28(2):484-495. [doi: 10.1016/j.chb.2011.10.020]
68. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
69. Henseler J, Ringle CM, Sinkovics RR. The use of partial least squares path modeling in international marketing. In: New Challenges to International Marketing. West Yorkshire: Emerald Group Publishing Limited; 2009:277-319.
70. Chin WW. Issues and opinion on structural equation modeling. MIS Q 1998;22(1):7-16 [FREE Full text]
71. Davis FD. DSpace@MIT. 1986. A technology acceptance model for empirically testing new end-user information systems: theory and results URL: https://dspace.mit.edu/handle/1721.1/15192
72. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York: Routledge; 2015.
73. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 1980;17(4):460-469. [doi: 10.1177/002224378001700405]
74. Kumari P, Mathew L, Syal P. Increasing trend of wearables and multimodal interface for human activity monitoring: a review. Biosens Bioelectron 2017 Apr 15;90:298-307. [doi: 10.1016/j.bios.2016.12.001] [Medline: 27931004]
75. Juhlin O, Zhang Y, Wang J, Andersson A. Fashionable Services for Wearables: Inventing and Investigating a New Design Path for Smart Watches. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. 2016 Presented at: NordiCHI'16; October 23-27, 2016; Gothenburg, Sweden. [doi: 10.1145/2971485.2971505]
76. Silina Y, Haddadi H. New Directions in Jewelry: A Close Look at Emerging Trends & Developments in Jewelry-Like Wearable Devices. In: Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: ISWC'15; September 7-11, 2015; Osaka, Japan p. 49-56. [doi: 10.1145/2802083.2808410]
77. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: an empirical smart wristband study. Comput Human Behav 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]
78. Hur HJ, Lee HK, Choo HJ. Understanding usage intention in innovative mobile app service: comparison between millennial and mature consumers. Comput Human Behav 2017 Aug;73:353-361. [doi: 10.1016/j.chb.2017.03.051]
79. Jones C, Ramanau R, Cross S, Healing G. Net generation or digital natives: is there a distinct new generation entering university? Comput Educ 2010 Apr;54(3):722-732. [doi: 10.1016/j.compedu.2009.09.022]
80. Peterson RA, Merunka DR. Convenience samples of college students and research reproducibility. J Bus Res 2014 May;67(5):1035-1041. [doi: 10.1016/j.jbusres.2013.08.010]
81. Ashraf R, Merunka D. The use and misuse of student samples: an empirical investigation of European marketing research. J Consumer Behav 2016;16(4):295-308. [doi: 10.1002/cb.1590]

82. Lucas WJ. Theory-testing, generalization, and the problem of external validity. Sociol Theory 2003;21(3):236-253 [FREE Full text] [doi: 10.1111/1467-9558.00187]
83. Atzmüller C, Steiner PM. Experimental vignette studies in survey research. Methodology 2010;6(3):128-138. [doi: 10.1027/1614-2241/a000014]
84. Hughes R, Huby M. The application of vignettes in social and nursing research. J Adv Nurs 2002;37(4):382-386. [doi: 10.1046/j.1365-2648.2002.02100.x] [Medline: 11872108]
85. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
86. Argent R, Daly A, Caulfield B. Patient involvement with home-based exercise programs: can connected health interventions influence adherence? JMIR Mhealth Uhealth 2018 Mar 1;6(3):e47 [FREE Full text] [doi: 10.2196/mhealth.8518] [Medline: 29496655]
87. Maddison R, Cartledge S, Rogerson M, Goedhart NS, Ragbir Singh T, Neil C, et al. Usefulness of wearable cameras as a tool to enhance chronic disease self-management: scoping review. JMIR Mhealth Uhealth 2019 Jan 3;7(1):e10371 [FREE Full text] [doi: 10.2196/10371] [Medline: 30609985]
88. Goncu-Berk G, Topcuoglu N. A healthcare wearable for chronic pain management. Design of a smart glove for rheumatoid arthritis. Des J 2017 Sep 6;20(sup1):S1978-S1988. [doi: 10.1080/14606925.2017.1352717]
89. Wieringa FP, Broers NJ, Kooman JP, van der Sande FM, van Hoof C. Wearable sensors: can they benefit patients with chronic kidney disease? Expert Rev Med Devices 2017 Jul;14(7):505-519. [doi: 10.1080/17434440.2017.1342533] [Medline: 28612635]

Abbreviations

AGFI: Adjusted Goodness of Fit
AVE: average variance extracted
BCC: Browne-Cudeck Criterion
BIC: Bayesian Information Criterion
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
CMIN/df: Minimum Discrepancy Degrees of Freedom
EFA: exploratory factor analysis
GFI: Goodness of Fit
mHealth: mobile health
NFI: Normed Fit Index Incremental Fit Index
RMSEA: root mean square error of approximation
TLI: Tucker Lewis Index

Edited by G Eysenbach; submitted 09.11.18; peer-reviewed by S Badawy, H Biermann, R Argent; comments to author 10.01.19; revised version received 18.03.19; accepted 06.04.19; published 09.08.19.

Please cite as:
Nelson EC, Verhagen T, Vollenbroek-Hutten M, Noordzij ML
Is Wearable Technology Becoming Part of Us? Developing and Validating a Measurement Scale for Wearable Technology Embodiment
JMIR Mhealth Uhealth 2019;7(8):e12771
URL: https://mhealth.jmir.org/2019/8/e12771/
doi:10.2196/12771
PMID:31400106

Original Paper

Accuracy of 12 Wearable Devices for Estimating Physical Activity Energy Expenditure Using a Metabolic Chamber and the Doubly Labeled Water Method: Validation Study

Haruka Murakami[1*], PhD; Ryoko Kawakami[1,2*], PhD; Satoshi Nakae[3,4], MSc; Yosuke Yamada[5], PhD; Yoshio Nakata[6], PhD; Kazunori Ohkawara[7], PhD; Hiroyuki Sasai[8], PhD; Kazuko Ishikawa-Takata[3], PhD; Shigeho Tanaka[3], PhD; Motohiko Miyachi[1], PhD

[1]Department of Physical Activity Research, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan

[2]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan

[3]Department of Nutrition and Metabolism, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan

[4]Graduate School of Engineering Science, Osaka University, Osaka, Japan

[5]Section of Healthy Longevity Research, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan

[6]Faculty of Health and Sport Sciences, University of Tsukuba, Ibaraki, Japan

[7]Graduate School of Informatics and Engineering, University of Electro-Communication, Tokyo, Japan

[8]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Motohiko Miyachi, PhD
Department of Physical Activity Research
National Institutes of Biomedical Innovation, Health and Nutrition
1-23-1 Toyama
Shinjuku
Tokyo, 162-8636
Japan
Phone: 81 3 3203 8061 ext 4201
Fax: 81 3 3203 1731
Email: miyachi@nibiohn.go.jp

Abstract

Background: Self-monitoring using certain types of pedometers and accelerometers has been reported to be effective for promoting and maintaining physical activity (PA). However, the validity of estimating the level of PA or PA energy expenditure (PAEE) for general consumers using wearable devices has not been sufficiently established.

Objective: We examined the validity of 12 wearable devices for determining PAEE during 1 standardized day in a metabolic chamber and 15 free-living days using the doubly labeled water (DLW) method.

Methods: A total of 19 healthy adults aged 21 to 50 years (9 men and 10 women) participated in this study. They followed a standardized PA protocol in a metabolic chamber for an entire day while simultaneously wearing 12 wearable devices: 5 devices on the waist, 5 on the wrist, and 2 placed in the pocket. In addition, they spent their daily lives wearing 12 wearable devices under free-living conditions while being subjected to the DLW method for 15 days. The PAEE criterion was calculated by subtracting the basal metabolic rate measured by the metabolic chamber and 0.1×total energy expenditure (TEE) from TEE. The TEE was obtained by the metabolic chamber and DLW methods. The PAEE values of wearable devices were also extracted or calculated from each mobile phone app or website. The Dunnett test and Pearson and Spearman correlation coefficients were used to examine the variables estimated by wearable devices.

Results: On the standardized day, the PAEE estimated using the metabolic chamber (PAEEcha) was 528.8±149.4 kcal/day. The PAEEs of all devices except the TANITA AM-160 (513.8±135.0 kcal/day; $P>.05$), SUZUKEN Lifecorder EX (519.3±89.3 kcal/day; $P>.05$), and Panasonic Actimarker (545.9±141.7 kcal/day; $P>.05$) were significantly different from the PAEEcha. None of the devices was correlated with PAEEcha according to both Pearson (r=−.13 to .37) and Spearman (ρ=−.25 to .46) correlation tests. During the 15 free-living days, the PAEE estimated by DLW (PAEEdlw) was 728.0±162.7 kcal/day. PAEE values of all

devices except the Omron Active style Pro (716.2±159.0 kcal/day; *P*>.05) and Omron CaloriScan (707.5±172.7 kcal/day; *P*>.05) were significantly underestimated. Only 2 devices, the Omron Active style Pro (r=.46; *P*=.045) and Panasonic Actimarker (r=.48; *P*=.04), had significant positive correlations with PAEEdlw according to Pearson tests. In addition, 3 devices, the TANITA AM-160 (ρ=.50; *P*=.03), Omron CaloriScan (ρ=.48; *P*=.04), and Omron Active style Pro (ρ=.48; *P*=.04), could be ranked in PAEEdlw.

Conclusions: Most wearable devices do not provide comparable PAEE estimates when using gold standard methods during 1 standardized day or 15 free-living days. Continuous development and evaluations of these wearable devices are needed for better estimations of PAEE.

(JMIR Mhealth Uhealth 2019;7(8):e13938) doi:10.2196/13938

KEYWORDS

physical activity; accelerometry; energy expenditure; indirect calorimetry; doubly labeled water

Introduction

Background

Physical activity (PA) has been reported to reduce the incidence of and mortality because of several noncommunicable diseases, including cardiovascular disease, stroke, and some types of cancer [1-3]. To promote or maintain PA, self-monitoring using pedometers and accelerometers has been considered effective [4]. However, the validity of estimating the amount of PA or PA energy expenditure (PAEE) detected using wearable devices has not been sufficiently established. Previously, we simultaneously examined the validity of total energy expenditure (TEE) estimated by 12 wearable devices during 1 standardized day in a metabolic chamber and 15 free-living days using the doubly labeled water (DLW) method [5]. This study allowed the ranking of daily individual TEE (ρ=.80-.88), but absolute values varied widely among devices and differed significantly from the criterion under free living. Moreover, it is better to estimate accurately not only TEE but also daily PAEE because TEE is mainly determined by the basal metabolic rate (BMR) rather than PA [6].

Several studies have tested the validity of wearable devices for estimating energy expenditure (EE) during some activities [7-14]. However, most have compared EE estimated by wearable devices and standard reference measures estimated by an expired gas analysis during very short structured activities in laboratories [7-9,11-13]. EE measured during such study designs also included resting EE (REE) or BMR, which do not reflect net PAEE. The BMR accounts for a substantial proportion of TEE and is relatively constant from day to day. In contrast, PAEE contributes to TEE to a lesser extent, but it is a fairly variable component that allows the opportunity to increase TEE [6]. Due to the relationship between the amount of PA and health outcomes, accurate estimations of the net PAEE using wearable devices are required, especially under free-living conditions that use wearable devices. Various wearable devices are available for consumer purchase [15], but little is known about their validity.

Objectives

In this study, we evaluated the validity of consumer-based and research-grade wearable devices for estimating PAEE values without the BMR or REE. We developed 2 designs: (1) standardized day for PAEE estimated using a metabolic chamber and (2) 15 free-living days for PAEE estimated using the DLW method.

Methods

Participants

A total of 21 healthy adults aged 21 to 50 years (9 men and 12 women) participated in this study. None of the participants had chronic diseases that could affect their metabolism or daily PA. Their body mass index (BMI) values were within the normal range (18.5-25.0 kg/m^2). Of 21 participants, 2 were excluded from all analyses: 1 because personal information in the JAWBONE UP24 (Jawbone, San Francisco, CA, USA) app during the 15 free-living days experiment had been set incorrectly, and the other because data from the metabolic chamber during the 1 standardized day experiment was incorrect because of instrument failure. Finally, 19 participants (9 men and 10 women) were included in this analysis. All procedures were reviewed and approved by the Ethics Review Board of the National Institute of Health and Nutrition (kenei-4-02). All participants provided written informed consent.

Wearable Devices

The consumer-based wearable devices used in this study were selected based on the following criteria: they were the most popular devices in Japan according to several marketing websites based on their sales ranking (eg, Amazon, Japan website[16] or kakaku website[17] as of December 1, 2014); the app could be displayed in Japanese on a mobile phone or website; and the clock settings of the app or device could be manipulated. We needed to change the clock setting from 9:00 am to 9:00 am the next day to 12:00 am to 12:00 am the next day to obtain the TEE for an entire day when participants used the metabolic chamber. A total of 8 wearable devices, including the Fitbit Flex (Fitbit, San Francisco, CA, USA), JAWBONE UP24, Misfit Shine (Misfit Wearables, Burlingame, CA, USA), EPSON PULSENSE (SEIKO EPSON, Nagano, Japan), Garmin Vivofit (Garmin, Olathe, KS, USA), TANITA AM-160 (TANITA, Tokyo, Japan), Omron CaloriScan HJA-401F (OMRON HEALTHCARE, Kyoto, Japan), and Withings Pulse O2 (Withings, Issy-les-Moulineaux, France), were selected for this study (Table 1). In addition, 4 research-grade wearable devices, namely, Omron Active style Pro (OMRON HEALTHCARE, Kyoto, Japan), Panasonic Actimarker EW4800 (Panasonic, Osaka, Japan), SUZUKEN Lifecorder EX

(SUZUKEN, Aichi, Japan), and ActiGraph GT3X (ActiGraph, Pensacola, FL, USA), were used in this study (Table 1). All devices had a built-in accelerometer. Of 12 wearable devices, 5 (Fitbit Flex, JAWBONE UP24, Misfit Shine, EPSON PULSENSE, and Garmin Vivofit) were placed on the nondominant wrist, 2 (TANITA AM-160 and Omron CaloriScan) were placed in a pocket, and 3 (Withings Pulse O2, Omron Active style Pro, Panasonic Actimarker, SUZUKEN Lifecorder EX, and ActiGraph GT3X) were placed on the waist. The position on the wrist or waist was randomly chosen for each participant, and each participant placed the devices in the same position throughout the experiments.

Table 1. Wearable devices used in the present study, basal metabolic rates extracted from each device, and information about invalid days and non-wearing time in 15 free-living days.

Number	Devices	Placement	Basal metabolic rates[a] (kcal/day), average (SD)	15 free-living days		
				Invalid days[b]	Nonwearing time in valid day	
					min/day, average (SD)	kcal/day[c], average (SD)
1	Fitbit Flex	wrist	1360.4 (195.2)	1	42.4 (18.4)	26.9 (23.4)
2	JAWBONE UP24	wrist	1312.6 (157.1)	0	40.1 (13.0)	25.4 (22.9)
3	Misfit Shine[d]	wrist	1708.0 (245.9)	15	40.4 (13.2)	26.1 (23.1)
4	EPSON PULSENSE[d]	wrist	1616.8 (179.8)	4	42.2 (13.5)	26.4 (22.3)
5	Garmin vivofit[d]	wrist	1630.2 (234.8)	0	39.4 (12.9)	25.2 (23.0)
6	TANITA AM-160[d]	pocket	1410.4 (211.5)	1	42.6 (14.3)	29.3 (29.0)
7	Omron CaloriScan[d]	pocket	1291.7 (186.2)	1	42.6 (14.3)	29.3 (29.0)
8	Withings Pulse O2[d]	waist	1608.9 (228.4)	1	45.5 (13.2)	33.5 (30.8)
9	Omron Active style Pro[d]	waist	1304.5 (188.5)	0	43.1 (13.8)	30.6 (31.3)
10	Panasonic Actimarker	waist	1327.5 (172.4)	0	43.1 (13.8)	30.6 (31.3)
11	SUZUKEN Lifecorder EX	waist	1327.4 (171.9)	0	43.1 (13.8)	30.6 (31.3)
12	ActiGraph GT3X[e]	waist	—[f]	2	42.9 (14.2)	30.5 (31.4)

[a]Basal metabolic rates were extracted from each app.

[b]Total invalid days in 19 participants during 15 days.

[c]The energy expenditure (kcal) in non-wearing time on a valid day was calculated based on time and METs reffered to the Compendium of Physical Activities.

[d]$P<.05$ vs BMR in metabolic chamber (1355.0±234.9 kcal/day).

[e]ActiGraph indicates only PAEE on its application.

[f]Not applicable.

Experimental Design

A total of 2 experiments were conducted to test the validity of the wearable devices: 1 used the metabolic chamber method during 1 standardized day, and the other used the DLW method during 15 free-living days. These 2 methods were used as the standard to determine TEE [18,19]. For the 1-day standardized experiment, participants visited the laboratory 2 hours before the start of the experiment (7:00 am) after an overnight fast of at least 10 hours. Then, height, weight, and body composition were measured. After setting and wearing 12 wearable devices, participants entered the metabolic chamber before 9:00 am and completed 24-hour metabolic chamber measurements (9:00 am to 9:00 am the next day) using a standardized protocol that included various activities common during daily life such as eating 3 meals, watching television (TV), using a computer, cleaning, and walking on a treadmill (Table 2). Each participant's energy intake for meal was calculated by multiplying each BMR by 1.6, which was the PA level (PAL) assumed for a standardized day. The meal was served 3 times per day, and the total energy intake was equally divided into 3 times. The participants were instructed to eat all the meals that were served, and they were not allowed to eat any other foods in the metabolic chamber. However, they were permitted to drink water freely. The average metabolic equivalents (METs) estimated using the compendium of physical activities [20] and previous studies [21-24] for this protocol was 1.37 METs, and the mean PAEE estimated using the estimated METs×hour and participants' weight was 447.0±66.8 kcal/day. Participants wore all the wearable devices during their waking hours without removing them. The 5 devices on the wrist were worn even while sleeping.

Table 2. Timetable for metabolic chamber on a standardized day.

Time	Activity
8:45	Entry in the room
09:00 – 09:30	TV watching
09:30 – 10:30	Breakfast; rice, chicken soup, macaroni salad, and sausage
10:30 – 11:00	Computer work
11:00 – 11:30	Reading a book on a stand
11:30 – 12:00	Folding the laundry
12:00 – 12:30	Cleaning
12:30 – 12:30	Walking (4.0 km/h), including 5 min of rest after walking
13:00 – 13:30	Walking (5.6 km/h), including 5 min of rest after walking
13:30 – 14:00	TV watching
14:00 – 15:00	Lunch; stir-fried vegetables & seafood on rice, cooked beans, egg, and miso soup
15:00 – 15:30	Computer work
15:30 – 16:00	TV watching
16:00 – 16:30	Desk work
16:30 – 17:00	Cleaning
17:00 – 17:30	Walking (4.0 km/h), including 5 min of rest after walking
17:30 – 18:00	Walking (5.6 km/h), including 5 min of rest after walking
18:00 – 18:30	TV watching
18:30 – 19:30	Dinner; rice, hamburg steak, salad, ham, and, vegetable soup
19:30 – 20:00	Computer work
20:00 – 20:30	Reading a book on a stand
20:30 – 21:00	Desk work
21:00 – 21:30	Computer work
21:30 – 22:00	TV watching
22:00 – 22:30	Folding the laundry
22:30 – 23:00	Readying oneself for sleep
23:00 – 07:00	sleep
07:00 – 07:15	lying
07:15 – 08:00	Supine posture
08:00 – 09:00	TV watching
9:10	Exit from the room

During the experiment involving 15 free-living days, participants visited the laboratory in the morning after an overnight fast of at least 10 hours and underwent measurements of height, weight, and body composition. After collecting baseline urine samples, DLW dosing was performed in the laboratory. A premixed dose containing approximately 0.06 g/kg of body weight of $^{2}H_2O$ (99.8 atom%; Cambridge Isotope Laboratories, MA, USA) and 1.4 g/kg of body weight of $H_2{}^{18}O$ (10.0 atom%; Taiyo Nippon Sanso, Tokyo, Japan) was administered orally to each participant. All participants collected their urine samples in air-tight parafilm-wrapped containers at the same time on days 1, 2, 3, 8, 9, 13, 14, and 15 after the baseline day (day 0) during free-living conditions.

Participants wore all the wearable devices when they were awake, but they did not wear them during water-related physical activities, physical activities during which the devices were difficult to wear, or when the battery was charging. Of 12 wearable devices, 5 were worn on the wrist even while sleeping. After 15 free-living days, all urine samples were collected and stored at −30ºC until they were analyzed. Dietary assessments using a brief self-administered diet history questionnaire [25] were conducted to calculate the food quotient (FQ) after 15 days. Logs for time awake, time asleep, nonwearing time, and PA during nonwearing time were completed for 15 days by each participant. PAEE during the nonwearing time was calculated based on the recorded time and METs that were referred to the *Compendium of Physical Activity* [20].

Data Reduction for Each Wearable Device

For the experiment involving 15 free-living days, the days were considered valid when participants wore the wearable devices for more than 10 hours/day [26]. However, we included 1 day when a participant slept for more than 14 hours and, therefore, did not wear the devices for more than 10 hours. The minimum number of valid days was defined as 10 days, and all participants fulfilled this requirement. The mean PAEE of valid days was used for the experiment involving 15 free-living days.

The PAEE for each device ($PAEE_{dev}$) was calculated by subtracting the BMR and 0.1×TEE as diet-induced thermogenesis (DIT) from TEE estimated by each device (TEE_{dev}). The PAL for each device (PAL_{dev}) was calculated by dividing the TEE by the BMR. The BMR for each device (BMR_{dev}) was calculated using the app. The SUZUKEN Lifecorder EX did not show the BMR_{dev} on the app, but the computation method for the BMR using the body surface area and coefficient of the BMR was provided in its instructions; therefore, we calculated the BMR according to those instructions. Because some devices did not show the individual predicted BMR in the device app, including the Fitbit Flex, Misfit Shine, Omron CaloriScan, and Withings Pulse O2, the TEE values of a day when the devices were stationary for the entire day were used as the BMR_{dev}. However, the Omron CaloriScan provided information, indicating that the DIT is included in the TEE when it was stationary for the entire day. Therefore, we did not subtract the DIT when PAEE was calculated using TEE. The ActiGraph GT3X showed only PAEE, not TEE; therefore, we used only the PAEE shown by the ActiGraph GT3X software.

Anthropometry and Body Composition

Height and body weight were measured on both experiment days, and each profile was used for each experiment. BMI (kg/m^2) was calculated, and body composition was determined using a bioelectrical impedance analysis (Inner Scan BC-600; TANITA).

Measurement of Energy Expenditure on a Standardized Day Using the Metabolic Chamber

An open-circuit, indirect metabolic chamber equipped with a bed, desk, chair, TV, toilet, sink, and treadmill was used to measure EE. The temperature and relative humidity in the room were controlled at 25ºC and 55%, respectively. Oxygen and carbon dioxide concentrations of the air supply and exhaust were measured using mass spectrometry (ARCO-1000A-CH; Arco System, Kashiwa, Japan). The flow rates of the exhausts from the chamber were measured using pneumotachography (FLB1; Arco System). Oxygen uptake (VO_2) and carbon dioxide output (VCO_2) were determined based on the concentrations of the inlet and outlet air flows from the chamber and the flow rate of the exhausts from the chamber, respectively. TEE from 9:00 am the first day until 9:00 am the next day was estimated from VO_2 and VCO_2 using Weir equation (TEE_{cha}). The BMR was measured in the supine position for 45 min during the morning (BMR_{cha}). The PAEE during 1 standardized day ($PAEE_{cha}$) was calculated by subtracting the BMR_{cha} and 0.1×TEE_{cha} from TEE_{cha}. The PAL during 1 standardized day (PAL_{cha}) was calculated by dividing the TEE_{cha} by the BMR_{cha}.

Measurement of Energy Expenditure During 15 Free-Living Days Using the Doubly Labeled Water Method

Gas samples for the isotope ratio mass spectrometer (IRMS) were prepared by maintaining the equilibration of the urine sample with gas. CO_2 was used to equilibrate ^{18}O, and H_2 was used to equilibrate ^{2}H. The platinum (Pt) catalyst was used for equilibration of ^{2}H. Gas samples for CO_2 and H_2 measurements were analyzed using IRMS (Sercon 20-20; Sercon Ltd, Crewe, UK). Each sample and its corresponding reference were analyzed in triplicate. The ^{2}H and ^{18}O zero-time intercepts and elimination rates (kd and ko) were calculated using the least-squares linear regression method on the natural logarithm of the isotope concentration as a function of the elapsed time from dose administration. Zero-time intercepts were used to determine the isotope pool sizes. A quality check was conducted according to the International Atomic Energy Agency book [27]. The memory effects of the IRMS were eliminated and checked using additional samples when the expected isotope ratio difference was high (eg, days 2-8), and the potential drift of the IRMS was corrected mathematically using standardized working criteria and checked for accuracy and precision using another working criterion at regular intervals in a series of measurements and between different measurement days. The samples obtained from 1 participant were analyzed in 1 series of measurements in 1 day to minimize the effects of day-to-day variation. The dilution space ratio of ^{2}H (Nd) and ^{18}O (No) of all 21 participants was 1.036±0.010 (range 1.021-1.056), which was an acceptable value according to a previous review of a large database [28]. Therefore, total body water (TBW) was calculated from the mean value or the isotope pool size of ^{2}H divided by 1.041 and that of ^{18}O divided by 1.007. The carbon dioxide production rate (rCO_2) was calculated as follows: rCO_2=0.4554×TBW×(1.007 ko−1.041 kd), for which we assumed that isotope fractionation applies only to breath water using equation A6 by Schoeller et al [29] with the revised dilution space constant provided by Racette et al [30]. The TEE (TEE_{dlw}) was calculated using a modified Weir formula based on the rCO_2 and FQ [31] as follows: TEE (kcal/day)=1.1 rCO_2+3.9 rCO_2/FQ.

The PAEE during free-living days ($PAEE_{dlw}$) was calculated by subtracting the BMR_{cha} and 0.1×TEE_{dlw} from TEE_{dlw}. The PAL during free-living days (PAL_{dlw}) was calculated by dividing TEE_{dlw} by BMR_{cha}.

Statistical Analysis

Data were expressed as mean (standard deviation). The Dunnett test, for which standard criteria were set as references, was used for comparing variables estimated by wearable devices during the use of the metabolic chamber method and the DLW method. The mean absolute percent errors (MAPEs) relative to the PAEE values estimated using standard methods were calculated to

provide an indicator of the overall measurement error. The Pearson and Spearman correlation coefficients were used to examine the relationship between standard criteria and variables estimated by wearable devices. Modified Bland-Altman plots [32] were used to test proportional biases between standard methods and devices, and the correlation coefficient of the standard criteria and the differences between the standard criteria and each device were examined for significance. During all analyses, $P<.05$ was considered statistically significant. All statistical analyses were performed with SPSS version 20.0 for Windows (IBM SPSS Japan Inc, Tokyo, Japan).

Results

Descriptive Results

Participants were aged 32.3±9.6 years. Their BMI and percentage body fat ranged from 18.5 to 24.8 kg/m^2 and from 14.8% to 32.2%, respectively. Although there was no invalid day during the standardized 1-day experiment, 25 invalid days were identified during the 15-day free-living experiment (Table 1), which corresponded to 8.8% of all experiment days (19 participants×15 days). Invalid days often occurred with the Misfit Shine because a few of these devices became loose without the knowledge of the participant and with EPSON PULSENSE because the battery quickly died. The average nonwearing time except for sleeping for each device ranged from 39.4±12.9 to 45.5±13.2 min/day, which corresponded to 25.2±23.0 to 33.5±30.8 kcal/day (Table 1). The most frequent activities during nonwearing time were bathing and showering (289 cases/19 participants×15days). There were 62 other activities including TV watching, deskwork, dressing, and exercise. The corresponding time and intensity for these activities were 5 to 450 min and 1.3 to 6.3 METs, respectively.

The BMR_{cha} was 1355.0±234.9 kcal/day. Several devices showed higher BMR_{dev} than BMR_{cha} ($P<.05$), including the Misfit Shine, EPSON PULSENSE, Garmin Vivofit, TANITA AM-160, and Withings Pulse O2 (Table 1).

Metabolic Chamber Study

During the standardized day, the $PAEE_{cha}$ was 528.8±149.4 kcal/day. All devices except the TANITA AM-160 (513.8±135.0 kcal/day; $P>.05$), SUZUKEN Lifecorder EX (519.3±89.3 kcal/day; $P>.05$), and Panasonic Actimarker (545.9±141.7 kcal/day; $P>.05$) showed significant differences in $PAEE_{dev}$ compared with $PAEE_{cha}$ (Figure 1). Moreover, 6 devices significantly underestimated values, whereas 3 devices significantly overestimated values. The Withings Pulse O2 (24.4±56.7 kcal/day) and Garmin Vivofit (29.5±34.0 kcal/day) showed large gaps in $PAEE_{cha}$, with MAPEs of 93.7±13.9% and 92.8±13.1%, respectively. Moreover, all devices showed systematic errors with high negative correlation coefficients on the Bland-Altman plots (Figure 1).

No devices showed a significant correlation with $PAEE_{cha}$ according to both Pearson and Spearman correlation tests (Figure 2). Regarding PAL, all devices except the TANITA AM-160 (1.51±0.07; $P>.05$), Panasonic Actimarker (1.56±0.08; $P>.05$), and SUZUKEN Lifecorder (1.55±0.04; $P>.05$) showed significant differences in PAL compared with PAL_{cha} (1.56±0.17; Tables 3 and 4). No devices showed a significant correlation with PAL_{cha} according to both Pearson and Spearman correlation tests (Tables 3 and 4). PAEE/body weight also showed similar results for PAL (Tables 3 and 4). Moreover, similar results were obtained in partial correlation test using body weight as a control variable.

Figure 1. Differences between PAEEcha (physical activity energy expenditure) or PAEEdlw and each PAEEdev on a standardized day and 15 free-living days. Note: (a) The correlations between PAEEcha or PAEEdlw in x axis and delta as PAEEdev - PAEEcha or PAEEdev - PAEEdlw in y axis were showed. (b) $P<.05$ vs PAEEcha or PAEEdlw. (c) $P<.05$ in correlations. PAEE: physical activity energy expenditure; MAPE: mean absolute percentage error; avg: average; SD: standard deviation.

A standardized day
(Measured by metabolic chamber: 528.8 ± 149.4 kcal/day)

	Difference in PAEE between each device and metabolic chamber (kcal/day) -1200 -800 -400 0 400	Estimated PAEE by each device (kcal/day) avg ± SD	MAPE (%) avg ± SD	Bland-Altman r [a]
Withings Pulse O2		24.4 ± 56.7 [b]	93.7 ± 13.9	-0.94 [c]
Garmin vivofit		29.5 ± 34.0 [b]	92.8 ± 13.1	-0.98 [c]
Misfit Shine		291.3 ± 89.0 [b]	50.5 ± 22.9	-0.86 [c]
EPSON PULSENSE		299.2 ± 144.7 [b]	52.8 ± 34.1	-0.74 [c]
JAWBONE UP24		321.7 ± 63.9 [b]	40.2 ± 16.0	-0.91 [c]
ActiGraph GT3X		423.9 ± 145.9 [b]	32.0 ± 20.6	-0.64 [c]
TANITA AM-160		513.8 ± 135.0	29.9 ± 27.6	-0.62 [c]
SUZUKEN Lifecorder EX		519.3 ± 89.3	26.5 ± 39.3	-0.82 [c]
Panasonic Actimarker		545.9 ± 141.7	31.7 ± 39.5	-0.65 [c]
Fitbit Flex		637.0 ± 115.9 [b]	39.7 ± 54.3	-0.74 [c]
Omron Active style Pro		737.0 ± 154.9 [b]	56.3 ± 55.3	-0.59 [c]
Omron CaloriScan		776.1 ± 138.2 [b]	62.9 ± 63.9	-0.62 [c]

15 free-living days
(Measured by DLW method: 728.0 ± 162.7 kcal/day)

	Difference in PAEE between each device and DLW method (kcal/day) -1200 -800 -400 0 400	Estimated PAEE by each device (kcal/day) avg ± SD	MAPE (%) avg ± SD	Bland-Altman r [a]
Withings Pulse O2		8.0 ± 63.9 [b]	98.4 ± 9.6	-0.94 [c]
Garmin vivofit		0.2 ± 42.9 [b]	100.2 ± 6.3	-0.97 [c]
Misfit Shine		167.7 ± 102.0 [b]	76.7 ± 13.9	-0.80 [c]
EPSON PULSENSE		270.8 ± 147.6 [b]	61.3 ± 21.3	-0.73 [c]
JAWBONE UP24		239.2 ± 91.8 [b]	65.9 ± 13.8	-0.86 [c]
ActiGraph GT3X		306.5 ± 172.5 [b]	57.9 ± 23.4	-0.50 [c]
TANITA AM-160		474.5 ± 179.2 [b]	36.3 ± 19.6	-0.47 [c]
SUZUKEN Lifecorder EX		503.6 ± 122.3 [b]	30.9 ± 16.5	-0.73 [c]
Panasonic Actimarker		535.4 ± 149.7 [b]	29.2 ± 15.3	-0.57 [c]
Fitbit Flex		567.8 ± 163.6 [b]	27.8 ± 15.9	-0.58 [c]
Omron Active style Pro		716.2 ± 159.0	19.4 ± 14.9	-0.53 [c]
Omron CaloriScan		707.5 ± 172.7	19.4 ± 15.2	-0.50 [c]

Figure 2. Correlation between PAEEcha (physical activity energy expenditure) and PAEEdev during 1 standardized day. Scattered plots between PAEEcha (x-axis) and PAEEdev (y-axis) during 1 standardized day. There was no significant correlation according to Pearson and Spearman tests. n.s.: nonsignificant.

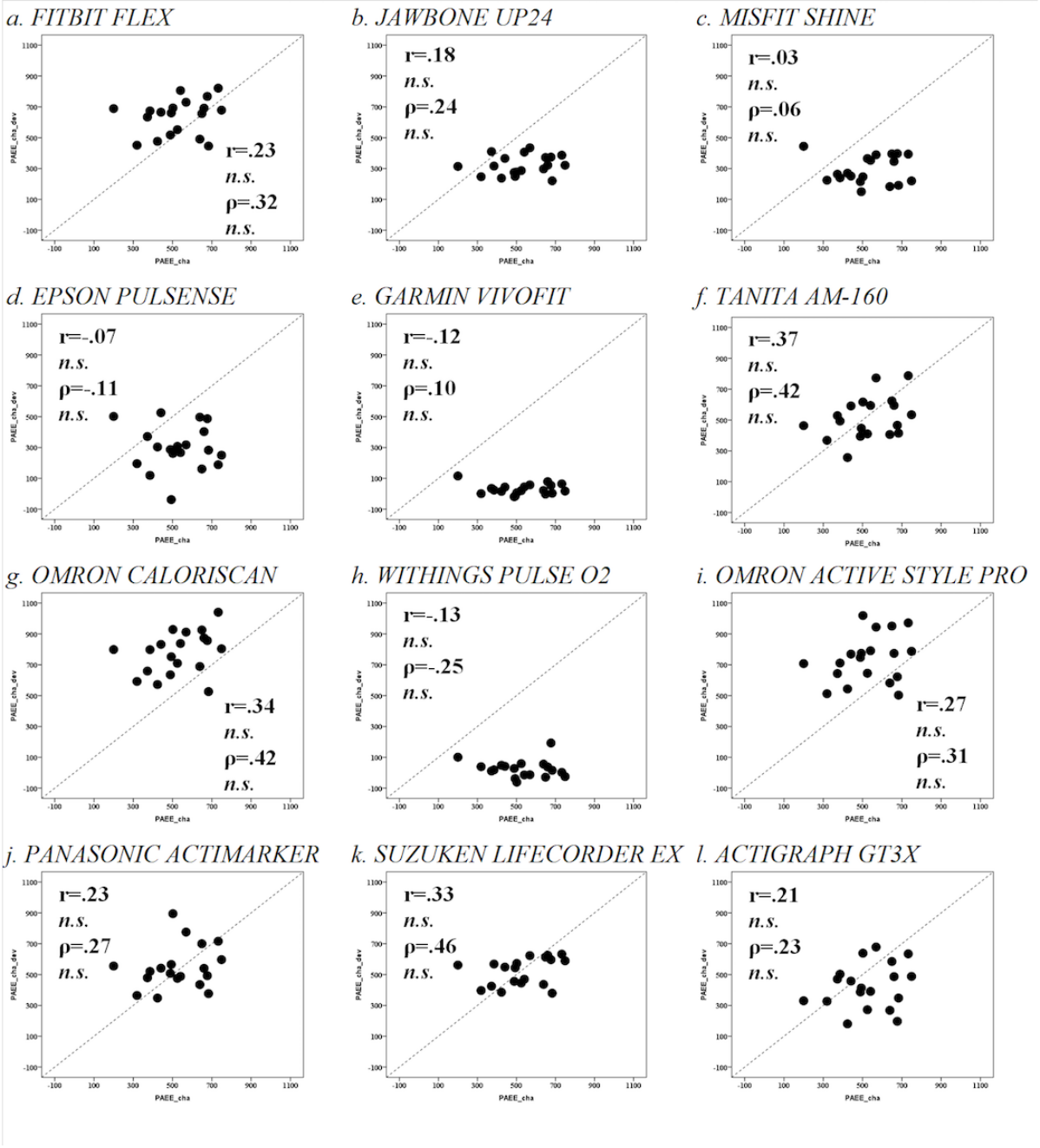


Table 3. The comparison between PAL^{a}_{cha} and PAL_{dev}, $PAEE^{b}_{cha}$/wt (physical activity energy expenditure) and $PAEE_{dev}$/wt.

Devices	A standardized day			
	PAL_{cha}: 1.56 ± 0.17		$PAEE_{cha}$/wt: 9.2 ± 2.4 kcal/kg/day	
	Value, average (SD)	Pearson correlation	Value, average (SD)	Pearson correlation
Withings Pulse O2	1.13 (0.04)[c]	0.08	0.5 (1.0)[c]	0.02
Garmin vivofit	1.13 (0.02)[c]	-0.27	0.5 (0.6)[c]	-0.37
Misfit Shine	1.30 (0.06)[c]	-0.19	5.1 (1.5)[c]	-0.25
EPSON PULSENSE	1.32 (0.11)[c]	0.08	5.4 (2.8)[c]	0.07
JAWBONE UP24	1.39 (0.06)[c]	-0.19	5.6 (1.0)[c]	-0.14
ActiGraph GT3X	1.47 (0.09)[c]	-0.29	7.2 (1.8)[c]	-0.26
TANITA AM-160	1.51 (0.07)	-0.13	8.8 (1.4)	-0.10
SUZUKEN Lifecorder EX	1.55 (0.04)	-0.30	9.0 (0.8)	-0.27
Panasonic Actimarker	1.56 (0.08)	-0.34	9.4 (1.5)	-0.26
Fitbit Flex	1.63 (0.06)[c]	-0.38	11.0 (1.2)[c]	-0.39
Omron Active style Pro	1.74 (0.07)[c]	-0.44	12.7 (1.4)[c]	-0.30
Omron CaloriScan	1.78 (0.07)[c]	-0.29	13.4 (1.2)[c]	-0.24

[a]PAL: physical activity level.

[b]PAEE: physical activity energy expenditure.

[c]$P<.05$ vs PAL_{cha} or $PAEE_{cha}$/wt.

Table 4. The comparison between PAL^{a}_{dlw} and PAL_{dev}, $PAEE^{b}_{dlw}$/wt and $PAEE_{dev}$/wt.

Devices	15 free-living days			
	PAL_{dlw}: 1.73 ± 0.21		$PAEE_{dlw}$/wt: 12.8 ± 3.1 kcal/kg/day	
	Value, average (SD)	Pearson correlation	Value, average (SD)	Pearson correlation
Withings Pulse O2	1.12 (0.04)[c]	-0.24	0.2 (1.1)[c]	-0.14
Garmin vivofit	1.11 (0.03)[c]	-0.08	0.0 (0.7)[c]	-0.07
Misfit Shine	1.22 (0.06)[c]	-0.02	2.9 (1.6)[c]	0.11
EPSON PULSENSE	1.30 (0.10)[c]	-0.31	4.7 (2.5)[c]	-0.05
JAWBONE UP24	1.31 (0.07)[c]	-0.30	4.1 (1.4)[c]	-0.10
ActiGraph GT3X	1.37 (0.13)[c]	0.11	5.2 (2.7)[c]	0.25
TANITA AM-160	1.48 (0.11)[c]	-0.01	8.1 (2.4)[c]	0.10
SUZUKEN Lifecorder EX	1.53 (0.08)[c]	-0.12	8.7 (1.5)[c]	0.07
Panasonic Actimarker	1.56 (0.10)[c]	0.26	9.3 (2.1)[c]	0.39
Fitbit Flex	1.57 (0.11)[c]	-0.08	9.8 (2.3)[c]	0.13
Omron Active style Pro	1.72 (0.10)	0.14	12.4 (1.9)	0.35
Omron CaloriScan	1.71 (0.09)	-0.07	12.2 (1.7)	0.11

[a]PAL: physical activity level.

[b]PAEE: physical activity energy expenditure.

[c]$P<.05$ vs PAL_{dlw} or $PAEE_{dlw}$/wt.

Doubly Labeled Water Study

During the 15 free-living days experiment, the $PAEE_{dlw}$ was 728.0±162.7 kcal/day. The PAEEs from all devices except the Omron Active style Pro (716.2±159.0 kcal/day; *P*>.05) and Omron CaloriScan (707.5±172.7 kcal/day; *P*>.05) were significantly underestimated (Figure 1). Only 2 devices, the Omron Active style Pro (r=0.46; *P*=.045) and Panasonic Actimarker (r=0.48; *P*=.04), showed significant positive Pearson correlations. In addition, 3 devices, the TANITA AM-160 (ρ=.50; *P*=.03), Omron CaloriScan (ρ=.48; *P*=.04), and Omron Active style Pro (ρ=.48; *P*=.04), can be ranked in $PAEE_{dev}$ (Figure 3). On the other hand, systematic biases indicated by Bland-Altman plots were observed for all devices with negative coefficients (Figure 1). Regarding PAL, all devices except the Omron Active style Pro (1.72±0.10; *P*>.05) and Omron CaloriScan (1.71±0.09; *P*>.05) showed significant differences in PAL_{dev} compared with PAL_{dlw} (1.73±0.21; Tables 3 and 4). No devices showed a significant correlation with PAL_{dlw} according to both Pearson and Spearman tests (Tables 3 and 4). PAEE/body weight also showed results similar to those of PAL (Tables 3 and 4). Moreover, similar results with partial correlation were obtained using body weight as a control variable.

Figure 3. Correlation between PAEEdlw (physical activity energy expenditure) and PAEEdev during 15 free-living days. Scatter plots for PAEEdlw (x-axis) and PAEEdev (y-axis) during 15 free-living days. Upper and lower values for r and ρ resulting from Pearson and Spearman correlation tests, respectively, are shown. n.s.: nonsignificant.

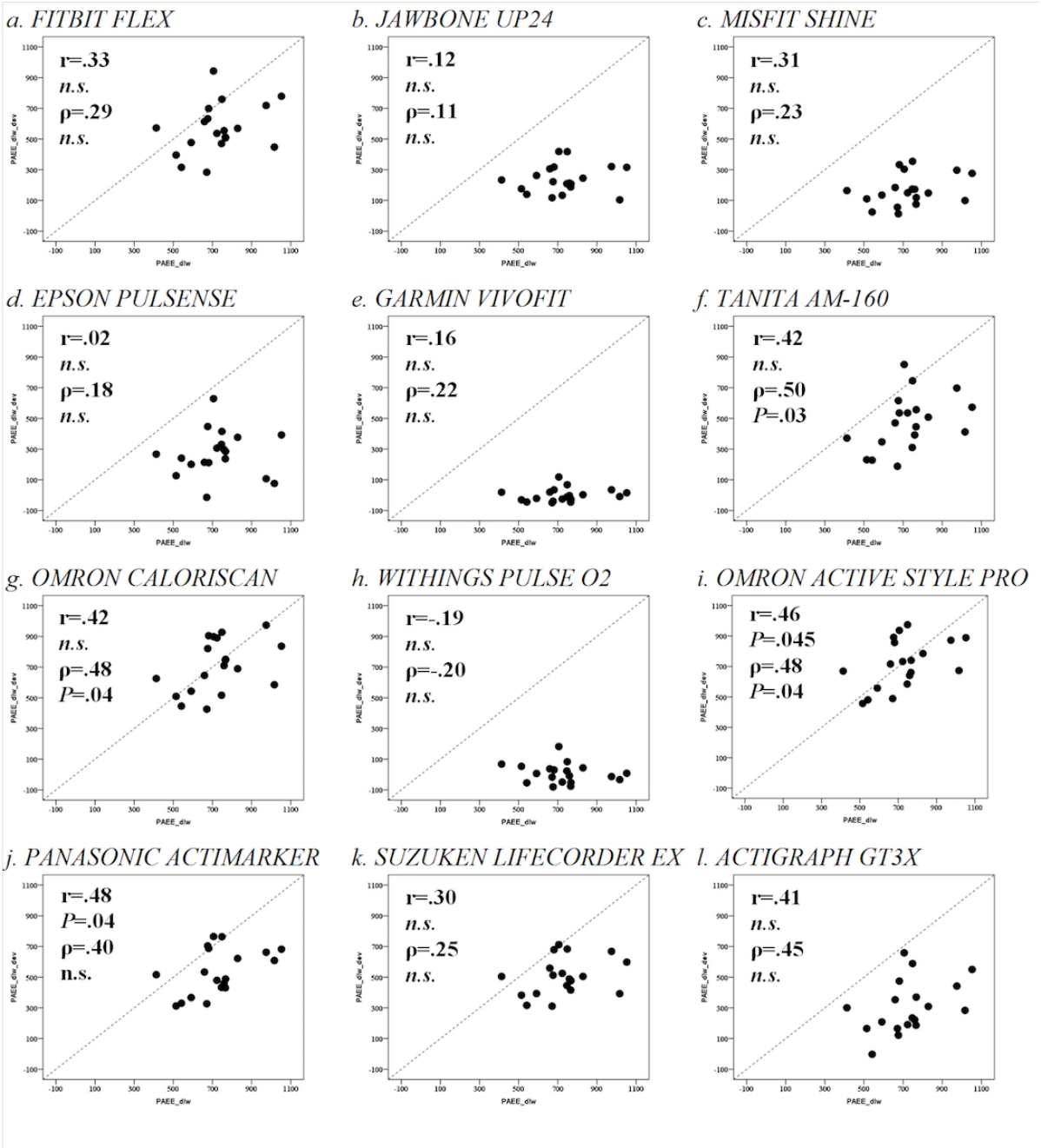


Table 5. The comparison between PAEE[a]$_{cha}$ and the unique PAEE parameters extracted from each consumer-based wearable device.

Number	Devices	Item	Standardized day (PAEE$_{cha}$: 528.8 ± 149.4 kcal/day）	
			Value, average (SD)	Pearson correlation
1	Fitbit Flex	N/A[b]	—[c]	—
2	JAWBONE UP24	active energy expenditure	503.3 (77.9)	0.30
3	Misfit Shine	N/A	—	—
4	EPSON PULSENSE	active energy expenditure	416.2 (173.1)	0.14
5	Garmin vivofit	exercise energy expenditure (web)	212.4 (44.4)	0.32
6	TANITA AM-160	active energy expenditure	726.8 (168.7)	0.39
7	Omron CaloriScan	active energy expenditure (web)	774.3 (137.1)	0.40
8	Withings Pulse O2	activity energy expenditure	318.4 (54.8)	0.23

[a]PAEE: physical activity energy expenditure.

[b]Fitbit and Misfit Shine were not available for unique PAEE parameters in their app and website.

[c]Not applicable.

Unique PAEE by Consumer-Based Devices

On the standardized day, we also compared the PAEE$_{cha}$ with the unique PAEE parameters obtained by 6 of the 8 consumer-based devices (Table 5). The absolute values from each device were not compared with PAEE$_{cha}$ because we could not find any information about these parameters and could not define the value as PAEE. None of the parameters showed a significant correlation with PAEE$_{cha}$.

Discussion

Principal Findings

We examined the validity of 12 consumer-based and research-grade wearable devices for estimating PAEE using a metabolic chamber and the DLW method as standard methods. On the standardized day, most of the wearable devices showed significant differences in PAEE when compared with PAEE$_{cha}$ (MAPE 26.5%-93.7%). Moreover, all wearable devices except the Omron CaloriScan and Omron Active style Pro significantly underestimated values during 15 free-living days (MAPE 19.4%-100.2%). These results were similar, even for PAL. The number of wearable devices with significant differences in PAEE compared with the standard criteria in this study was greater than the number of devices with significant differences in TEE in our previous study using same 12 devices; we found that only 2 devices during the standardized day and 4 devices during 15 free-living days showed significant differences in TEE compared with the standard criteria [5]. These results showed that wearable devices had lesser accuracy when estimating PAEE than TEE, which included the BMR.

Comparison With Previous Studies

Several studies have evaluated the validity of EE estimated by wearable devices during some activities [7-14]. Most of these studies were conducted during very short structured activities in laboratories. For the most studied device (Fitbit), there were many inconsistent results such as overestimated EE [4,33,34], underestimated EE [11,12], and comparable EE [8]. It has also been reported that the EE estimations based on the Fitbit were largely different depending on the activity types performed during those studies [8,12]. These discrepancies may have been dependent on the differences in the standard criteria, EE assessment method, and selected activities. In this study, the PAEE estimated by the Fitbit Flex was somewhat comparable with standard PAEEs during a standardized day and during 15 free-living days in consumer-based wearable devices, which was consistent with the results of the Fitbit Zip [11]. Furthermore, in this study, the JAWBONE UP24 underestimated PAEEs during both experiments, which was consistent with the results of previous studies [7,11]. However, the Misfit Shine and Garmin Vivofit underestimated PAEE during this study but overestimated PAEE during previous studies [7,9]. Attention is necessary when directly comparing the present results of this study with the previous results because what was used to evaluate PAEE was slightly different. We evaluated TEE−BMR−TEE×0.1 as PAEE (ie, net EE with PA); however, most previous studies that evaluated EE included the BMR or REE during experimental activities as PAEE. We also compared the unique indices of PAEE provided by several devices as PAEE$_{cha}$ (Table 5). These were indicated on the app as active EE or exercise EE. However, no parameters were significantly correlated with PAEE$_{cha}$. Most evidence that demonstrated the relationship between PA and risk reduction of disease based on epidemiological studies were described as the amount of PA but not as the TEE. Therefore, it is important to accurately assess daily PAEE in terms of preventive medicine and public health.

Underestimation Under Free Living

In a comparison of the results of the standardized day and those of 15 free-living days, all wearable devices except the Omron CaloriScan and Omron Active style Pro underestimated PAEE for 15 free-living days, whereas 6 devices underestimated PAEE, and 3 devices overestimated PAEE on the standardized day. Because TEE measurements using the metabolic chamber have been reported as not significantly different from TEE measured by DLW methods on the same days [35], our results were not caused by different criteria for the TEE assessment.

Underestimation by most devices during 15 free-living days may have been partly caused by the nonwearing time. We calculated the average PAEE during the nonwearing time ($PAEE_{nonwear}$) by multiplying the nonwearing time by MET corresponding to the PA performed [20] based on the daily log recorded by participants. Even if $PAEE_{nonwear}$ derived from each wearable device were added to each $PAEE_{dev}$, PAEE would have remained underestimated. This means that many types of PA are underestimated during free-living days.

It has been reported that cycling and washing laundry are underestimated by wearable devices [8,12]. Moreover, standing that does not produce acceleration may be classified as sedentary behavior [36]. These types of PA during free-living days may have caused underestimation of PAEE in this study. Although early consumer-based wearable devices for estimating PA relied on movement sensors alone (eg, accelerometers), more recently developed wearable devices integrate several physiological or geographical outputs, including heart rate, skin temperature, galvanic skin response, and a global positioning system [37]. PAEE that cannot be captured by an accelerometer may be accurately estimated using these multisensor wearable devices in the future. Another reason for the underestimation of PAEE during free-living days could have been transition in postures (eg, sit-to-stand), transition in directions, and acceleration and deceleration during movements. Recent studies have suggested that significant additional EE is associated with changing directions and/or changing postures [38-41], and those transitions are often observed during free-living days [42,43]. However, those elements were not usually considered to establish and validate PA monitors. To assess actual PAEE during daily life, it is necessary to continuously evaluate the validity of these sensors for estimating PAEE.

Perspectives

Wearable devices can be powerful tools that provide not only individual information but also large-scale population data on a global scale. Most wearable devices can connect to the internet through an app on a user's mobile phone and collect data. Using 68 million days of step count data from 717,517 users of the Argus Smartphone app, Althoff et al [44] showed that inequality in PA within a country was associated with the prevalence of obesity in the population. Moreover, multiple aspects of health behavior need to be monitored simultaneously and continually because our health outcomes resulted from various health behaviors that included not only PA but also daily diet, smoking, and sleep [45]. Under such circumstances, it is important to be able to properly evaluate the multilateral health behavior and physiological parameters globally. However, some problems have been highlighted by the continuous wearing of such a device. One-third of owners of a consumer-based wearable device stopped using it within 6 months [15]. Therefore, it is necessary to enhance continuity and strive to maintain and improve health outcomes through various other approaches.

Limitations

There were some limitations to this study. First, the sample size was small and restricted to normal-weight individuals; therefore, results cannot be generalized to obese or lean people. Comprehensive validation extending to other populations with various PALs is required. Because it was expected that some types of PA were underestimated and some were overestimated by wearable devices, different PA situations may lead to different results. Second, we could not examine the validity of all wearable devices for all types of activity during a standardized day. Different settings using different intensities and other types of activities may lead to different results. We also need to confirm the results in different settings or examine the validity of each activity performed during a standardized day to reveal the causes of the underestimation and overestimation. Third, BMR values estimated by several wearable devices were obtained as whole-day values with stable situation. This was not supposed by the manufacturers; therefore, we might have used BMR incorrectly for several devices, which might have led to erroneous estimations of PAEE because it was calculated by subtracting the BMR from the TEE. Therefore, comparisons of absolute values of PAEE for these devices in this study must be interpreted with caution.

Conclusions

In conclusion, most wearable devices showed PAEEs that were significantly different from those estimated using gold standard methods during a standardized day and 15 free-living days. It is possible that the PAEE of some PA is underestimated during free-living situations by wearable devices. The development of wearable devices that can accurately estimate PAEE will lead people to use them as motivational tools. Moreover, this will allow researchers to precisely understand PA in an observational study or intervention study, thereby leading to public health recommendations based on scientific evidence.

Acknowledgments

This research was supported by the Practical Research Project for Lifestyle-related Diseases including Cardiovascular Diseases and Diabetes Mellitus from the Japan Agency for Medical Research and Development (AMED). AMED 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.

Authors' Contributions

HM, RK, and MM had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors contributed to the study concept and design and drafted the manuscript. HM, RK, and SN collected data or samples and analyzed.

Conflicts of Interest

ST reported receiving research funding from Omron Health Care Inc. No other disclosures were reported.

References

1. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the global burden of disease study 2013. Br Med J 2016 Aug 9;354:i3857 [FREE Full text] [doi: 10.1136/bmj.i3857] [Medline: 27510511]
2. Liu L, Shi Y, Li T, Qin Q, Yin J, Pang S, et al. Leisure time physical activity and cancer risk: evaluation of the WHO's recommendation based on 126 high-quality epidemiological studies. Br J Sports Med 2016 Mar;50(6):372-378. [doi: 10.1136/bjsports-2015-094728] [Medline: 26500336]
3. Samitz G, Egger M, Zwahlen M. Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies. Int J Epidemiol 2011 Oct;40(5):1382-1400. [doi: 10.1093/ije/dyr112] [Medline: 22039197]
4. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
5. 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 1;176(5):702-703. [doi: 10.1001/jamainternmed.2016.0152] [Medline: 26999758]
6. Luís GJ, María MJ, Barbany M, Contreras J, Amigó P, Salas-Salvadó J. Physical activity, energy balance and obesity. Public Health Nutr 2007 Oct;10(10A):1194-1199. [doi: 10.1017/S1368980007000705] [Medline: 17903330]
7. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
8. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
9. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
10. 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;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
11. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014 Sep;46(9):1840-1848. [doi: 10.1249/MSS.0000000000000287] [Medline: 24777201]
12. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Kozey KS, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. J Phys Act Health 2015 Feb;12(2):149-154. [doi: 10.1123/jpah.2012-0495] [Medline: 24770438]
13. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):pii: E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
14. Tucker WJ, Bhammar DM, Sawyer BJ, Buman MP, Gaesser GA. Validity and reliability of Nike + Fuelband for estimating physical activity energy expenditure. BMC Sports Sci Med Rehabil 2015;7:14 [FREE Full text] [doi: 10.1186/s13102-015-0008-7] [Medline: 26751385]
15. Wright SP, Hall BT, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Dec 1;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
16. Amazon. Selling Rankings URL: https://www.amazon.co.jp/gp/bestsellers/sports/2201382051/ref=zg_bs_nav_sg_2_14315501 [accessed 2014-12-01]
17. [Price.com]. [Activity meter mail order price comparison] URL: https://kakaku.com/keitai/activity-meter/ [accessed 2014-12-01]
18. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
19. Westerterp KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. Front Physiol 2013;4:90 [FREE Full text] [doi: 10.3389/fphys.2013.00090] [Medline: 23637685]
20. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett Jr DR, Tudor-Locke C, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]

21. Ohkawara K, Oshima Y, Hikihara Y, Ishikawa-Takata K, Tabata I, Tanaka S. Real-time estimation of daily physical activity intensity by a triaxial accelerometer and a gravity-removal classification algorithm. Br J Nutr 2011 Jun;105(11):1681-1691. [doi: 10.1017/S0007114510005441] [Medline: 21262061]
22. Taguri E, Tanaka S, Ohkawara K, Ishikawa-Takata K, Hikihara Y, Miyake R, et al. Validity of physical activity indices for adjusting energy expenditure for body size: do the indices depend on body size? J Physiol Anthropol 2010;29(3):109-117 [FREE Full text] [doi: 10.2114/jpa2.29.109] [Medline: 20558969]
23. Midorikawa T, Tanaka S, Kaneko K, Koizumi K, Ishikawa-Takata K, Futami J, et al. Evaluation of low-intensity physical activity by triaxial accelerometry. Obesity (Silver Spring) 2007;15(12):3031-3038 [FREE Full text] [doi: 10.1038/oby.2007.361] [Medline: 18198312]
24. Newton Jr RL, Han H, Zderic T, Hamilton MT. The energy expenditure of sedentary behavior: a whole room calorimeter study. PLoS One 2013;8(5):e63171 [FREE Full text] [doi: 10.1371/journal.pone.0063171] [Medline: 23658805]
25. Kobayashi S, Honda S, Murakami K, Sasaki S, Okubo H, Hirota N, et al. Both comprehensive and brief self-administered diet history questionnaires satisfactorily rank nutrient intakes in Japanese adults. J Epidemiol 2012;22(2):151-159 [FREE Full text] [doi: 10.2188/jea.je20110075] [Medline: 22343326]
26. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
27. Assessment Of Body Composition And Total Energy Expenditure In Humans Using Stable Isotope Techniques. Austria: International Atomic Energy Agency; 2009.
28. Sagayama H, Yamada Y, Racine NM, Shriver TC, Schoeller DA, DLW Study Group. Dilution space ratio of 2H and 18O of doubly labeled water method in humans. J Appl Physiol (1985) 2016 Jun 1;120(11):1349-1354 [FREE Full text] [doi: 10.1152/japplphysiol.01037.2015] [Medline: 26989221]
29. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jéquier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. Am J Physiol 1986 May;250(5 Pt 2):R823-R830. [doi: 10.1152/ajpregu.1986.250.5.R823] [Medline: 3085521]
30. Racette SB, Schoeller DA, Luke AH, Shay K, Hnilicka J, Kushner RF. Relative dilution spaces of 2H- and 18O-labeled water in humans. Am J Physiol 1994 Oct;267(4 Pt 1):E585-E590. [doi: 10.1152/ajpendo.1994.267.4.E585] [Medline: 7943308]
31. Black AE, Prentice AM, Coward WA. Use of food quotients to predict respiratory quotients for the doubly-labelled water method of measuring energy expenditure. Hum Nutr Clin Nutr 1986 Sep;40(5):381-391. [Medline: 3771290]
32. Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. Stat Med 2008 Feb 28;27(5):778-780. [doi: 10.1002/sim.3086] [Medline: 17907247]
33. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
34. Murakami H, Tripette J, Kawakami R, Miyachi M. 'Add 10 min for your health': the new Japanese recommendation for physical activity based on dose-response analysis. J Am Coll Cardiol 2015 Mar 24;65(11):1153-1154 [FREE Full text] [doi: 10.1016/j.jacc.2014.10.080] [Medline: 25790890]
35. Speakman JR. The history and theory of the doubly labeled water technique. Am J Clin Nutr 1998 Oct;68(4):932S-938S. [doi: 10.1093/ajcn/68.4.932S] [Medline: 9771875]
36. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
37. 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]
38. Hatamoto Y, Yamada Y, Higaki Y, Tanaka H. A novel approach for measuring energy expenditure of a single sit-to-stand movement. Eur J Appl Physiol 2016 May;116(5):997-1004. [doi: 10.1007/s00421-016-3355-5] [Medline: 27017496]
39. Hatamoto Y, Yamada Y, Sagayama H, Higaki Y, Kiyonaga A, Tanaka H. The relationship between running velocity and the energy cost of turning during running. PLoS One 2014;9(1):e81850 [FREE Full text] [doi: 10.1371/journal.pone.0081850] [Medline: 24497913]
40. McNarry MA, Wilson RP, Holton MD, Griffiths IW, Mackintosh KA. Investigating the relationship between energy expenditure, walking speed and angle of turning in humans. PLoS One 2017;12(8):e0182333 [FREE Full text] [doi: 10.1371/journal.pone.0182333] [Medline: 28796796]
41. Miles-Chan JL, Fares EJ, Berkachy R, Jacquet P, Isacco L, Schutz Y, et al. Standing economy: does the heterogeneity in the energy cost of posture maintenance reside in differential patterns of spontaneous weight-shifting? Eur J Appl Physiol 2017 Apr;117(4):795-807. [doi: 10.1007/s00421-017-3563-7] [Medline: 28260201]

42. Smith L, Hamer M, Ucci M, Marmot A, Gardner B, Sawyer A, et al. Weekday and weekend patterns of objectively measured sitting, standing, and stepping in a sample of office-based workers: the active buildings study. BMC Public Health 2015 Jan 17;15:9 [FREE Full text] [doi: 10.1186/s12889-014-1338-1] [Medline: 25595219]
43. Grant PM, Dall PM, Kerr A. Daily and hourly frequency of the sit to stand movement in older adults: a comparison of day hospital, rehabilitation ward and community living groups. Aging Clin Exp Res 2011;23(5-6):437-444. [doi: 10.1007/BF03325239] [Medline: 22526075]
44. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Dec 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
45. Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med 2012;9(1):e1001160 [FREE Full text] [doi: 10.1371/journal.pmed.1001160] [Medline: 22291576]

Abbreviations

AMED: Japan Agency for Medical Research and Development
BMI: body mass index
BMR: basal metabolic rate
DIT: dietary-induced thermogenesis
DLW: doubly labeled water
EE: energy expenditure
FQ: food quotient
IRMS: isotope ratio mass spectrometer
MAPEs: mean absolute percent errors
MET: metabolic equivalent
PA: physical activity
PAEE: physical activity energy expenditure
PAL: physical activity level
rCO $_2$: carbon dioxide production rate
REE: resting energy expenditure
TBW: total body water
TEE: total energy expenditure
VCO $_2$: carbon dioxide output
VO $_2$: oxygen uptake

Edited by G Eysenbach; submitted 07.03.19; peer-reviewed by G Behrens, P Lee; comments to author 18.04.19; revised version received 04.06.19; accepted 09.06.19; published 02.08.19.

Please cite as:
Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, Sasai H, Ishikawa-Takata K, Tanaka S, Miyachi M
Accuracy of 12 Wearable Devices for Estimating Physical Activity Energy Expenditure Using a Metabolic Chamber and the Doubly Labeled Water Method: Validation Study
JMIR Mhealth Uhealth 2019;7(8):e13938
URL: https://mhealth.jmir.org/2019/8/e13938/
doi:10.2196/13938
PMID:31376273

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org